

Scientific and technical

The Horizontal Gaze Nystagmus test: fraudulent science in the American courts

JL Booker
Central Texas Analytical Consultants, PO Box 509, Eday, TX 78524-0509, USA
Science & Justice 2004 44 133 - 139 Original paper submitted 18 June 2002 revised paper accepted 18 March 2004

The Horizontal Gaze Nystagmus (HGN) test was conceived, developed and promulgated as a simple procedure for the determination of the blood alcohol concentration of drivers suspected of driving while intoxicated (DWI). Bypassing the usual scientific review process and touted through the good offices of the federal agency responsible for traffic safety, it was rushed into use as a law enforcement procedure, and was soon adopted and protected from scientific criticism by courts throughout the United States. In fact, research findings, training manuals and other relevant documents were often held as secrets by the state. Still, the protective certification of its practitioners and the immunity afforded by judicial notice failed to silence all the critics of this deeply flawed procedure. Responding to criticism, the sponsors of the test traveled the path documented in this paper that led from mere (if that word can ever truly apply to a matter of such gravity) carelessness in research through self-serving puffery and finally into deliberate fraud - always at the expense of the citizen accused.

Le test du Nystagmus du regard horizontal (HGN) a été conçu, développé et promu comme une procédure simple pour déterminer la concentration d'alcool dans le sang de conducteurs suspects de conduire en état intoxiqués (DWI). En passant à côté du processus de revue scientifique habituel, et introduit au travers des bons offices de l'agence fédérale responsable pour la sécurité routière, ce test a rapidement été introduit comme une procédure judiciaire et a été rapidement adoptée et protégée de toute critique scientifique par les tribunaux à travers les Etats-Unis. En fait, les résultats de recherche, les manuels de formation et d'autres documents pertinents ont souvent été tenus secrets par l'Etat. Malgré tout, la certification protectrice de ses praticiens et l'immunité donnée par la notice judiciaire n'ont pas suffi à faire taire toutes les critiques de cette procédure profondément défectueuse. Afin de répondre à la critique, les promoteurs du test ont suivi les chemins documentés dans cet article qui ont conduit d'une simple négligence (si ce mot peut vraiment s'appliquer à une affaire d'une telle gravité) basée sur du vent, et finalement à une fraude délibérée - toujours au détriment du citoyen accusé.

Editorial Comment: The Editor is particularly grateful for the helpful comments made by the several referees with scientific, legal and medical backgrounds who have reviewed this paper.

Der 'Horizontal Gaze Nystagmus (HGN)-Test' wurde erdnen, entwickelt und bekannt gemacht als einfache Methode zur Bestimmung der Blutalkoholkonzentration bei Personen, die verdächtigt wurden, ein Kraftfahrzeug im betrunkenen Zustand zu führen (DWI). Unter Umgehung der üblichen wissenschaftlichen Prüfung und durch Werbung, die mit tatkräftiger Unterstützung der für die Verkehrssicherheit zuständigen Bundesbehörde erfolgte, wurde diese Methode zügig als Strafverfolgungsmittel eingeführt und bald von Gerichten in den gesamten Vereinigten Staaten übernommen und vor wissenschaftlicher Kritik in Schutz genommen. De facto wurden Forschungsergebnisse, Übungsanleitungen und andere relevante Dokumente durch den Staat oftmals geheim gehalten. Doch weder die schützende Zulassung der Anwender noch die per Gerichtsbescheid gewährleistete Immunität konnten die Kritik an dieser zuleist fehlerhaften Prozedur zum Schweigen bringen. Als Reaktion auf diese Kritik beschritten die Sponsoren des Tests den Weg, der in dieser Arbeit beschrieben ist. Dieser führte von bloßer (wenn dieses Wort überhaupt auf eine Sache von demartig gravierender Tragweite angewendet werden darf) Nachlässigkeit in der Forschung über eigenmächtige Marktbesetzung letztlich hin zu vorätzlichem Betrug - immer auf Kosten der angeschuldigten Bürger.

El test de nistagmo de mirada horizontal (HGN) se concibió, desarrolló y promulgó como un simple procedimiento para la determinación de alcohol en sangre de conductores sospechosos de conducir en estado de embriaguez. Una vez pasado el correspondiente proceso de revisión científica y dirigido a través de los buenos offices de la agencia federal responsable de la seguridad del tráfico, fue urgido su aplicación como procedimiento de aplicación de la Ley y pronto fue adoptado y protegido por los tribunales de los estados Unidos. De hecho, los hallazgos de investigación, los manuales de entrenamiento y otros documentos relevantes eran con frecuencia mantenidos en secreto. Sin embargo, la certificación protectora de los que lo practicaban y la inmunidad judicial permitida no consiguieron silenciar todas las críticas de este procedimiento con profundos fallos. En respuesta a las críticas, los partidarios del test siguieron el camino documentado en este trabajo que lleva del merec si es que esta palabra pueda aplicarse realmente a un asunto de tanta gravedad) descuido en la investigación a través de un desoservido engaño y finalmente a un fraude deliberado siempre a expensas del ciudadano acusado.

© The Forensic Science Society 2004
Key words Forensic science, field sobriety tests, DWI, driving, vehicle, intoxication, Horizontal Gaze Nystagmus, HGN.

journal name & citation

J. Booker
The Horizontal Gaze Nystagmus test: Molecular science in the American courts

Introduction

The Horizontal Gaze Nystagmus (HGN) test is the most widely applied toxicological procedure in the United States of America. Developed between 1974 and 1983 in a federally-funded program sponsored by the National Highway Transportation Safety Administration (NHTSA), an agency created by the Department of Transportation to enact and enforce the 1966 National Traffic and Motor Vehicle Safety Act, the HGN test flew directly in the face of established medical and toxicological opinion that simple clinical tests were inadequate for the task of establishing intoxication. In fact, when the test was developed, the American Medical Association (AMA) had just published its comprehensive review of the subject of 'Alcohol and the Impaired Driver' (which, interestingly, was not cited in the publications touting the HGN test) that contained the statement, 'Conscientious physicians recognized that each of the accepted signs or symptoms of intoxication might well be a symptom of some disease or condition of the accused entirely unrelated to the ingestion of alcohol. Indeed, Herman A. Heise, a former Chairman of the AMA Committee on Medicolegal Problems, and an outstanding authority on the subject of determining alcohol intoxication, has compiled a list of over 100 pathological conditions which produce one or more of the symptoms generally accepted as signs of intoxication. Only by careful differential diagnosis could intoxication be distinguished from these illnesses.' [1]

Although its merit as a scientific test might have been disputed, its value as a political matter is not in doubt: important voting constituencies that supported the American president elected in 1980 were the 'religious right', a fundamentalist coalition of religious groups that eschew the use of alcohol, the 'moral majority', a political group that also espoused prohibition, and the soon-to-be-important Mothers Against Drunk Driving (MADD) which still advocates similar views. Supplemented by epidemiological studies that resulted in a recommendation by the AMA for acceptance of 0.10 w/v% as the limit of intoxication [2] and corroborative opinions by most toxicologists based on the Grand Rapids Study [3], impaired drivers were perceived as a national problem to be dealt with - preferably sooner rather than later. In that context, the HGN was rushed into service without ever seeing peer-review and critique, and for the last two decades it has been taught to police officers as one of three tests described in variants of a federally-produced manual titled, *DWI Detection and Standardized Field Sobriety Testing Student Manual*.

In 1996 the fundamental competence of the scientific basis of the standardized honory of field sobriety tests was challenged with the exposure of an internal paradox that implied the work was poorly researched and inadequately reviewed prior to its adoption by the National Highway Transportation Safety Administration [4]. In 1998 the integrity of the statistical evaluation of the original research upon which the validity of the tests rested was unfavorably reviewed [5]. In 2001 new research indicated that the Horizontal Gaze Nystagmus (HGN), the cornerstone of the test battery was fundamentally flawed and that the HGN test was improperly conducted by more than 95% of the police officers who used it to examine drivers suspected of driving while

intoxicated (DWI) [6]. This summary critique demonstrates that it is scientifically meritless and that the United States Department of Transportation indulged in deliberate fraud in order to mislead the law enforcement and legal communities into believing the test was scientifically meritorious and overvaluing its worth in the context of criminal evidence.

Because of peer-review and the universal expectation that experimental procedures and data will be produced to substantiate research claims, true fraud in scientific matters is extremely rare [7,8,9]. Its early stages are often indistinguishable from unpopular, mistaken, or simply eccentric science [10], so it is essentially invisible within the body of scientific research until it is exposed through a process of unfavorable review and the inability of other scientists to duplicate experimental results or independently confirm the reported findings. In the context of forensic science, it 'is impelled through our courts by a mix of opportunity and incentive' [11], the opportunity arising from the adversarial nature of the proceedings and the lack of scientific knowledge among the legal participants and the incentive usually being a rational motive such as greed, fame, power, or some other careerist goal [12, 13].

The HGN test, which is accepted by the courts as evidence equal in merit to chemical testing and protected from review and criticism by judicial notice [14], avoided this review because the bulk of the publications describing and reviewing the procedure are found outside peer-reviewed literature and because the test is taught, certified and used as an enforcement function of police agencies, not as a scientific function. In the State of Texas for example, breath alcohol test results must be accompanied by the testimony of a degreee scientist with specialized training in the instrumentation and protocols of the program and the effects of alcohol [15], but HGN test results may be introduced through a police officer who has a certificate showing completion of a short course of instruction taught by non-degreee instructional personnel.

The Elements of Fraudulent Science

The accusation of intentional fraud is so serious, it seems reasonable that at least three elements should be demonstrated before a scientific work can be declared to be deliberately fraudulent:

- In the underlying research there must be a considerable lack of competence in the form of significant errors, internally inconsistent test results, misrepresentation of experimental results or redaction of unfavorable data during its development.
- Unfavorable review and criticism must elicit significant efforts to hide evidence of flawed research, especially efforts such as those recognized and defined as 'pathological science' by Langmuir [16] or as 'pseudo-science' by Iastrucci [17].
- The advocates of the work must promote it using deliberate deception and misrepresentation of fact intended to deceive and mislead the non-scientifically trained audience.

JL Bookow

The Horizontal Gaze Nystagmus test: fraudulent science in the American courts

Incompetence

In the matter of the field sobriety tests, particularly the HGN test, lack of fundamental competence can be illustrated by the procedure used for the measurement of the 45° angle of deviation of vision; reduction of unfavorable data during its development is demonstrated by development of the vertical gaze nystagmus (VGN) test.

Measuring an Angle of 45°

The signal element of the HGN test is the criterion of discernible nystagmus appearing before the eyes are deviated 45° laterally. During the development of the HGN test, police officers were expected to estimate the lateral angle of nystagmus onset with an accuracy of plus or minus 3° on the basis of having been trained using a protractor appliance [16], but with the early, widespread rejection of the test as a method for determining an exact blood alcohol concentration based on the angle of nystagmus onset, the training procedure was modified slightly so that the simpler criterion of having officers estimate a 45° angle of deviation after being trained with a simple template replaced the more demanding task of estimating the angle accurately [19].

In the mid-1990s training in the estimation of the critical 45° angle was changed with the adoption of a procedure in which the officer simply moved the stylus on which the subject's eyes were focused to a position directly in front of the tip of the subject's shoulder as an indicator of 45°. After 1999 the written training materials describe in detail this testing protocol in which the point of the shoulder serves as a reference point for determining a 45° angle and only briefly mention the template measurement training method [20,21].

The officer's first instruction for conducting the test, 'Position the stimulus approximately 12-15 inches from the subject's nose and slightly above eye level,' has remained unchanged since the inception of the test. The distance from eye to tip of the shoulder is approximately eight inches (from a low of approximately seven inches for small framed men and many women to as much as nine inches for large framed men), thus the deviation angle is never 45° but an angle of approximately 30° with a range from 25° to 37°, therefore the modification of the test protocol to incorporate the use of the tip-of-the-shoulder procedure is certain to produce inaccurate estimates of a 45° angle.

Although the estimate of the 77% reliability of the HGN test as an indicator of a BAC greater or less than 0.10% was developed for a procedure using a true 45° angle, that estimate of reliability is still cited as applying to tests conducted using the shoulder tip criterion. Error introduced by the erroneous measurement of the critical 45° angle is offered as a positive feature in the testing protocol, reasoning that the officer is looking for onset of nystagmus prior to 45°, and because all the angles included in the arc between 25° and 37° appear prior to a 45° deviation of the eyes - this inaccurate measurement actually constitutes a margin of safety for the defendant, assuring that he will benefit from the error.

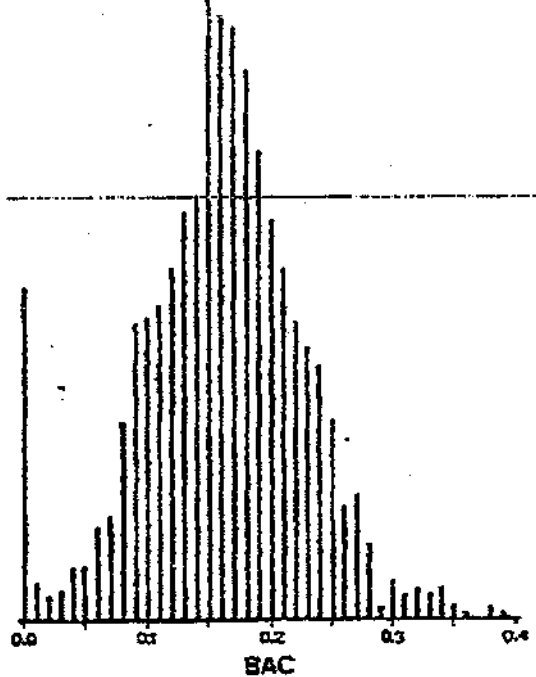
Even if there were a place in science for situative incompetence,

that reasoning highlights the fact that arrest results are demonstrably inconsistent with the original research findings: Good *et al.* [22] summarized the results of the 1981 report by Tharp *et al.* [23], stating that it had shown BAC estimates to be accurate to ±0.02% (Tharp *et al.* actually reported ±0.03%) using the formula:

Angle of onset of gaze nystagmus = 51° - 105° x percent BAC,

apparently created as a composite of four equations generated by Tharp *et al.* Using this formula - or, for that matter, any of the four formulae found in the original report - an onset angle of 30° corresponds to a BAC value of approximately 0.20% (25° corresponds to a BAC value of approximately 0.25%, and 37° corresponds to a BAC value of approximately 0.14%). If, therefore, the original research is to be believed and tests are administered in the field according to the current methodology, the critical 'onset prior to 45°' clues should rarely appear among test subjects with BAC values less than 0.20% and never for subjects below 0.10%. (Arrests made in the absence of a nystagmus onset less than 45° - either accurately or inaccurately measured - are so rare that they are effectively non-existent.) Arrest data such as those shown in Figure 1 show that for more than 2000 arrests in Texas approximately 17% of the arrests were of subjects with BAC values of from 0.00% to 0.09% and that approximately 63% of the arrests were of subjects with BAC

Figure 1 Blood alcohol concentrations of 2,068 Texas drivers arrested for DWI (1998-2002).



JL Boelter

The Horizontal Gaze Nystagmus test: fraudulent science in the American courts

values of from 0.10% to 0.19%. A similar compilation of arrest data published in the 2000 version of the NHTSA Instructor's Manual [24] shows that 31% of the arrests were of subjects with BAC values of from 0.00% to 0.09%, and that approximately 49% of the arrests were of subjects with BAC values of from 0.10% to 0.19%. Having 80% of the field data fall outside the range of error predicted from controlled laboratory studies exemplifies one of the characteristics ('unsystematic in the respect that its various parts do not necessarily relate to and support each other by virtue of internal consistency.') of pseudoscience as posited by Lastrucci.

The Vertical Gaze Nystagmus test

The origin of the VGN test is somewhat obscure. The 1992 Instructor's manual says, '[V]ertical nystagmus was not included in SEST battery during the original research, however, it is a reliable indicator of high alcohol for that individual, or PCP impairment' [25]. The only reference to any PCP literature found in the NHTSA research documents is a paper by Liden *et al.* [26] that unambiguously shows vertical nystagmus to have been present in only four of the nine cases of PCP poisoning discussed, explaining the reason for the phrase, 'selected sections' in the NHTSA citation of that work. It also explains why the student manual teaches the practitioner a specific procedure to use when administering the VGN test, but does not discuss an evaluation of the test apart from the vague, pejorative reference to 'high doses of PCP' or how the authors of the method arrived at the conclusion that 'Vertical nystagmus usually occurs with PCP' [27].

Attempts to hide flawed research

The second element of fraudulent science, an attempt to hide evidence of flawed research, can best be illustrated using Langmuir's criteria of 'pathological science.' Of the six criteria Langmuir suggested, five demonstrably apply to the HGN test:

1. *Fantastic theories contrary to experience are suggested.* The HGN test is based upon the assumption that any physical or mental impairments that serve to produce 'clues' from which intoxication may be inferred are caused by alcohol. There is no requirement that any evidence, much less substantive evidence, of the presence of alcohol is a predicate condition for advancing to a conclusion regarding the amount of alcohol that is present. The original developers explained away some of the false positive results obtained during their developmental studies as possibly being due to 'drugs other than alcohol, the result of brain damage, of illness, or of unknown etiology' [28], but by the time the training manual was published, each test, individually, and in combination was evaluated with respect to alcohol being the sole source of debilitation.

2. *There are claims of great accuracy.* After 1983, NHTSA's estimates of the accuracy of the HGN test with respect to a BAC of 0.10% were based on the finding of four clues. Several reasons existed to doubt the claimed 77% accuracy of the HGN test (80% when combined with walk-and-turn 'clues'), the most notable being a peculiar internal paradox [4], the paucity of published experimental data, and lack of any explanation of the

developmental methodology other than 'a table was developed.' The promoters of the HGN test have recently encouraged prosecutors to believe that 'experienced law enforcement officers were correct ninety-six percent of the time in determining a .10 BAC or more using the HGN test' [29], a finding that defies belief considering that under laboratory conditions using specially trained, experienced observers and a protractor to measure the angle the claimed accuracy of the HGN test was only 88%, and less than that (78%) when conducted without that appliance [30]. Considering that tests conducted at roadside seldom, if ever, conform to the minimum standards of the written protocol, that from 17-31% of arrested subjects produced chemical tests less than 0.10%, and that drugs other than alcohol, brain damage, illness, or debilitation of unknown etiology exist in the field as well as in the laboratory, the claimed accuracy of 96% is unbelievable.

3. *The maximum effect that is observed is produced by a causative agent of barely detectable intensity, and the magnitude of the effect is substantially independent of the cause.* Goldberg established a very important principle in testing for the effects of alcohol. 'If the test is to be applied for practical purposes, and no basal values are available, the variability between individuals should be slight as compared with the departures from normal due to alcohol.' Given that half the population of sober adults exhibit end position nystagmus and that all the possible clues associated with the HGN test, especially the critical 'onset prior to 45" manifest at blood alcohol concentrations half (or less than half) that for which the test was developed, it is apparent that the HGN test is subject to a significant number of false positive results.

4. *The effect is of a magnitude that renders close to the limit of detectability or, many measurements are necessary because of the very low statistical significance of the results.* Because people with BACs greater than 0.15% generally exhibit obvious signs of impairment [31] and can be identified without the use of the field sobriety tests, it serves a useful purpose only for discriminating among suspects with BACs less than 0.15%, a fact acknowledged in the developmental work by Sharp *et al.* [32] but ignored during subsequent validation studies. Based on arrests and chemical test data (Figure 1), the field sobriety tests had an error rate of approximately 40% among people whose BACs were between 0.00% and 0.14%. Data published by NHTSA [24] indicate an even higher rate of error (ca. 53%) among this group. This high rate of undetectable error is unavoidable [5,33].

5. *Criticisms are met by ad hoc excuses thought up on the spur of the moment.* The intended use of the field sobriety tests has never been in doubt. With almost Biblical solemnity the authors of the NHTSA student manual stated that they were to be used 'only when the tests are administered in the prescribed, standardized manner, and only when the standardized clues are used to assess the subject's performance; and, only when the standardized criteria are employed to assess that performance.' They also stated that the HGN test alone could be used to correctly estimate whether a person's BAC was above or below

JL Gasker
The Horizontal Gaze Nystagmus test: fraudulent science in the American courts

values of from 0.10% to 0.19%. A similar compilation of arrest data published in the 2000 version of the NHTSA Instructor's Manual [24] shows that 31% of the arrests were of subjects with BAC values of from 0.00% to 0.09% and that approximately 49% of the arrests were of subjects with BAC values of from 0.10% to 0.19%. Having 80% of the field data fall outside the range of error predicted from controlled laboratory studies exemplifies one of the characteristics ('unsystematic in the respect that its various parts do not necessarily relate to and support each other by virtue of internal consistency') of pseudoscience as posited by Lastrucci.

The Vertical Gaze Nystagmus test

The origin of the VGN test is somewhat obscure. The 1992 Instructor's manual says, '[V]ertical nystagmus was not included in SFST battery during the original research, however, it is a reliable indicator of high alcohol for that individual, or PCP impairment' [25]. The only reference to any PCP literature found in the NHTSA research documents is a paper by Lidon et al. [26] that unambiguously shows vertical nystagmus to have been present in only four of the nine cases of PCP poisoning discussed, explaining the reason for the phrase, 'selected sections' in the NHTSA citation of that work. It also explains why the student manual teaches the practitioner a specific procedure to use when administering the VGN test, but does not discuss an evaluation of the test apart from the vague, pejorative reference to 'high doses of PCP' or how the authors of the method arrived at the conclusion that 'Vertical nystagmus usually occurs with PCP' [27].

Attempts to hide flawed research

The second element of fraudulent science, an attempt to hide evidence of flawed research, can best be illustrated using Langmuir's criteria of 'pathological science.' Of the six criteria Langmuir suggested, five demonstrably apply to the HGN test:

1. *Fanciful theories contrary to experience are suggested.* The HGN test is based upon the assumption that any physical or mental impairments that serve to produce 'clues' from which intoxication may be inferred are caused by alcohol. There is no requirement that any evidence, much less substantive evidence, of the presence of alcohol is a predicate condition for advancing to a conclusion regarding the amount of alcohol that is present. The original developers explained away some of the false positive results obtained during their developmental studies as possibly being due to 'drugs other than alcohol, the result of brain damage, of illness, or of unknown etiology' [28], but by the time the training manual was published, each test, individually, and in combination was evaluated with respect to alcohol being the sole source of debilitation.

2. *There are claims of great accuracy.* After 1983, NHTSA's estimates of the accuracy of the HGN test with respect to a BAC of 0.10% were based on the finding of four clues. Several reasons existed to doubt the claimed 77% accuracy of the HGN test (80% when combined with walk-and-turn 'clues'), the most notable being a peculiar internal paradox [4], the paucity of published experimental data, and lack of any explanation of the

developmental methodology other than 'a table was developed.' The promoters of the HGN test have recently encouraged prosecutors to believe that 'experienced law enforcement officers were correct ninety-six percent of the time in determining a .10 BAC or more using the HGN test' [29], a finding that defies belief considering that under laboratory conditions using specially trained, experienced observers and a promotor to measure the angle the claimed accuracy of the HGN test was only 88%, and less than that (78%) when conducted without that appliance [30]. Considering that tests conducted at roadside seldom, if ever, conform to the minimum standards of the written protocol, that from 17-31% of arrested subjects produced chemical tests less than 0.10%, and that drugs other than alcohol, brain damage, illness, or debilitation of unknown etiology exist in the field as well as in the laboratory, the claimed accuracy of 96% is unbelievable.

3. *The maximum effect that is observed is produced by a cumulative agent of barely detectable intensity, and the magnitude of the effects is substantially independent of the cause.* Goldberg established a very important principle in testing for the effects of alcohol, 'If the test is to be applied for practical purposes, and no basal values are available, the variability between individuals should be slight as compared with the departures from normal due to alcohol.' Given that half the population of sober adults exhibit end position nystagmus and that all the possible clues associated with the HGN test, especially the critical 'onset prior to 45°' manifest at blood alcohol concentrations half (or less than half) that for which the test was developed, it is apparent that the HGN test is subject to a significant number of false positive results.

4. *The effect is of a magnitude that remains close to the limit of detectability or, many measurements are necessary because of the very low statistical significance of the results.* Because people with BACs greater than 0.15% generally exhibit obvious signs of impairment [31] and can be identified without the use of the field sobriety tests, it serves a useful purpose only for discriminating among suspects with BACs less than 0.15%, a fact acknowledged in the developmental work by Theop et al. [32] but ignored during subsequent validation studies. Based on arrests and chemical test data (Figure 1), the field sobriety test had an error rate of approximately 40% among people whose BACs were between 0.00% and 0.14%. Data published by NHTSA [24] indicate an even higher rate of error (ca. 53%) among this group. This high rate of undetectable error is unavoidable [5,33].

5. *Critikisms are met by ad hoc excuses thought up on the spur of the moment.* The intended use of the field sobriety tests has never been in doubt. With almost Biblical solemnity the authors of the NHTSA student manual stated that they were to be used 'only when the tests are administered in the prescribed, standardized manner, and only when the standardized clues are used to assess the subject's performance; and, only when the standardized criteria are employed to assess that performance.' They also stated that the HGN test alone could be used to correctly estimate whether a person's BAC was above or below

0.10% in 77% of the subjects tested. Without question, the HGN test was promoted and accepted for stand-alone use for the purpose of discriminating between BACs greater and less than 0.10%. Recently, however, prosecutors have been advised [34], 'Although the HGN test is an excellent indicator of impairment, the test results alone are not used to convince a jury that a defendant was impaired. Combined with other evidence of impairment, such as erratic driving, odor of alcoholic beverage, glassy or blood-shot eyes or unsatisfactory performance on other SFSTs, HGN is strong evidence of impairment.'

There are several compelling reasons for impeaching this argument:

a. 'Impairment' (also called 'loss of normal use of mental and physical faculties', 'not having normal use of mental and physical faculties', and 'under the influence') describes a simple evaluation based on (1) the tests performed on the driver at the time of arrest, (2) an assumption of the presence of alcohol, and (3) an imagined response to the tests in the absence of alcohol, a concept that necessarily ignores other possible causes of apparent impairment such as fatigue, debilitation, diabetes or other illness, and individual variations in the ability to perform various tests.

HGN has never been proffered as an indicator of impairment, much less 'an excellent indicator.' At best, it has been promulgated for the entirely different purpose of indicating the presence and chemical concentration of alcohol in a person's blood. It is statistically correlated with a blood alcohol concentration. In fact, the officers are taught that when illegal per se laws were adopted they would not be required to 'establish that the defendant was "under the influence" but that it would be 'sufficient to show that the defendant's BAC was 0.10 or more' [35]. In jurisdictions such as Texas where police officers receive proficiency certificates and are therefore licensed to offer testimony in courts, it is because they have produced documented correlations of actual arrests and chemical concentration tests: they are never required to address the issue of impairment in order to be certified.

b. Alcohol on human breath is undetectable by odour, and the odour of alcoholic beverage is a notoriously poor indicator of the presence of alcohol [36].

c. If it is dependent on the presence or absence of glassy or blood-shot eyes, one must assume the quality of the HGN test as evidence changes seasonally with the air's pollen and mold spore content, but this issue has never been addressed in any of the HGN literature.

d. 'Unsatisfactory performance on other SFSTs' is directly addressed by Anderson *et al.* [37] and in all versions of the student and instructor training manuals: All other field tests are irrelevant if the HGN examination reveals four or more clues and, with the abandoning of the 'decision table,' there is no procedure for addressing composite performance factors.

e. Citing erratic driving as a confusing observation for the HGN test is a remarkable example of circular reasoning because the test was originally designed to discriminate between alcohol-induced intoxication and 'driving errors [which] may be attributable to impairment other than alcohol intoxication, such as a woman who has just had her purse stolen and is too upset to concentrate on driving; a diabetic person in need of insulin; a married couple arguing; an elderly man driving too carefully, etc.' [38].

The Intention to Deceive

Resources Favorable to the HGN Test

Because of its gravity, the accusation of fraud must rest upon unambiguous, unimpeachable evidence. That evidence is 'Appendix F' in *Horizontal Gaze Nystagmus: The Science and the Law: A Resource Guide for Judges, Prosecutors and Law Enforcement* [39], a publication directed toward lay readers. 'Appendix F' has been attached to documents submitted to United States Federal Court to demonstrate that the HGN test meets the Daubert standard [40] and establish the merit of the battery of standardized field tests in criminal prosecutions.

Appendix F contains a compilation of publications headed, 'Publications favourable to HGN.' Within this is a list headed, 'Other research studies and articles,' that cites 60 publications, 22 of which were cited as references relied upon in the development of the standardized battery of tests and, therefore, can hardly be said to constitute a favorable review of the battery or of any of its component tests. Although it constitutes 'bolstering' in the legal sense, it does not constitute fraud in the scientific sense.

The elements of fraud are found in the content of several of the publications dating well after the field sobriety tests were promulgated for general use. Papers by Averbuch-Heller [41], Burton [42], Hwang [43], are particularly instructive in assessing whether the list of publications in Appendix F are favourable to the HGN test. These publications share common features:

a. None contains any reference to the HGN test or to any of its component elements (breakdowns of smooth tracking, end-position nystagmus, or appearance of nystagmus prior to deviation of gaze 45° from straight ahead).

b. None of the authors used the HGN test or any identifiable procedure that resembles the HGN test or its component elements as a diagnostic procedure or an examination technique.

c. None contains any references to publications that describe or review the HGN test or any of its component elements.

d. None contains any references to publications authored by any of the principal authors of the NHTSA publications, Southern California Research Institute (SCRI) publications (Burns and Moskowitz and Therp *et al.*) or other publications describing or commenting upon the HGN test.

e. None contains any reference to nystagmus being a symptom

JL Bookler

The Horizontal Gaze Nystagmus test: fraudulent science in the American courts

of ethanol intoxication or consumption. One of the papers (Hwang) states the serum alcohol concentration of the person about whose medical condition it was written was zero [45].

Discussion

In approximately half of DWI arrests, videotapes of the officer and suspect are made at the time of the traffic stop. These tapes are often either exculpatory or of such poor quality that they have no probative value, and chemical tests performed at a later time are not acceptable *per se* as evidence of intoxication while driving [44]. Since arrests are rarely made on the basis of egregious impairment, the arresting officer's essentially unrefutable opinion of intoxication based on the field sobriety tests is the only evidence available to the state that the suspect was intoxicated when stopped. Admissible lay testimony and undocumented, therefore unimpeachable observations are matters for the courts, but the underlying worth of the field sobriety tests falls within the purview of scientists.

The state's argument for the field sobriety tests does not rest on proof of merit, but upon *qui tacet consentit* reasoning – that these tests have been so widely accepted they must have been subjected to some kind of review prior to adoption in the many jurisdictions where they are used, that somewhere along the way someone would have spotted the flaws and shortcomings. Considering that the student manual was originally considered to be a confidential state document and was only obtained through an 'Open Records Act' request, silence from the scientific community cannot be considered an endorsement of the program. Moreover, certification by the state is limited to law enforcement personnel.

Even if the field sobriety tests failed to attract the attention of scientists, numerous capricious, unexplainable changes from the original protocol suggest that the agencies sponsoring and enforcing the use of the tests failed to exercise due diligence in evaluating them; for example,

1. Thorp *et al.* produced a test-evaluation procedure that relied on finding both nystagmus and balance clues [45]. The procedure was abandoned by Anderson *et al.* without explanation in 1983 [37].

2. Anderson *et al.* also produced a patient recognition decision table that combined walk-and-turn clues and nystagmus clues to determine the probability of intoxication [37]. This procedure was the cornerstone of the tests until it was abandoned without explanation in 1999.

3. Anderson *et al.* reviewed earlier laboratory results and produced estimates of the reliability of the individual tests [46]. The training materials discarded the fact that these values were estimates and students were taught that they should expect the same results in field use [47].

4. Contradictory changes in the test procedures were incorporated without explanation.

a. In the report issued by Thorp *et al.* a caveat accompanied the walk-and-turn test noting that it might not be suitable for subjects over the age of 65 [48]; in 1992, that suitability profile was changed to 60 years of age and more than 50 pounds overweight [19]; in 2000 the original parameters were re-adopted [20].

b. For the one-leg stand test, the 1983 Thorp *et al.* caveat was for subjects over the age of 65 or more than 50 pounds overweight [48]; in 1992, the age changed to 60 years of age and the caveat about subjects more than 50 pounds overweight remained [19]; in 2000 the age returned to 65 years and the weight warning was dropped [20]. (Paradoxically, each of the altered statements was prefaced, 'The original research indicated,' which appears to be a reference to age 60 and no weight consideration [49].)

5. The physical tests parameters have also changed. For example, earlier versions of the training manuals (Instructor and student) insisted that the walk-and-turn test required a 'clearly visible' line [19]; the latest version omits this requirement, so most subjects being tested in the field are now walking (and occasionally stepping off) invisible, imaginary lines.

6. The original documents were quite specific in matters of procedure and conditions, but the current publications say that the effect of individual officers changing these parameters should be left to a lay jury to decide ('may have some effect on the evidentiary weight given to the results') [20] although they provide no information to guide the evaluation process.

The substance of fraud is demonstrable; the motive is speculative. In this instance, one other salient fact may provide a clue addressing motive: Thorp *et al.* cite a 1974 NHTSA publication stating that the drivers arrested for DWI had an average blood alcohol concentration of 0.17%, and stated that the goal of the program in which the HGN test was developed was directed toward seeing that value lowered [50]. Two decades later that goal has not been achieved despite the money, time and effort invested by the federal and many local governments (Figure 1).

References

1. Committee on Medicolegal Problems of the American Medical Association. *Alcohol and The Impaired Driver, A Manual on the Medicolegal Aspects of Chemical Tests for Intoxication*. Chicago: American Medical Association, 1968. (revised and reissued in 1971).
2. The House of Delegates of the American Medical Association. *Policy Statement*. Chicago: American Medical Association, November 30, 1950.
3. Borkshire JL, Crowther H, Shurtain R, Zel W and Jernin R. *The Flow of The Drinking Driver in Traffic Accidents*. Indiana: Indiana University Press, 1964.
4. Bookler JL. The Field Test Paradox. *Voice for the Deafened* 1986; 25(2): 9-10.
5. Bookler JL. The Application of the 'Known and Proven Rate of Error' Criterion to the Standardized Battery of Field Sobriety Tests. *Voice for the Deafened* 1988; 27(2): 24-27.
6. Bookler JL. *Impairment Nystagmus: An Indicator of Ethanol Intoxication*. *Science & Justice* 2001; 41: 114-116.
7. Kaufman D. *Field vs. Evidence*. *Science* 1987; 236: 141.