Horizontal Gaze Nystagmus: A Review of Vision Science and Application Issues

Steven J. Rubenzer, Ph.D. and Scott B. Stevenson, Ph.D.

ABSTRACT: The Horizontal Gaze Nystagmus (HGN) test is one component of the Standardized Field Sobriety Test battery. This article reviews the literature on smooth pursuit eye movement and gaze nystagmus with a focus on normative responses, the influence of alcohol on these behaviors, and stimulus conditions similar to those used in the HGN sobriety test. Factors such as age, stimulus and background conditions, medical conditions, prescription medications, and psychiatric disorder were found to affect the smooth pursuit phase of HGN. Much less literature is available for gaze nystagmus, but onset of nystagmus may occur in some sober subjects at 45° or less. We conclude that HGN is limited by large variability in the underlying normative behavior, from methods and testing environments that are often poorly controlled, and from a lack of rigorous validation in laboratory settings.

KEYWORDS: forensic science, driving while intoxicated, DUI, sobriety testing, horizontal gaze nystagmus, DWI, HGN, driving under the influence, operating while intoxicated, OWI

The Standardized Field Sobriety Tests (SFSTs) have become an important part of driving while intoxicated (DWI) enforcement since they were introduced in the 1980s. Consisting of three standardized psychophysical tests, failure on the SFSTs is used to establish probable cause to arrest and demand a breath test. The defendant’s performance on the SFSTs may also be introduced in most states as circumstantial evidence that the defendant is impaired by alcohol (1,2).

Of the three tests, Horizontal Gaze Nystagmus (HGN) has generated the most interest, both from scientific and legal perspectives. The other tests, Walk and Turn and One Leg Stand, arguably do not require any specialized knowledge to interpret, as many courts have held (1–4). Primarily, defendants are scored on behaviors that reflect lack of balance and coordination, symptoms of intoxication that have long been recognized. Legal tradition holds that any lay person can testify as to whether another person appeared intoxicated or not and that such judgments require no special expertise. In contrast, HGN’s indications of intoxication are more subtle and not common knowledge. Further, HGN has roots in laboratory science and clinical medicine. For these reasons and others, HGN has often been regarded as a scientific test requiring expert testimony before admitting it as evidence. Although this might seem to require testimony from a behavioral or medical scientist, some courts have taken judicial notice of HGN as a scientific test and have not permitted police officers to qualify as experts based on specialized training. Other courts do not deem HGN to be a scientific test (1–4).

HGN is controversial (5–8) and has been the subject of considerable advocacy by prosecutors and their experts and criticism by defense lawyers and their experts. Not surprisingly, there has been a polarization of opinion. At the time of this article, there have been no comprehensive scientific reviews of HGN from the perspective of eye movement science. This article will attempt to fill this void, focusing on laboratory studies of functional eye movement and gaze, including those that employed alcohol. It will not attempt to address physiology or diseases of the eye or nervous system in depth. We will begin with a brief description of the visual system and HGN. We will then address the empirical studies of HGN as a sobriety test, partisan arguments that support or criticize its use, and in the main part of the article, discuss empirical findings in the visual science literature that bear on its reliability and validity. Finally, we summarize our analysis and discuss the limitations of National Highway Traffic Safety Administration’s (NHTSA) (1,2) training program for police officers and the implications for use of HGN in a law-enforcement environment.

Overview of the Visual System

The retina is the tissue at the back of the eye on which light is focused and detected. The most sensitive portion of the retina is the fovea, a specialized area that is densely packed with receptors and allows maximum resolution and clarity of images. Animals with a fovea must be able to move the eye to a target of interest, then maintain the gaze to keep the image on the fovea. People are able to change the direction of their gaze in several ways, some of which are reflexive and others which are mostly voluntary. The smooth pursuit (SP) system allows the viewer to smoothly track a steadily moving object, as long as it does not go too fast, thus keeping the image on the fovea. In this way, a motorist can read a road sign, even as it moves relative to the body and the rest of the visual field. Generally, the smooth pursuit system is reported to be able to track smoothly moving objects up to a rate of 30°/sec (9–11), although texts (12,13), a review (14), and authors of individual studies (15–19) report that wide individual differences exist.
SP provides the means to maintain fixation on a moving target. In contrast, saccades are the primary means of changing eye fixation, either to a different object or to one that has moved too quickly or unpredictably to be followed by SP. Saccades are very rapid but require substantial preparation time to initiate—about 200 msec. Saccades are sometimes said to be “ballistic,” in that once launched, they cannot be recalled. But once under way, they are fast—up to 1000/sec—10 to 30 times as fast as smooth pursuit. Saccades take nearly twice as long to initiate as a smooth pursuit from the center of the visual field to the periphery, or about 30°/sec (1,2). This motion is designed to assess breakdown of smooth pursuit, as manifested by the eyes falling behind the target and saccades to bring the eyes back to the target. Next, the officer is to hold the stimulus as far to the side as the subject can focus and look for distinct (large amplitude) and sustained (at least 4 sec) nystagmus at maximum deviation (DSNMD). Lastly, the officer returns the stimulus to the midline and slowly moves it laterally, at approximately 10°/sec, looking for onset of sustained gaze nystagmus before 45° of lateral deviation. In the most recent training materials, NHTSA instructs students to estimate 45° by moving the stimulus to the side a distance equal to the distance to the subject. A secondary guideline, which was the only one before 2006, is to move the stimulus to align with the tip of the shoulder. This approach will usually result in gross underestimation of 45° (20,21), although this works to the advantage of the suspect.

Each procedure is repeated once. The subject is scored one point for each of the “clues” described for each eye; the presence of four clues is taken as evidence of intoxication. There is no guidance about how to score a clue that is present during one of the two administrations. Students are taught that several medical conditions (brain tumors, brain damage, disease of the inner ear) may produce nystagmus but are told these are uncommon among suspects they will encounter. Environmental conditions (wind, dust, etc. irritating the suspect’s eyes; visual distractions) are also noted as potential problems.

**Prosecution and Defense Claims**

There have been many claims about HGN from advocates and critics of HGN. In this section, we report some of them. The fact that an argument is cited should not be mistaken as indicating our support for that position.

The American Prosecutors Research Institute (22) asserted that HGN is the most reliable SFST and encouraged police officers to testify in support of this claim. HGN administration is said to be simple, including estimating 45° from the midline (23). Properly trained police officers are said to be able to distinguish HGN from other abnormalities of eye movement (24). One DWI Resource Prosecutor asserted that HGN signs are indicative of “visual dysfunction” (25), while another prosecutor asserted “HGN is not just an indicator of impairment; HGN is impairment” (26). Lastly, and unlike the other SFSTs, HGN has been touted as immune to practice effects (27–29).

In one of the first critiques by a defense attorney, Pangman (30) noted that the officer’s scoring cannot be verified and could be altered after seeing the results of a portable breath test. He noted that unlike field applications, HGN laboratory studies employed protractors and chin rests, that there are a number of other recognized causes of nystagmus, and that lay persons may mistake normal saccades for nystagmus. He cited NHTSA sources which report that the angle of onset of nystagmus (AON) decreased after midnight in drinking subjects, and citing a respected medical report that the angle of onset of nystagmus indistinguishable from alcohol gaze nystagmus if they deviate their eyes more than 40° to the side” (30, p. 2). Defense attorney Mimi Coffey (31), citing prominent eye movement researchers, noted that there are over 40 recognized types of nystagmus and asserted it is unrealistic that a police officer can distinguish these forms from alcohol-induced nystagmus. In *State v. Dahood*, medical eye specialists opined that DSNMD occurs in as many as 80–90% of normal subjects, has no value as a sign of pathology, and should be eliminated from the HGN test (7). One ophthalmologist testified that HGN is invalid because the stimulus
is held above eye level, thus involving muscles other than the lateral and medial rectus muscles that are primarily responsible for lateral eye movements (7). While NHTSA prescribes that each pass for breakdown of SP take 2 sec to go from midline to far gaze, for a speed of about 30/sec, this same physician asserted that the limit of lateral eye movement is 85° and that the proper speed should be 20/sec. Elsewhere he argued that a pass from midline to far gaze should take 4-sec (J. Citron, personal communication)—twice the lateral eye movement is 85°/C17620.

Rubenzer (8,32) pointed out HGN’s lack of validity data pertaining to mental, physical, or driving impairment and noted that the interrater and test–retest reliabilities cited for HGN are inadequate by conventional standards.

Empirical Reliability and Validity

It is well established that moderate amounts of alcohol (e.g., 0.08% blood alcohol concentration [BAC]) result in breakdown of SP and increased nystagmus when the eyes are turned out away from primary position (12,13,33). However, HGN is a highly specific implementation of these principles and must be evaluated on its own merits as a sobriety test. There are numerous other causes of nystagmus other than alcohol (13,33), and the role of anxiety, fatigue, and environmental conditions in HGN performance has not been thoroughly examined.

Since their inception, HGN and the other NHTSA SFSTs have been validated against estimated BAC rather than indications of mental, physical, or driving impairment. However, HGN is not admissible to establish a precise BAC, or in most jurisdictions, even whether the defendant is above or below the legal standard (i.e., 0.08% BAC). Some jurisdictions have statutes that require specific biological tests, and there is concern over the lack of precision in estimating BAC and the fact that the officer’s scoring cannot be verified (34).

HGN has been empirically evaluated in a number of laboratory and field studies. Substantial correlations with BAC are typically obtained (r = 0.51–0.77), as well as moderate levels of classification accuracy (8). However, none of the studies have been conducted in a truly blind manner. The laboratory studies excluded old or medically impaired subjects, were conducted during daylight hours, and did not invoke fear of arrest (7,8,32,35). Three large field studies, sponsored by NHTSA, reported high accuracy rates for the SFSTs and HGN (36–38). However, there are numerous limitations to these studies: Officers had the benefit of observing driving errors, the inside of the defendant’s vehicle, and the defendant’s demeanor. Stops appear to have been prompted by driving errors and circumstances (late weekend nights in close proximity to bars) that would constitute preselection of a high risk group. Not surprisingly, from 72 to 80% of those stopped were above the legal BAC limit. In two of the studies, officers were supervised in about half of the stops to ensure they performed the SFSTs correctly, and in the third, they had access to portable breath tests that may well have influenced their scoring. Lastly, in all three studies, officers were volunteers, experienced, highly motivated, and had just undergone refresher training. Thus, the accuracy rates reported for these studies may not be applicable to typical DWI stops (8,32) and should not be attributed to the field sobriety tests alone (39).

Rubenzer (8) reported that HGN clearly performed best among the SFSTs in studies such as those cited earlier, showing an average correlation with BAC of 0.65 across nine studies. Sensitivity to BACs above 0.08% or 0.10% was generally excellent (0.72–1.00) and specificity was usually good, although there were some clear exceptions. Likelihood ratios (an index of diagnostic power created by dividing sensitivity by the false-positive rate [FPR]) averaged 3.4–5.5 for BAC criteria of 0.04–0.10%, with better figures at 0.04%, although this was based on only two studies. These are respectable figures, but ones that are probably inflated by methodological problems in the studies. Variations in the speed or angle of administration for the smooth pursuit phase (i.e., horizontal vs. diagonal) do not appear to affect correlation with BAC (40), but the effects of such variations on diagnostic statistics were not examined. A recent study (41) found that false positives were not increased by variations from standard procedures, but false positives were unacceptably high (0.57–0.77) in all conditions. The position of the subject (standing, sitting, lying down) appears to have little effect (42–44), and administration of HGN on a boat provided positive results in two studies (45,46). Two unpublished studies found that HGN showed incremental validity over other observations (43) or field sobriety tests (45).

Diagnostic statistics such as sensitivity and FPR are potentially very informative, but when studies are not conducted blind or are otherwise flawed, their value is much reduced. In such circumstances, measures of reliability may be a better gauge of a test’s functioning and potential. Both the interrater and test–retest reliability figures reported for HGN (r = 0.59–0.71) (23,42,46) are modest for tests that provide the basis for arrest and, often, evidence of impairment in legal proceedings. All figures are well below the 0.80 standard advocated by Heilbrum (47) and far below the “bare minimum” of 0.90 recommended by Nunnally and Bernstein (48).

Thus, and as delineated elsewhere (8,32), the limitations of HGN and its supporting literature from a behavioral science perspective include: (i) Minimal evidence that it is related to driving or behavioral impairment, (ii) lack of true double-blind studies to provide an unbiased estimate of its relation to BAC, (iii) interrater reliability below accepted standards, (iv) susceptibility to medical conditions and some medications (addressed in detail below), and (v) the prevalence of potentially confounding factors present at many DWI stops, including anxiety, fatigue, and circadian rhythms. Laboratory studies of eye movement functioning have addressed some of these issues and raised others.

HGN and Visual Science

This next section will discuss vision science findings relevant to HGN. There have been a great many studies of smooth pursuit, far fewer investigations of gaze nystagmus. Almost none refer specifically to DSNMD. This clue and the angle of onset of nystagmus will be discussed together, as they presumably share common mechanisms.

Smooth Pursuit

Description and Functional Parameters—When confronted with a target that begins to move at a moderate speed, the smooth pursuit system initially responds with a stereotyped movement of the eyes in the same direction as the movement of the object, but this brief, initial motion is not tailored to the target’s velocity. The eyes then make a catch-up saccade to bring the image of the object onto the fovea (49) and begin the calibrated tracking of the object, generally matching its speed and direction (50). However, the match may not be precise, and saccades supplement the SP system to improve tracking performance (51). All six ocular muscles contribute to eye position in any direction (52,53), as they maintain tension at all times except when a saccade is made in the opposite direction of their angle of operation (54).
Eye researchers originally used pendulums to gauge SP, taking advantage of its natural sine wave motion. In the 1950s, electronics began to be used to present stimuli, and this allowed examination of a wide variety of stimulus parameters. While sine waveforms continue to be widely used, many researchers use constant velocity movements referred to as ramps, because a graph of the eyes’ position by time results in a straight, sloping line to the right. Triangle waves refer to two ramp motions, one away from midline, the other back toward it, joined without a gap (see Fig. 1). Researchers will sometimes merge multiple sine wave patterns to create pseudorandom stimuli to reduce predictability of the motion. Some researchers recommend the use of ramp motions (55), others sinusoidal (56), and some stress the need for examination of both (13).

The maximum speed of smooth pursuit tracking varies greatly across people, stimulus conditions, and how smooth pursuit is defined. While some researchers have described the maximal speed of smooth pursuit as 30–45°/sec, others have argued for both lower and higher figures. One researcher noted that some subjects could not “keep up” at 20°/sec (57), whereas another reported a steady decline in the adequacy of SP throughout the range of speeds examined (1.7–20.8°/sec) (58). Others have argued that much higher values are possible (59–63). Perhaps as a result of practice or a superior nervous system, professional baseball players have been able to keep their eye on a fast pitch as it approaches the plate with an angular velocity of 90°/sec (59).

There are several ways of gauging the adequacy of SP, which contributes to the different maximum values claimed. The most prevalent index is gain, which is the ratio of the eye speed to that of the target. At low speeds, such as 10°/sec, gain will approach 1.00, as the eye can match the target degree by degree. At higher speeds, gain drops precipitously. Adequate smooth pursuit is typically represented by gains of 0.90 or above. Unfortunately, reporting practices vary across research groups, so that maximum tracking speed estimates provided are not comparable from one study to another. Some authors remove catch-up saccades from the eye tracing record and calculate gain on only the smooth pursuit movements, while others do not. This can make a substantial difference, because saccades are much faster than SP and occur when the SP system is falling behind. In one study, subjects were able to track a target at up to 90°/sec with saccades, but only 60°/sec without them (64). At the highest target speeds, no attempt is made to track the stimulus, and the eye remains in its original position (9). Two methodological issues deserve comment. High sensitivity recording equipment will detect saccades that are overlooked by lesser instruments, so researchers should report the resolution of their observations. Second, gain should be assessed only on the smooth pursuit portion of the record, after removing all saccades.

A second method of gauging SP is to note the number and size of saccades the eyes makes in attempting to keep up with a fast target. One study suggested that this may be the most sensitive measure of SP performance (65), and unlike gain, it may be gauged without special equipment. Unfortunately, many authors do not report the size of saccades observed in the eye records, so it is unclear whether they would be visible without oculographic equipment, and their relevance to HGN is not clear. Other researchers, particularly in psychiatry, utilize a simple judgment rating scale (66) or the mean square deviation of the eyes’ path from that of the stimulus. Such studies have reported impaired SP tracking in 6–8% of normal populations (67,68), but another study reported a figure of 20% (69).

Smooth pursuit movements are nearly always accompanied by saccades, although at low target speeds, they may be too small to detect without recording equipment (58). While many studies examined the speed of smooth pursuit, there is surprisingly little normative data available regarding the presence of nystagmus or presence of catch-up saccades for subjects at different ages and for different target speeds and stimulus conditions. However, several authors have reported findings that, if valid, would render the use of SP as a sobriety indicator problematic. Flom et al. (70) reported that above frequencies corresponding to 30°/sec, tracking movements were entirely saccadic. Moser et al. (71) reported observing 6.8 saccades per 20 sec of observation time in sober subjects, while figures increased to 9.8 and 12.5 for subjects at 0.05% and 0.10% BAC. Stimulus speed was 15°/sec, and all saccades were 1.5° or larger, but no further information is given. It is not clear whether saccades of 1.5° would be visible to a police officer at roadside, or scored if they were observed. Schalen (18) reported sizeable (3–10°) saccades occurred 6 times per minute at 10°/sec, with 11, 20, and 33 such saccades observed at 20, 30, and 40°/sec. Ross et al. (72) reported an average of 50.5 catch-up saccades per minute among a group of 37 normals, pursuing a target at 16.7°/sec. The average saccade was 2.5° (SD = 0.5), presumably large enough to be observed by the naked eye. However, in contrast to the previous findings, several authors have reported tracking at 30°/sec or more with few or no sizable saccades (11,73,74).

The smooth pursuit pass for HGN requires 2 sec per eye, or 4 sec to go from maximum deviation from one side to the other. Thus, to observe one such normal saccade during a typical 2-sec pass (one eye), a rate of 30 saccades per minute would be required, on the average, and this would produce two saccades during the pass across the whole visual field. Some studies have exceeded the 30 saccades per minute rate (70,72), but most have not (11,18,71,73,74). The actual rate of saccades per minute during HGN SP under the influence of alcohol has never been reported.

Age results in a decline in several aspects of smooth pursuit performance (12,75–80), possibly as a result of atrophy of cerebral cortical neurons or loss of cerebellar Purkinje cells. As people age, they react less quickly to the initial stimulus movement (78), show reduced gain, and require more catch-up saccades to track adequately (12,76–78,80). As one researcher stated, “the diagnosis of abnormal pursuit must be qualified by the age of the patient.
Smooth pursuit is an age-dependent motor system (80), p. 465. While younger subjects are able to maintain gains of over 0.90 to targets moving 30–40°/sec, the gain for elderly subjects fell below 0.90 when targets exceeded 5°/sec in one study (80). A significant decline in maximum gain may occur as early as after age 30 (76; see Fig. 2). There is some evidence that women, particularly older women, perform less well than men the same age (77).

Qualities of the target also influence SP. Bright targets elicit greater eye acceleration, at shorter latencies, than do dim targets (13). If ambient light is poor, parfoveal tracking may be preferred, at which time rods are more efficient photoreceptors than cones (13). Thus, quality of illumination may be important, and police officers might be advised to use a light-emitting (but nonaversive) stimulus during nighttime DWI investigations. Maximum velocity for nasal-ward eye movement is typically higher than for movements toward the temple (51), and higher for targets at the center of the visual field than for eccentric targets (11). Smooth pursuit is relatively robust to stroboscopic (intermittent) illumination (81,82), but blinking causes a brief reduction in eye speed, followed by a catch-up saccade. Blinking causes the contraction of all the eye muscles, and the disruption to SP may be the result of a central process (decreased activity in omnipause neurons) rather than vision loss during the eye closure (83). The predictability of a target’s motion also greatly influences the accuracy of smooth pursuit, so that a steady velocity of stimulus movement is important (13). Time to change from one speed to another unanticipated speed requires about 133 msec (49), and such a change often generates a small saccade (10). It should be noted that these studies utilized mechanical or electronic presentation of stimuli that could change speed instantaneously, which presumably would not be the case for a human administrator.

The role of attention in smooth pursuit performance is complex and not fully understood, in large degree because pursuit movements are a combination of voluntary (SP) and reflexive (OKR) responses to target motion. Some researchers use the term “smooth pursuit” to describe the putative mechanisms of voluntary tracking, while others use the term to describe the behavior of following a small target. Deficits in nonvoluntary attention have been proposed as the reason for the deficits observed in schizophrenia and attention deficit disorder (ADD), and attention areas of parietal lobe are implicated in target selection for pursuit. However, smooth following eye movements also involve a visual reflex, dependent primarily on a relatively simple cortico-pontine-cerebellar pathway, linking visual sensory areas involved in the processing of motion signals to motor regions of the cerebellum, via the pontine nuclei (84).

Because SP has been conceptualized by some as an automatic process, it has been theorized that distracter tasks will draw attention away from SP, allowing it to function unencumbered by disruptive intentional processes. Results have been decidedly mixed. Some distractions have been observed to increase blinking, saccadic intrusions, and “velocity arrests” (disruption of SP) in some studies (85–88), but to improve smooth pursuit in others (89–91). Barnes and Crombie (92, p. 550) reported SP was “heavily influenced by the presence of any unintended static peripheral cues, hence the need to operate in conditions of complete blackout.” On the other hand, when the subject is required to pay attention to a quality of the stimulus (i.e., changes in color), SP is enhanced (93), as it is with larger or high contrast stimuli (13). And while HGN might seem a largely nonverbal task, the instructions subjects are given influence the quality of tracking (18). Several researchers have commented on the need to prompt older or fatigued subjects to pay close attention and try hard to obtain optimal results.

The presence of a patterned background, or competing stimuli, tends to lower the gain of smooth pursuit by about 10–20%, particularly if the background contains sharp visual boundaries perpendicular to the direction of eye movement or meaningful background images (58,94,95). Disruption is the greatest for distractant material in the same visual plane, less so for material behind or in front of the stimulus (96). Interference appears to be greater when the target is toward the periphery (92,94,97,98). If the background moves in the opposite direction of the target, interference is increased; if it moves in the same direction, SP may be facilitated (95,98).

Fatigue is sometimes represented as a cause of diminished SP, although often without supportive references or ones that fail to support the assertion (12,15,99). Barnes and Crombie (92, p. 550) stated that the nystagmus they observed was “susceptible to changes in arousal, responses diminishing with increased drowsiness,” but presented no quantitative data. Another study examined a single subject after 30 h without sleep and reported that tracking was almost entirely saccadic (15). Subjects showed 10% fewer saccades in the morning than in the afternoon in another investigation (100). An unpublished study reported that 24 h of wakefulness did not affect smooth pursuit in HGN (101). However, two published studies (102,103) reported degradation in smooth pursuit after substantial sleep deprivation (40 h in one), but little loss of performance until sleep loss exceeded 24 h or subjects reported high levels of sleepiness. It should be noted that there are several possible types of fatigue: muscle, mood state, and fatigue that is often presumed by time of day, time without sleep, or amount of previous activity. While it is unlikely that muscle fatigue plays a role in HGN, the others may be relevant.

One of the claimed benefits of HGN is resistance to training or practice. Some studies have found no effect of retesting SP in schizophrenic or psychotic patients (104,105), but others have reported a powerful effect of active training or biofeedback (12,106,107). None has examined the ability to suppress nystagmus under the influence of alcohol, however.

Thus far, we have only discussed SP in terms of drift and catch-up saccades. However, a number of researchers have described other types of saccades that may supplement, or interfere with, SP. Leading or anticipatory saccades take the eye ahead of the moving target and may reflect a loss of crucial inhibitory control. Some
have suggested that the saccades produced by normals in SP tasks are not catch-up saccades, but leading ones (88), whereas many researchers do not distinguish among various types. One study that did distinguish them found the most fast eye movements were not catch-up saccades but were in the direction of target travel (100).

**Physiology and Effects of Medical Conditions and Drugs**—A substantial number of brain centers control smooth pursuit. These include the neural integrator, a group of structures located in the nucleus prepositus hypoglossi. Situated in the lower brainstem, the neural integrator coordinates the velocity and position changes required for all conjugate (paired) eye movements. Other subcortical structures that support horizontal SP include the Dorsolateral Pontine Nucleus, the brainstem, and the cerebellum. Cortical areas include primary visual areas of the occipital lobe, motor areas in the Frontal Eye Fields, motion processing areas in the Middle Temporal and Medial Superior Temporal areas, and the VI and attention-related areas of the occipital lobe (84).

Many medical conditions affect SP, including Parkinson’s disease, progressive supranuclear palsy, cerebellar disorders, hepatic encephalopathy, Alzheimer’s disease, and large cerebral lesions (13). Hartje et al. (108) reported that patients with lesions (vascular, neoplastic) in one brain hemisphere or diffuse brain damage (traumatic, inflammatory, degenerative diseases, epileptic) showed many large saccades during smooth pursuit. None of these patients had evidence of cerebellar or brainstem dysfunction, indicating cortical damage was sufficient to disrupt SP. Another group of researchers (109) found 76% of patients with a confirmed diagnosis of multiple sclerosis, and 25% of patients with optic neuritis, showed impaired smooth pursuit, which manifested as greatly increased number of saccades to stimuli moving more than 35°/sec. Abnormal SP was found in 46% of patients with generalized vascular disease (long-standing hypertension, diabetic vasculopathy, arteriosclerosis), 69% of 32 patients with localized eye disorders (cataract, glaucoma, scleritis, optic atrophy, leukemia, confusion of the globe, retinal aneurism or detachment), and 73% of 325 patients with the diseases of the central nervous system such as Parkinson’s disease and Alzheimer’s dementia (74). Citek (110) reported that hypoglycemia will produce disruption of SP, but no other eye symptoms. Aside from neurological or muscular dysfunction, physical obstructions, such as a tumor in the eye socket, can impair smooth pursuit. In Brown’s syndrome, the action of the superior oblique muscle is limited by a restriction in the orbit or by resistance at the point where its tendon passes through the trochlea. Barbiturates and other depressants impair SP, as do medications used to treat pain, seizures, agitation, mood swings, anxiety, and insomnia (see Table 1), several of which are common conditions.

An extensive list of factors that may potentially produce nystagmus has been articulated in court cases such as *Schultz v. State* (111), but this list is not complete and a number of the causes cited have not been demonstrated to cause eye movement abnormalities.

Some psychiatric disorders also affect SP. The first such study noted a parallel between eye behavior and psychiatric symptomology: Manics had a tendency to overshoot the target when making corrective saccades, whereas depressives showed “overdamped” responses (112). This and many studies have reported low gain and smooth pursuit interrupted by saccades in schizophrenics and their first degree relatives (68,113,114), but the degree of psychosis, rather than diagnosis, appears to be the strongest predictor of poor SP (66,68). Findings for patients with bipolar disorder (manic depression) have been mixed (66,115). About 50% of acutely ill manic patients show impaired smooth pursuit, while impairment in

<table>
<thead>
<tr>
<th>Chemical Class or Name</th>
<th>Common Trade Names</th>
<th>Street Names</th>
<th>Use</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benzodiazepines</td>
<td>Xanax, Valium, Ativan, Klonopin, Restoril, Serax, Rohypnol, Halcion, Librium, Dalmane, Mogadon, Ambien</td>
<td>Downers, Tranks, Blues, Yellows</td>
<td>Treatment of anxiety, muscle spasms, insomnia, agitation, seizures</td>
<td>SP</td>
</tr>
<tr>
<td>Phenytoin</td>
<td>Dilantin, Phenytek</td>
<td>NA</td>
<td>Treatment of seizures</td>
<td>SP, GN</td>
</tr>
<tr>
<td>Carbamazepine</td>
<td>Tegretol, Biston, Calepsin, Carbatrol, Epitol, Equeetro, Finlepsin, Sirtal, Stazepine, Telesmin, Timonil</td>
<td>NA</td>
<td>Treatment of seizures</td>
<td>SP, GN</td>
</tr>
<tr>
<td>Barbiturates</td>
<td>Phenobarbital, Seconal, Amytryl</td>
<td>Barbs, Downers, Reds, goofballs, Yellow Jackets, Blue Devils</td>
<td>Treatment of agitation, pain, anxiety</td>
<td>SP, GN</td>
</tr>
<tr>
<td>Lithium carbonate</td>
<td>Carbolith, Cibalith-S, Duralith, Eskalith, Lithane, Lithizine, Lithobid, Lithonate, Lithotabs</td>
<td>–</td>
<td>Treatment of bipolar disorder, mood swings</td>
<td>SP, GN</td>
</tr>
<tr>
<td>Nicotine</td>
<td>NA</td>
<td>–</td>
<td>Recreational</td>
<td>SP*</td>
</tr>
<tr>
<td>Narcotics</td>
<td>Morphine, Codeine, Dihaludit, Percodan</td>
<td>M, Morph, smack, junk, horse</td>
<td>Treatment of pain</td>
<td>SP</td>
</tr>
<tr>
<td>Choral hydrate</td>
<td>Noctec</td>
<td>“Mickey Finn”</td>
<td>Produce sleep</td>
<td>SP</td>
</tr>
<tr>
<td>Nitrous oxide</td>
<td>NA</td>
<td>Buzz Bomb</td>
<td>Calm patients, analgesia</td>
<td>SP</td>
</tr>
<tr>
<td>Phencyclidine</td>
<td>NA</td>
<td>PCP, wet, fry, angel dust</td>
<td>Animal tranquilizer</td>
<td>GN¹</td>
</tr>
</tbody>
</table>

Adapted from Leigh and Zee (13), Table 12–11.

SP, smooth pursuit; GN, gaze nystagmus.

*Causes square wave jerks; its effects on SP are variable.

¹Causes nystagmus even at primary position.
such patients in remission may be because of lithium treatment (115). Studies have found disrupted SP in schizotypal patients and in subjects deemed “at risk” for schizophrenia (67,105,116). A mixed group of “26 neurotics undergoing therapy” showed larger and much more frequent saccades than did the normal control group, although considerably less than the neurologically impaired sample (108). Abnormal SP has been found associated with self-report depression scores (114), but not elevated neuroticism scores (116) or experimentally induced stress/anxiety (90). Children with ADD show reduced smooth pursuit efficiency and saccadic intrusions, although the only study to examine ADD in adults did not find significant impairment in most subjects (72). Lastly, 42.4% of chronic alcoholics were found to have impaired smooth pursuit, marked by prominent saccadic movements, compared to 20.0% of age-matched controls (69).

There is some evidence that stimulant drugs can lessen SP impairment in affected populations. Nicotine tends to improve SP in schizophrenics (117,118), and Ritalin did so in one study of ADD children (119), but not another (120). Thus, it is possible that the use of common stimulant drugs in combination with alcohol may mask breakdown of SP and pose a substantial challenge to law-enforcement use of HGN.

Nicotine is one such drug, and its effects are complex. It is associated with a peculiar pattern of movement called bow tie nystagmus, named for the shape the eye movement takes. The movement was produced at primary position in the dark and disappeared when the subject fixated on a target (121) and thus is not likely to be confused with HGN. Another study reported that smoking produced various forms of nystagmus in 16 of 25 subjects, apparently at primary position (122). Smoking a single cigarette has been reported to impair SP (123), and nicotine ingestion can lead to an increase in square wave jerks ([123,124]; see Fig. 3), a pattern of saccades that could easily be mistaken for alcohol-induced nystagmus. Several other authors reported no effect of nicotine in normals (vs. schizophrenics) (117,125,126), and two studies reported that nicotine improved SP slightly, albeit in only one eye (118,127).

Reliability—The test–retest reliability of SP measures has been thoroughly examined in schizophrenics and found to be adequate for research purposes, but only a few studies have addressed performance in normals or drinkers. Ettinger et al. (128), using gain and frequency of saccades as dependent variables, found good retest reliability for SP at 36 and 48 sec among sober subjects, but not for movements at 12 and 24 sec. Use of catch-up saccades as a dependent measure yielded lower reliabilities than for gain. Using a criterion of interclass correlation (ICC) > 0.40 as indicative of good reliability, all values met this standard except for gain at 12 and 24 sec and catch-up saccades at 12 sec. The ICC value for gain at 36 sec exceeded the cited 0.75 standard for excellent reliability. Cronbach’s alpha (used to assess internal consistency) was above 0.75 for all variables except for catch-up saccades at 12 sec (see Table 2). It was not reported whether the reliability differences across conditions were statistically significant. Another set of researchers (129) also found test–retest reliabilities were somewhat better for faster moving stimuli and somewhat better for retesting immediately (r = 0.73) or after 2 weeks (r = 0.77) than after 2 years (r = 0.65).

However, the usual meaning of test–retest and internal measures of reliability may be inverted in the case of SP used as a sobriety test. Because all subjects should perform quite well when sober, with few saccadic intrusions, there should be few reliable
differences between individuals or trials, and this is reflected by low reliability figures.\(^2\) Conversely, reliable individual differences were present for the higher stimulus of 36\(^\circ\)/sec or faster, indicating that stimulus speed is a critical variable in correct performance of HGN. The reliability paradox does not apply to interrater reliability (which Ettinger et al. did not examine), as agreement among qualified observers is essential, at least for within the parameters (e.g., stimulus speed) that HGN is likely to be performed.

Mundt et al. (130) did not observe Ettinger et al.’s pattern of results, finding low test–retest reliability whether sober subjects were presented targets at either 20 or 40\(^\circ\)/sec (\(r = 0.35\)). The dependent variable was number of saccades multiplied by their size. Test–retest correlations when subjects were dosed to 0.08\% BAC were 0.64 and 0.65, but scores when not drinking were strong predictors of performance when drinking (\(r = 0.51–0.56\); see Table 3). There was a considerably greater increase in error scores (from sober to dosed condition) for pursuit at 20\(^\circ\)/sec than for 40\(^\circ\)/sec (not shown in table), probably because the higher speed generated saccades in more people in the sober condition.

Influence of alcohol—A number of researchers have reported that SP impairment begins by 0.05% BAC (71,131–138), although the amount of impairment at this level may be minimal (131) and in one study, SP deficits had largely disappeared as BAC fell to 0.047% several hours after drinking (16). Several groups have reported that the degree of SP impairment increases from lower to higher BAC levels (70,71,131–134). Flom et al. (70) reported that breakdown of SP began at 0.02% and increased steadily through 0.12%. Another group (135) reported significant decrements in SP at 0.03% BAC. Wilkinson et al.’s (134) charts indicate the substantial impairment of SP at 5 min after the consumption of alcohol, at which time BAC was only about 0.015%. No studies were located that reported normal SP at BACs > 0.05% BAC.

Gaze Nystagmus

The natural resting position of the eyeball in a normal, awake individual is facing straight ahead, called primary gaze. Large deviations from primary gaze are unusual in daily life, as people naturally turn their head to fixate on or follow targets that are more than 15–20° off the midline (13). To look to the side, as required in the HGN test, the oculomotor system must overcome the substantial elastic, restorative forces of the muscles and connective tissues that support the globe. This requires a finely calibrated and very steady source of nerve innervation under control of the neural integrator. A “leaky integrator” is unable to maintain steady eye control, causing the eye to drift back toward primary gaze and resulting in gaze nystagmus. The effort required to maintain gaze away from the midline increases in direct proportion to the angle of deviation, 1.2 g/degree (49). Alexander’s law, originally applied to vestibular-induced nystagmus, holds that the amplitude of nystagmus increases as the angle of deviation is increased (13,137). Thus, the HGN clue of DSNMD appears closely related to angle of onset of nystagmus (AON) and perhaps redundant with it.

In HGN, the clues of DSNMD and onset of nystagmus before 45° rely on eliciting nystagmus as the eyes are turned away from the midline. Readers should be aware that terminology can be confusing in this area. A number of researchers refer to “endpoint nystagmus” but are not referring to nystagmus at maximum deviation as one might expect. Rather, it simply means nystagmus when the eyes are deviated to the side, sometimes as little as 25°.

Although an authoritative text (13) stated that the gaze fixation system is relatively robust to changes in target size, luminance, color, and distance, no references were cited in support of this conclusion. While many parameters for SP have been investigated, GN has received relatively little attention. In the following section, we will discuss the few findings available regarding possible confounding factors before moving on to detailed discussions of normative studies, reliability, and the effects of alcohol.

Many medical conditions and medication, as discussed previously, can cause nystagmus during either smooth pursuit or lateral gaze (see Table 1). There is very little literature on the effect of psychiatric disorders on gaze steadiness. One study found adult ADD patients had difficulty maintaining fixation, often making inappropriate saccades off-target (138), while results for persons at risk for schizophrenia are mixed (105,139).

Several unpublished studies found no effect of circadian rhythms or fatigue on gaze nystagmus in subjects who are not drinking (23,101,140), although alcohol (0.10% BAC) decreased the angle of onset of nystagmus after midnight in two of them. However, one was based on only five subjects and the other ten, five that may have been subjects in the first. In contrast, the only peer-reviewed study (141) reported that 55% of subjects showed DSNMD in one or both eyes after sleep deprivation of 24.5 h and 13–14 h of continuous mental and physical activity.
The prevalence of gaze nystagmus in normal groups has been examined by several groups. The HGN developers (23) reported that 50–60% of individuals show some small, transient nystagmus at maximum deviation, a figure congruent with at least one medical source (142), but Booker (141) reported 19% of sober subjects showed DSNMD in one or both eyes. Coates (143) declared that the angle of onset of nystagmus (AON) in normal population is 40–45° but did not cite any data or references. A series of studies of 115 normal subjects (144) found that no subject had onset of nystagmus at 40° or less of lateral deviation, and a subsequent investigation led the authors to conclude that nystagmus at <40° should be viewed as a sign of a disturbed equilibrium in the oculomotor system. This conclusion was seconded in another large study (145), which found the average AON was 52.25° in both eyes, with a standard deviation of 6.5°. However, examination of the data (see Table 4) shows nearly 10% of subjects showed AON at <45°, and an additional 9.2–14.5% showed AON at this point. Sixty-three percent of 260 subjects showed some nystagmus at maximum deviation; 13.5 percent of these showed nystagmus in only one eye. Most surprisingly, older subjects showed less nystagmus at maximum deviation than younger subjects. Several other studies that examined gaze nystagmus (146–148) used much smaller samples. Schmidt and Kommerell (148) reported that some of six subjects displayed onset of nystagmus at 45–50° of lateral deviation within a few seconds, and all did with enough time. Further, the amplitude of nystagmus was large and persistent—possibly large enough to qualify as DSNMD. With the exception of Booker, none of the earlier mentioned studies examined the presence of DSNMD in sober subjects, and we were unable to find any empirical work that reported such findings.

The maximum angle of eye deviation from the midline varies among subjects, with average figures of 55–59° for age groups up through 50 (149,150), and a standard deviation of about 6° (149). Each eye is able to gaze 2–5° further to its own side (abduction) than across the midline (adduction). Older subjects showed reduced maximum range, both toward and away from the midline (149,150).

Only one study has examined the concordance of SP deficits and gaze nystagmus. From an original sample of 623 patients with one or both deficits, 52 patients were selected based on several exclusion criteria. The authors reported that “a substantial gaze-evoked nystagmus of more than 3–4°/sec at 40° lateral gaze is always [italics added] combined with a SP deficit” ([151], p. 387). Twenty-five percent of subjects had SP deficient without gaze nystagmus. The authors noted that the neural integrator is essential for all conjugate eye movements, while lesions to cortex, brain stem, and cerebellum can disrupt SP without impairing gaze holding.

Reliability—Reliability of gaze nystagmus measures was examined by the same two research groups that provided such data for SP. Ettinger et al. (128) examined gaze stability at primary position and at 12° left and right gaze in sober subjects. Both internal and test–retest reliabilities were low (r = 0.45 and 0.55, ICC = 0.54). However, the relevance of these figures to HGN is unclear. When fixation stability is used as a sobriety test, large angles (30–65°) of eccentric gaze are utilized, resulting in substantial effort to maintain gaze well outside its natural range. Further, Ettinger et al.’s subjects did not consume alcohol, and as explained previously, low test–retest reliabilities may be desirable. Mundt et al. (130) examined the number and size of saccades when sober and drinking subjects were asked to focus on targets at 35, 40, and 45° eccentric. The test–retest correlation for subjects tested and re-tested at c. 0.08% BAC. Sober performance predicted performance when dosed with alcohol (r = 0.45). No studies have directly examined any of the various types of reliability (internal, test–retest, interrater) specifically for DSNMD or AON (as opposed to generalized gaze nystagmus) following consumption of alcohol.

Effects of alcohol—In the seminal study on gaze nystagmus and alcohol, Aschan (152) reported that most subjects showed the onset of gaze nystagmus at 40° lateral gaze for 0.06% BAC. While subjects tended to show the same thresholds for the ascending and descending limb of the BAC curves, there was substantial intersubject variation. Aschan noted that gaze nystagmus increased in intensity with increased eccentric fixation, consistent with Alexander’s law, and appeared at smaller angles when one eye was covered.

Only one study has examined the HGN clue of the onset of nystagmus before 45°. Burns et al. (personal communication, 2000) attempted to identify the BAC threshold from this clue using laboratory equipment and blind testing procedures. All subjects above 0.08% BAC were accurately classified, whereas three of nine subjects that averaged 0.06% BAC were misclassified.

Unlike the other HGN clues, DSNMD was introduced in the second SFST laboratory study (23) without explanation or references. The only empirical study located that directly evaluated this HGN clue found reason for caution in its use. Booker (141) did not report diagnostic statistics for subjects while intoxicated, but reported that 62% of subjects continued to show DSNMD in at least one eye immediately after all alcohol had cleared from the blood. Only one rater participated, so interrater reliability could not be examined.

Several studies reported that AON correlated with BAC 0.70 or more using basic instruments (23,153,154). NHTSA researchers (23) examined this variable, but for field use recommended a dichotomous rating of whether or not the AON was <45° or not. However, according to their data, this decision point yields an estimated BAC of about 0.06%. Officers were directed to estimate this angle by alignment with the shoulder, as the researchers apparently deemed the use of a cardboard protractor as too cumbersome in the field. However, the correlation of the officers’ estimates with the actual angle was only 0.58. When a continuous AON was measured mechanically, it achieved a correlation of ~0.71 to ~0.72 with BAC and higher classification accuracy than any test, or the entire battery of tests, used in the study. Two other studies observed the correlations of ~0.76 (154) and ~0.88 (153) with BAC in actual traffic stops. ER physicians also observed a highly linear relationship of AON with BAC extending up to 0.40% BAC (153). This is important, as many behavioral indicators of intoxication show marked insensitivity in subjects who have developed a tolerance to alcohol, whereas it appears AON does not.

A fourth study on this issue was conducted in the 1980s and cited in several important HGN critiques and court cases. The following quote from State v. Witte (155) cites Pangman (30), who is in turn referring to Norris (personal communication, 1985):

### TABLE 4—Angle of onset of nystagmus in 131 subjects.

<table>
<thead>
<tr>
<th>Angle from Midline</th>
<th>Percent of Subjects Showing Nystagmus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left (%)</td>
</tr>
<tr>
<td>35</td>
<td>0.8</td>
</tr>
<tr>
<td>40</td>
<td>9.2</td>
</tr>
<tr>
<td>45</td>
<td>9.2</td>
</tr>
<tr>
<td>50</td>
<td>33.6</td>
</tr>
<tr>
<td>55</td>
<td>27.5</td>
</tr>
<tr>
<td>60</td>
<td>16.0</td>
</tr>
<tr>
<td>65</td>
<td>3.8</td>
</tr>
<tr>
<td>70</td>
<td>0.0</td>
</tr>
</tbody>
</table>
The data in the study revealed that there was virtually no correlation between the actual value of blood alcohol concentration and the predicted value based upon the angle of onset of nystagmus. However, a correlation did develop between the breath alcohol reading and the level predicted by the alcohol gaze nystagmus. Interestingly, the study concluded that this was caused by the very subjective nature of the test itself: “Since the police officers are the ones operating the breath testing equipment, it appears that, at least in some of the cases, an already known breath alcohol value may have influenced the determination of the angle of onset.”

This summation is incorrect and misleading. In fact, an $R^2$ of 0.53 ($r = 0.73; n = 38$) was reported when AON was used to predict blood or urine alcohol levels (proportions of blood and urine analysis not reported), whereas a considerably higher value ($R^2 = 0.81; r = 0.90; n = 88$) was observed when AON was correlated with breath test results. While the difference in correlations is significant, the former value is hardly near to zero and it may be because of the problems inherent in urine alcohol analysis rather than observer contamination.

Conclusions from Visual Research and HGN’s Use in DWI Investigations

Eye movement research has shed some light on the HGN issues noted at the beginning of the article and raised some new ones. A review of the vision science literature finds: (i) The NHTSA description of SP as a marble rolling on a glass table does not take into account the irregularities and small saccades often found in the SP of sober subjects, (ii) the NHTSA-recommended speed for SP may be too fast to produce smooth pursuit velocity without visible saccades for some people when sober, even when performed perfectly under favorable conditions, (iii) age has a considerable effect on SP, (iv) increased stimulus speed during SP is associated with not only more saccades, but also higher test–retest reliability of SP measures, (v) visual background characteristics can lower SP quality, (vi) predictable stimuli are followed better than unpredictable targets, (vii) a number of medical conditions and prescription drugs interfere with SP and increase GN, (viii) psychosis and possibly other psychiatric conditions, such as mania and depression, affect SP, (ix) performance when sober on SP and GN tasks strongly predicts performance while drinking in the only study to examine the issue, and (x) test–retest reliability of both SP and gaze nystagmus may be too low for forensic use even when instruments are used and needs further investigation. HGN’s interrater reliability was identified as a serious shortcoming of manual HGN administration in the first author’s previous review (8). Several other potentially important variables (effect of low temperature, wind) have not been addressed at all. Fatigue and anxiety are potentially confounding factors in most DWI stops and have received very little rigorous study. Currently, one published study found that anxiety does not affect smooth pursuit. One unpublished study found no effect of fatigue on any aspect of HGN, while two published studies found decrements in SP when subjects were substantially sleep deprived. Another found DSNMD in 55% of those subjected to experimentally induced fatigue. Two studies, one unpublished, suggest that minor variations from the standard administration may not affect HGN’s correlation with alcohol, but the effect on diagnostic statistics was addressed only in the unpublished article, which also found high false-positive errors for all conditions.

However, one criticism of HGN can be dismissed as unsupported by either rational or empirical evidence: Administering HGN slightly above the eye line, as specified in the standard protocol, does not invalidate the test by involving additional eye muscles: all six are involved during all SP movement. Further, this is the only protocol that has received any empirical evaluation.

The results reported here contrast somewhat with the first author’s review of HGN sobriety test studies (32). In that review, HGN performed better than other sobriety tests at predicting estimated BAC, although none of the studies were truly blind, and the findings were regarded as circumstantial evidence of validity. In contrast, the eye movement research reviewed here suggests many potential problems. In considering the laboratory vision research, there are important differences between studies of smooth pursuit conducted in laboratories and those that examine SP as part of a field sobriety test. In laboratory studies, the stimulus is presented mechanically or electronically with high precision to the center of the visual field rather than across a span of 120°. Usually, a high contrast stimulus is used against a blank or featureless background. Often many trials are conducted and initial or flawed trials may be discarded. Subjects in laboratory studies are typically screened for health or psychiatric problems that might impair performance and they have no reason to fear “failing” the procedure. Although not specified, presumably most testing is done during the day or early evening. Many of these differences would seem to make SP more challenging in field conditions than in the laboratory settings, making the discrepancy between laboratory eye studies and those of HGN studies all the more puzzling. It is possible that HGN scoring is more affected by the external cues of intoxication than other sobriety tests, and this factor leads to higher, but contaminated, correlations and diagnostic statistics. There are two plausible reasons. HGN scoring appears more difficult and subjective (interrater reliability coefficients are lower) than for other SFSTs, and the fact that the subject’s eye movements are not recorded may lead to a lessened sense of accountability: The defendant’s performance on other sobriety tests is usually recorded on video, and the jury can compare the officer’s account with their own judgments. Either factor may produce scoring errors influenced by observations other than formal HGN clues. Alternatively, perhaps real-life stimuli and backgrounds (as opposed to laser spots and visually sterile environments), and the fact that subjects are assessed on their first few SP settings, facilitate performance over that observed in laboratory settings. Possibly officers are able to distinguish HGN clues from other eye movement irregularities, a claim advanced by one HGN advocate (24).

As noted in the introduction, interrater reliability is probably a serious problem in field HGN, and even the modest figures reported may underestimate the problem. Pearson’s $r$, when used to judge interrater reliability, is primarily sensitive to agreement in the rank order of subjects across raters: It does not account for differences in the values assigned. In other words, two examiners could differ greatly in terms of how many subjects score four clues or more but agree which subject is highest, second highest, lowest, etc. They would nonetheless show high interrater correlation on Pearson’s $r$. Several authors (156–158) have recommended the use of an intraclass correlation coefficient (ICC) to gauge agreement on continuous measures because it takes account of both ranking similarity and agreement on the absolute level.

There are numerous possible sources of HGN interrater disagreement. The original NHTSA researchers reported that officers had considerable difficulty in accurately estimating an angle of 45° within 3° (23). There is also subjectivity in the judgment of when nystagmus has occurred, as its amplitude may vary from far less than 1° to more than 20°. There is little agreement among professionals about the threshold of observation: One author asserted...
nystagmus of 0.1° is detectable (143), while another group reported that nystagmus of 0.3–0.5° was not visible to the naked eye (159). The second author’s experience with optometric clinicians suggests that 1° is the smallest amplitude of jerk nystagmus that clinicians can reliably detect and classify. Predictability and reliability of scoring will break down for nystagmus that is near the threshold of observation. Many people have nondistinct endpoint nystagmus when sober, and there are no studies that assess how well officers can distinguish this from DSNMD. A high quality video recording of the eyes during HGN could preserve the evidence and allow inspection by the defense lawyer, the jury, and experts, but also risks revealing nystagmus that would not be visible to the naked eye if magnification is used. Interrater reliability appears to be less problematic in research when eye tracings are used.

Another factor contributing to poor interrater reliability may be the diffculty in administering the SP phase with the correct motion and uniform velocity. In video tapes of police-administered HGN, it is common to see noticeable changes in stimulus speed and arcing in the path of the stimulus in either horizontal or vertical planes. Moving a stimulus at a constant speed in a straight line is not a natural motion, requires movement along three joints (shoulder, elbow, and wrist), and is not easy to do. The variations among different test administrators may have a direct effect on HGN’s validity. If the officer takes the recommended 2 sec for each pass (from midline to extreme gaze), this translates to an average speed that is barely within many people’s smooth pursuit capacity. Any variations from this average will result in some portion of the pass being faster than the prescribed rate, quite possibly causing the eye to lag the stimulus and to catch up using a saccade. These problems, and the difficulties officers have in estimating angle of deviation with passable accuracy, suggest the need for greater training or the use of practical roadside implements. Limitations in the standardization and reliability of scoring limit the validity and classification accuracy that a test can achieve (48).

While interrater reliability and calibration problems might be addressed through instrumentation, limited test–retest reliability is both more interesting and possibly less easily resolved. Even when recorded by instrument, substantial variations in eye movement variables are observed over time, even when subjects are sober at both assessment periods. The general effects of alcohol also tend to be unreliable within and across episodes (130,160,161). Although eye movement performance and alcohol responsivity may covary, this is only one possibility and has never been tested.

An essential aspect of all scientific measurement is that the instruments used are carefully calibrated. This means that the values returned by the instrument have been verified against accepted standards, and through repeated measures, the variability in measurement as a result of instrument noise is known. In the case of HGN testing during SFST, the “instrument” is the arresting officer, who makes a subjective assessment of whether nystagmus is present or “distinct.” Presumably, this varies with the number and size of saccades, but no studies have examined the HGN administrator’s judgments of saccade frequency and amplitude against an objective eye tracker recording. The “instrument,” in the case of HGN, is uncalibrated.

In a chapter for medical readers, noted nystagmus researchers (33) pointedly questioned the use of nystagmus in sobriety testing:

Unfortunately, that alcohol can produce horizontal gaze-evoked nystagmus has led to a “roadside sobriety” test conducted by law-enforcement officers. Nystagmus as an indicator of alcohol intoxication is fraught with extraordinary pitfalls: many normal individuals have physiologic end-point nystagmus; small doses of tranquilizers that wouldn’t interfere with driving ability can also produce nystagmus; nystagmus may be congenital or consequent to structural neurologic disease; and often a neuro-ophthalmologist or sophisticated oculographer is required to determine whether nystagmus is pathologic. Such judgments are difficult for experts to make under the best conditions and impossible to make accurately under roadside conditions. It is unreasonable to have cursorily trained law officers using the test, no matter how intelligent, perceptive, and well meaning they might be. As noted, meticulous history taking and drug-screening blood studies are often essential in evaluating patients with nystagmus. (pp. 26–27)

There appear to be no easy answers to several of these objections, at least at the time of the traffic stop. Evaluation of the patient by an optometrist or ophthalmologist could uncover some of these problems, but only after the defendant has undergone the distress and inconvenience of arrest.

The importance of potentially confounding medical conditions for HGN assessment depends on three factors: (i) the base rate of the condition in the driving public, especially during late night hours, (ii) the percentage of persons with the condition that show eye symptoms, and (iii) whether the eye abnormalities produced closely resemble alcohol-caused nystagmus. Some potential causes are likely to be quite rare among drivers suspected of DWI (e.g., Brown’s syndrome). Others might be quite common, but produce symptoms in only a small percentage of those with the condition. Still others may rarely be mistaken for alcohol-induced nystagmus.

Nystagmus can be caused by a variety of things, some external such as visual motion or body rotation, others internal such as vestibular imbalance, neural damage, or chemical toxicity. Adherence to NHTSA’s administration standards will control for some possible causes, such as tilting the head (which tilts the semicircular canals) or a moving visual background. Various schemes have been proposed to enumerate distinct types of nystagmus (e.g., the 49 types of nystagmus cited by Dell Osso and Daroff; [33]). However, the vast majority can be described as visual-induced (optokinetic), vestibular (rotational, caloric), gaze-evoked (end-point), or idiopathic (infantile nystagmus). While some have used such lists to assert that there are 49 types of nystagmus other than alcohol-induced, this is not correct: Nystagmus found in HGN qualifies for several of the terms in the list (acquired, horizontal, jerk, associated). Lengthy lists mix descriptive and etiological terms, and some forms of nystagmus are unlikely to be mistaken for HGN. For example, lid nystagmus involves only the eyelids. However, normal end-point nystagmus, square wave jerks, infantile nystagmus, and optokinetic nystagmus could be mistaken for HGN under some circumstances, as might a nystagmus with a vestibular or congenital cause.

Perhaps more important, some medical conditions that damage the cerebellum or neural integrator may disrupt SP and cause gaze nystagmus that is indistinguishable from that caused by alcohol. For gaze nystagmus, there are three likely presentations depending on whether it is congenital or caused by imbalance in the neural integrator or in the vestibular system. Conditions that affect the neural integrator will present gaze nystagmus with the same waveform that is observed when alcohol is the cause: a decelerating drift from the original position of gaze followed by a saccade back to the original position (Fig. 3a). There is nothing for a police officer to discriminate, as the expected waveforms are identical. A congenital nystagmus or one caused by the vestibular system might be able to be distinguished from alcohol-induced nystagmus with training, but this has not yet been demonstrated in a peer-reviewed publication.
While the effects of alcohol on SP and gaze nystagmus are well-established, there is no direct empirical support for DSNMD, the most controversial of the HGN clues (7). Here, research is needed to establish whether it is related to BAC, and if so, whether it provides incremental validity over AON. Booker’s findings that it is sensitive to fatigue and recent alcohol use after BAC has reached 0.00% require replication, but cast a heavy shadow over this indicator.

HGN has been studied almost exclusively for predicting BAC, not behavioral impairment. While substantial correlations have been reported, no study has reported an average error of estimate of <0.03% (Standard Error of the Estimate = 0.0376; [162]), so that a 95% confidence interval spans a range of 0.147% BAC. Partly for this lack of precision, HGN is not admissible to prove a specific BAC in any jurisdiction. Rather, it is admissible to establish probable cause or as circumstantial evidence of behavioral impairment. But there are no studies demonstrating HGN’s validity for gauging physical, behavioral, or driving deficiencies. Only one study (163) reported correlations of HGN with other sobriety tests and number of computer-administered cognitive tasks. While moderate correlations were observed ($r = 0.30$–$0.60$), all were substantially less than HGN’s correlation with BAC ($r = 0.77$). Thus, there is very minimal evidence of HGN’s validity on the loss of mental ability, and none at all on loss of physical or driving ability.

It is unclear whether nystagmus is an indication of visual impairment. Regarding the significance of gaze nystagmus, the authors of one respected text stated that it “does not produce great functional disability since the eyes are used mostly near the central position” ([13], p. 254). Both for lateral gaze and during SP, the presence of nystagmus does not necessarily impair vision. A number of studies on congenital forms of nystagmus suggest that as long as 50 msec of saccade is obtained, visual acuity is maintained (164–168). Alcohol at moderate doses reduces the maximum velocity of saccades but does not affect acuity (169), although this may be because acuity testing is usually performed at primary gaze, not with the eyes deviated or following a moving target. Lastly, these effects have been indexed by BAC, not nystagmus. However, one study (170) reported that nystagmus (not further described) provided incremental validity over BAC in predicting deficits in visual search. This was true of groups above and below 0.08% BAC, as well as the total sample.

**NHTSA Training and Its Sufficiency**

There are numerous problems with the standard NHTSA training for HGN. Perhaps most notable, a total of perhaps 3 h classroom and demonstration time, out of a 3 day course, is devoted to HGN. Typically, this consists of viewing video tapes demonstrating alcohol-induced nystagmus and group examinations of subjects, most of whom are dosed with alcohol. The discussion of other types of nystagmus covers about one page, while the entire subject of medias suggests that are independent of base rate and are often quite substantial. Correct arrest rates, which are equivalent to positive predictive power, will always exceed the base rate (172). Thus, an 88% correct arrest rate in one NHTSA study (38) is less impressive when they are explicitly told that medical conditions that can cause nystagmus are rare among the driving population (1,2).

NHTSA publications often emphasize the overall accuracy rates or the “correct arrest” rates. Both take advantage of the high base rates of intoxication in many of the NHTSA study samples. NHTSA publications typically do not report the false-positive rates, which are independent of base rate and are often quite substantial. Correct arrest rates, which are equivalent to positive predictive power, will always exceed the base rate (172). Thus, an 88% correct arrest rate in one NHTSA study (38) is less impressive when they are explicitly told that medical conditions that can cause nystagmus are rare among the driving population (1,2).

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Highest correlation with BAC (blood alcohol concentration) of any established field sobriety test</td>
<td>1. Lacks specificity for alcohol; sensitive to a number of medical and psychiatric conditions and some prescription drugs</td>
</tr>
<tr>
<td>2. Most empirical support of any sobriety test</td>
<td>2. Many potential confounding factors not thoroughly investigated</td>
</tr>
<tr>
<td>3. Sensitive to low levels of BAC</td>
<td>3. Often performed incorrectly in the field</td>
</tr>
<tr>
<td>4. Appears to remain calibrated to BAC even in alcohol-tolerant drinkers</td>
<td>4. Not demonstrated to be related to behavioral or driving impairment</td>
</tr>
<tr>
<td>5. Easily administered, requires no special equipment</td>
<td>5. Environmental conditions at stop differ from laboratory testing, may reduce validity or create false positives</td>
</tr>
<tr>
<td>6. Requires little verbal instruction, could be administered to non-English speakers if nonverbal instructions developed</td>
<td>6. Ability of field officers to distinguish between alcohol-induced nystagmus and other eye movements not demonstrated</td>
</tr>
<tr>
<td>7. Lacks face validity, so juries give it less weight than psychomotor tests</td>
<td>8. In jurisdictions where considered a scientific test, may face admissibility hurdles</td>
</tr>
</tbody>
</table>

The initial error in SP tracking is not mentioned in NHTSA training materials. The initial, fast passes to check for equal tracking of the eyes may well produce nystagmus in sober people, and without explicit warning, police officers may be biased by observing it immediately before performing the assessment for breakdown of SP.

Breakdown of SP is also described inaccurately by NHTSA as being like a marble rolling on sandpaper (1,2), an analogy that suggests vertical, not horizontal, perturbations. Alcohol-induced nystagmus creates a specific waveform, whether occurring during SP or gaze: There is a slow, decelerating drift off the target followed immediately by a saccade back to the target. Officers are not taught to look for this specific waveform, either in their text or in video examples (1,2). Whether such training can teach officers to make such distinctions on commonly occurring, potentially confounding eye movements should be determined immediately, and if successful, incorporated into NHTSA training.

The implicit emphasis of NHTSA training is on detecting intoxication, and police officers may never evaluate a substantial number of persons who have not been drinking before being assigned to traffic duty. An alternative model might define the purpose of testing as to distinguish those who are legally intoxicated from those that are not. Police officers who do not test many sober people may never come to question NHTSA’s inaccurate description of normal SP and may regard any jerking movement of the eye as alcohol-induced nystagmus. This risk is increased when they are explicitly told that medical conditions that can cause nystagmus are rare among the driving population (1,2).

**TABLE 5—Advantages and disadvantages of HGN as a field sobriety test.**
daytime or sobriety checkpoint stops, false-positive errors may greatly outnumber valid arrests.

Strengths and Limitations

While HGN has many problems, it is not without its virtues. Table 5 lists its advantages and disadvantages relative to other non-chemical field sobriety tests. The only real rival to HGN, when BAC is the criterion, is AON. It has shown higher levels of correlation and accuracy of classification in several studies, although the optimal formula has varied somewhat across studies. Although there is less empirical support for AON than HGN, it has several advantages. Unlike HGN, it can provide a continuous scale of test performance that is linearly related to BAC. While HGN is likely to “top out” at BAC levels of 0.08–0.12%, AON does not. An AON of 20° could provide stronger evidence of intoxication than an HGN score of 6. Secondly, because GN relies on fewer brain structures and the variables that affect it than for SP.

Conclusion

While the sobriety testing literature provides circumstantial evidence of HGN’s validity when BAC is used as a criterion, the eye movement literature raises serious questions about its use as a road-side sobriety test. Primary among these is how many sober subjects’ SP at 30°/sec is free of observable saccades. Roadside testing entails several factors that probably disadvantage the SP system, including the presence of a meaningful background (e.g., the officer’s face), effects of fatigue, fear, or circadian rhythms, unpredictable stimulus speed, and testing at eccentric view angles. There is very minimal data to support the validity of HGN for gauging mental, physical, visual, or driving impairment. The clue of breakdown of SP is significantly biased against older people, and the age at which this disadvantage becomes significant has not been established. HGN appears potentially vulnerable to false-positive errors from a number of medical and situational conditions as well as prescription drugs. NHTSA does not adequately address these and other issues in its training materials and curriculum. Lastly, peer-reviewed research needs to address many of the points made in this article, including the effects of anxiety and fatigue on HGN, and the relationship between HGN, alcohol, and driving performance.

References

33. Dell’Osso LF, Daroff RB. Nystagmus and saccadic intrusions and oscillations. In: Tasman W, Jaeger EA, editors. Duane’s clinical
14 JOURNAL OF FORENSIC SCIENCES


Additional information and reprint requests:
Steve Rubenzer, Ph.D., A.B.P.P.
11914 Astoria, Suite 490
Houston, TX 77089
E-mail: rubenzer.steve@att.net

16 JOURNAL OF FORENSIC SCIENCES