Published in final edited form as:

J Pediatr Ophthalmol Strabismus. 2015 ; 52(2): 114–118. doi:10.3928/01913913-20150114-04.

Reliability and Application Variability of a Commercially Available Infrared Videonystagmography Unit

A Gerling¹, S Leu, PhD², A Morton³, J Eschegoyen, MD³, A Nalbandian, PhD⁴, and R Lingua, MD^{5,#}

¹Gavin Herbert Eye Institute, University of California-Irvine, Irvine, CA 92697

²Biostatistics, Epidemiology & Research Design Unit, University of California-Irvine Institute for Clinical and Translational Science, Irvine, CA 92697

³University of California-Irvine, School of Medicine, Irvine, CA 92697

⁴Discovery Foundation, University of California-Irvine, Irvine, CA 92697

⁵Director, Pediatric Ophthalmology and Strabismus, Department of Ophthalmology, Gavin Herbert Eye Institute, University of California-Irvine, Irvine, CA 92697

Abstract

Purpose—Nystagmus is a condition of involuntary eye movement. The causes for nystagmus may be congenital, idiopathic, or acquired. Considerable debate exists on the therapeutic potential of various surgical techniques. Currently, there are neither standardized protocols nor devices to record quantitative data on patients with clinical nystagmus undergoing various procedures at multiple centers to facilitate randomized prospective clinical trials.

Methods—The authors evaluated the reliability and variability of a commercially available infrared videonystagmography unit, by recording eye movement waveforms elicited from normal volunteers (n=117, 13 patients, 9 trials), by different examiners (A, B, and C). Five movement characteristics were examined including saccadic latency, velocity and precision and pursuit gain and velocity.

Results—The overall test/retest variability was low, where the median coefficient of variation of the three testers for all five eye movement categories was less than 15%, and less than 10% of the coefficients of variation calculated were more than 20%. However, there was a significant difference interobserver variability for all outcomes, except saccade latency.

Conclusions—The between-tester analysis was found to have greater variability than the test/retest reliability analysis. Future studies at multiple sites, using videonystagmography measurements should aim to have each patient repeatedly tested by the same tester. In anticipation of multicenter, randomized, prospective clinical trials of surgical procedures for nystagmus, standardized protocols for nystagmographic data collection and analysis must be validated both within and among participating centers.

#Correspondence to: Robert Lingua, MD, Clinical Professor, Director of Pediatric Ophthalmology and Strabismus, Gavin Herbert Eye Institute, University of California-Irvine, 850 Health Sciences Road, Irvine, CA 926, Tel: 949-824-7615, Fax: 949-824-2073, rlingua@uci.edu.

Introduction

Numerous surgical approaches have been reported for the treatment of nystagmus^{1,2}. A standardized method for the quantitative evaluation of these surgical improvements is lacking, which has led to a clinical standard of describing only qualitative improvements in nystagmus³. Several techniques have been developed for measuring and recording eye movements for the study of nystagmus in laboratory settings. For the past 40 years, scleral search coils (involving the use of a wire-embedded contact lens that generates recordable changes in the surrounding magnetic field as it moves with the eye) have served as the gold standard for human patients. Although precise spatially ($<1^\circ$) and temporally ($<1\text{ms}$), scleral search coils have limited clinical application in children due to their invasive nature^{4,5}.

An alternate strategy to this scleral search coil technique involves a video recording technology, called infrared videonystagmography (VNG). Several studies have compared the simultaneous recordings of VNG and scleral search coil methods in both human⁶⁻⁸ and non-human^{9,10} patients. These studies found that although the search coils were better for measuring torsional movements, the two systems gave comparable accuracy for eye position, especially in the horizontal axis⁹.

In the field of audiology, VNG systems are commonly used to diagnose vertigo and balance disorders¹¹. In ophthalmology, the use of objective eye movement recording for oculomotor assessment remains relatively rare¹². Data from the equipment manufacturers claim accuracy, though independent verification by ophthalmologists is lacking. To assess the value of using VNG systems to objectively quantify the surgical effect of a nystagmus procedure, it is important to define the reliability and variability of that particular system to quantitate movement parameters in any setting.

We report on the smooth pursuit and saccade eye movement waveforms elicited from normal patients while viewing standardized calibrated targets. Serial trials were performed on each patient to determine test/retest reliability. Additionally, multiple examiners tested each patient to determine between-tester variability. If system reliability can be confirmed by low variability among testers, and high reliability within a tester in normal subjects, then the system may be further used to accurately quantify the results of surgical effects in patients with nystagmus.

Patients and Methods

In this report, we investigated the eye movement waveforms elicited from normal patients and examined the reliability and variability of a commercially available infrared VNG unit (Interacoustics VO405, Interacoustics, Eden Prairie, MN). This study was approved by Institutional Review Boards at University of California-Irvine. After obtaining statistical consultation on the study design, 13 patients (at least 18 years of age) were recruited. Any patient with a history of eye surgery, eye trauma, or nystagmus was excluded. Three of the researchers for this study (AG, DM, JE) acted as testers after undergoing training with the VNG system, as described by the manufacturer (see manual for Interacoustics).

Interacoustics VNG

The VO405 Interacoustics VNG apparatus is commercially available and is used to record and analyze eye movement characteristics of induced nystagmus (caloric irrigation of the ear canal, or rapid changes in posture) in patients with dizziness and imbalance. It consists of a set of goggles with infrared cameras to simultaneously record bilateral eye movement (corneal light reflex through transparent prism goggles), a projector to display calibrated targets for the subject to view, and a computer software program to display, record, and analyze the eye movement data (Interacoustics).

The VNG recording software quantitates and displays characteristics of the fast and slow components of eye movement. The software is capable of analyzing components of spontaneous nystagmus (no light), nystagmus during gaze/fixation (light), pursuit, saccade and optokinetic nystagmus. To determine the VNG's reliability in the study of nystagmus, we presented subjects without clinical nystagmus two different stimulus paradigms: smooth pursuit and saccade targets from the VNG software to simulate the slow and fast components of a jerk nystagmus. In each testing, subjects were instructed to face the projection screen and follow the projected dot with their eyes while keeping their head still. In smooth pursuit, the dot would appear in the center of the screen and smoothly oscillate right to left 20°. In the saccade program, the dot would appear in the center of the screen and after a few seconds disappear and reappear at a new location in the horizontal plane prompting a rapid eye movement to regain fixation on the target. Each sequence lasted approximately 20 seconds. All patients were positioned 70 cm (the testing distance between the patient's eyes and test object) from the target screen, as per manufacturer's instructions. Calibrations were performed as recommended by the manufacturer. Simultaneous binocular video recordings of the eyes and two-dimensional linear tracings of the movement were displayed by the software. Data were analyzed quantitatively and qualitatively as numerical data was immediately applied to these eye movements by the proprietary software program.

Data Collection

Each of the three testers (A, B, and C) performed three consecutive VNG sessions on each subject for a total of nine sessions per patient. Each session consisted of one 20-second smooth pursuit trial (slow movement) followed by one 20-second saccade trial (fast movement). Between each session, the tester recalibrated the equipment by physically removing and refitting the goggles on the patient to simulate variability between multiple patient visits with the same tester. To minimize the effect of subject learning or fatigue, the order of testers were varied by randomly assigning each subject to one of the following tester sequences: ABC, ACB, BAC, BCA, CAB, and CBA. There were at least two subjects assigned to each order. This study design was created by consultation with a statistician (University of California-Irvine, Institute for Clinical and Translational Science).

The measures of fast and slow eye movements were generated by the software and displayed numerically. Data points included saccadic velocity in saccadic refixation (degrees per second), latency (time to initiation), precision (saccadic acquisition of target), gain (accuracy in pursuit of targets), and slow phase velocity of pursuit (degrees per second). Each eye was

analyzed independently and data were separated for movements of the right eye into right and left gaze and for the left eye into right and left gaze (Table 1).

Statistical Analysis

For test/retest reliability, the three VNG sessions performed by the same tester for the same study patient were summarized using coefficient of variation (CV = standard deviation/mean) for each outcome of both eyes and both moving directions. To evaluate possible tester difference and session order effect, a linear mixed model was applied for each outcome, adjusting for eyes and movement direction. In the linear mixed model, a covariance matrix was constructed to account for the correlation between eyes and between moving directions, and the 9 repeated sessions were modeled with a compound symmetry covariance structure. For this study, we performed a total of 35 hypotheses tests (7 tests for each outcome). To control for the multiple hypothesis tests, the false discovery rate adjustment method¹³, which controls the expected false discovery among all significant findings, was applied. When applicable, both original *P* value and false discover rates (ie, adjusted *P* value) are presented. All analyses were performed with SAS 9.3 (SAS Institute, Cary, NC) and the false discovery rate level was set at 0.05.

Results

Test/Retest Reliability

Thirteen CVs were obtained from the three trials performed by the same tester for each outcome by eyes, moving direction, and tester (Figure 1). The median and maximum CVs were 8.3% and 31.1% for latency, 8.1% and 39% for velocity, 7.7% and 34.2% for precision, 6.1% and 53.1% for gain, and 6.9% and 31.2% for SPV. Of the 260 CVs computed for each tester, 35 (13.5%), 7 (2.7%), and 35 (13.5%) CVs were more than 20% for tester A, B, and C, respectively (Figure 1).

Interobserver Variability

Results from the linear mixed model showed that after adjusting for eyes and moving direction, the order of the tester or trial session did not affect any of the outcomes. However, there was a significant difference in interobserver variability for all outcomes except latency. Although testers A and B were not greatly different from each other, tester C depicted significantly higher measures for velocity, precision and SPV, and lower measures for gain (all $P < 0.001$, all false discovery rates < 0.005) versus testers A and B. To visualize the difference of interobserver variability, the difference between the observed value and the mean value of each study subject was calculated. A box-and-whisker plot of the difference by each individual tester is illustrated in Figure 2.

Discussion

This report examined the eye movement waveforms elicited from normal patients to determine the reliability and variability of a commercially available infrared VNG unit. To calculate the amount of variability produced by a VNG system to track eye movements, it must be acknowledged that there are inherent confounding components which may be

difficult to quantify such as the inherent variability of the device or software, and any variability in the response of the patient (attention and ability to follow protocol). At this point, a study paradigm that tests human patients cannot be created where these variables are uniformly controlled.

Our results showed that the overall test/retest variability was low, where the median CV of the three testers for all five eye movement categories was less than 15%, and less than 10% of the CVs calculated were more than 20%. Although the CV values differed among testers, the CVs were not greatly different among the various VNG outcomes. With low test/retest variability, the VNG system (machine and human combined) is considered reliable to produce repeated measurements with consistency when operated by the same person. Results of future movement study data showing changes greater than 20% before and after treatment can confidently be attributed to treatment effect rather than VNG system variability, when patients are tested by the same examiner.

Although variability within each tester on the same patient was low, interobserver variability performing tests on the same patient was higher than preferred. Based on the results of the linear mixed model, there was significant interobserver variability for all categories measured, except latency. This disparity can be attributed to actions directly performed by each tester including how they placed the goggles on the patients as well as each tester's manual alignment of the cameras before calibration by the system.

Since there was a significant between-tester difference, future studies using repeated VNG measures should aim to have each patient examined by the same tester for all visits. Multiple testers taking measurements of the same patient across multiple visits could lead to a higher variability in the data. Also, testers confounded with the study condition, such as one tester in the preoperative setting and one in the postoperative setting, or one tester for one type of procedure and another for a different procedure, could lead to biased conclusions.

Acknowledgments

This study was partially supported by the National Center for Advancing Translational Sciences, National Institutes of Health (Grant UL1TR000153) and Research to Prevent Blindness.

References

1. Lee J. Surgical management of nystagmus. *Journal of the Royal Society of Medicine*. May; 2002 95(5):238–241. [PubMed: 11983764]
2. Lee JP. Surgical management of nystagmus. *Eye*. 1988; 2(Pt 1):44–47. [PubMed: 3410140]
3. Abel LA. Infantile nystagmus: current concepts in diagnosis and management. *Clinical & experimental optometry : journal of the Australian Optometrical Association*. Mar; 2006 89(2):57–65.
4. Collewijn H, van der Mark F, Jansen TC. Precise recording of human eye movements. *Vision research*. Mar; 1975 15(3):447–450. [PubMed: 1136166]
5. Robinson DA. A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field. *IEEE transactions on bio-medical engineering*. Oct.1963 10:137–145. [PubMed: 14121113]
6. DiScenna AO, Das V, Zivotofsky AZ, Seidman SH, Leigh RJ. Evaluation of a video tracking device for measurement of horizontal and vertical eye rotations during locomotion. *Journal of neuroscience methods*. May; 1995 58(1-2):89–94. [PubMed: 7475237]

7. van der Geest JN, Frens MA. Recording eye movements with video-oculography and scleral search coils: a direct comparison of two methods. *Journal of neuroscience methods*. Mar 15; 2002 114(2): 185–195. [PubMed: 11856570]
8. Traisk F, Bolzani R, Ygge J. A comparison between the magnetic scleral search coil and infrared reflection methods for saccadic eye movement analysis. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie*. Aug; 2005 243(8):791–797.
9. Kimmel DL, Mammo D, Newsome WT. Tracking the eye non-invasively: simultaneous comparison of the scleral search coil and optical tracking techniques in the macaque monkey. *Frontiers in behavioral neuroscience*. 2012; 6:49. [PubMed: 22912608]
10. Stahl JS, van Alphen AM, De Zeeuw CI. A comparison of video and magnetic search coil recordings of mouse eye movements. *Journal of neuroscience methods*. Jun 30; 2000 99(1-2):101–110. [PubMed: 10936649]
11. Jacobson, GP.; Shepard, NT. Balance function assessment and management. San Diego, CA: Plural Pub; 2008.
12. Bedell HE, Stevenson SB. Eye movement testing in clinical examination. *Vision research*. Sep 20.2013 90:32–37. [PubMed: 23416869]
13. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*. 1995; B(57):289–300.

Figure 1.

Box-and-whisker plot for coefficient of variation (CV) from three trials performed by the same tester for the same subject (n=13). Circles represent individual CV (%); boxes represent the 1st quartile (Q1), median, and the 3rd quartile (Q3); the lower whisker extends to the minimum observation; the upper whisker extends to the maximum observation below upper fence, (where upper fence = $Q3 + 1.5 \times (Q3 - Q1)$); and observations above the upper fence are outliers. The dashed line shows a reference line of 20% CV.

Figure 2.

Box-and-whisker plot, for each tester ($n=39$), of the difference between the observed value and the mean value of each study subject. Circles represent individual difference; boxes represent the 1st quartile (Q1), median, and the 3rd quartile (Q3); the lower whisker extends to the minimum observation above the lower fence, (where lower fence = $Q1 - 1.5 \cdot (Q3 - Q1)$); the upper whisker extends to the maximum observation below upper fence, (where upper fence = $Q3 + 1.5 \cdot (Q3 - Q1)$); and observations beyond the lower or upper fences are outliers. The dashed line shows the reference line of 0.

Distribution of the Videonystagmography Outcomes for Both Eyes and Both Moving Directions (n=117, 13 Patients, 9 Trials)^a**Table 1**

Outcome	Eyes	Moving Direction	Minimum	25 th percentile	Median	75 th percentile	Maximum	Mean	Standard Deviation
Latency (sec)	OD	to Left	98	133	149	161	208	148.8	21.4
		to Right	103	135.5	157	167.5	200	152.5	21.5
	OS	to Left	98	133.5	148	158.5	206	147.2	20.2
		to Right	103	135.5	158	169.5	210	154.2	22.0
Velocity (°/sec)	OD	to Left	342	604	654	737	908	663.6	103.4
		to Right	318	598.5	661	736.5	942	669.0	105.1
	OS	to Left	359	606	668	736.5	959	664.4	101.0
		to Right	282	605	663	722.5	939	665.9	100.2
Precision* (%)	OD	to Left	61	102.5	111	122	183	112.2	16.4
		to Right	57	100	109	117	144	108.6	13.6
	OS	to Left	59	102	110	119.5	189	110.5	16.9
		to Right	56	104	112	120	211	112.7	17.2
Gain* (%)	OD	to Left	25	74.5	83	88	98	79.6	12.1
		to Right	35	75.5	84	88.5	97	80.6	11.7
	OS	to Left	23	64	81	85	95	74.7	13.5
		to Right	36	71	82	87	104	77.9	13.1
SPV (°/sec)	OD	to Left	13	19	21	23	34	21.2	3.6
		to Right	12	20	22	23	36	22.0	3.7
	OS	to Left	13	19	20	22	35	20.9	3.7
		to Right	12	20	22	24	36	22.6	3.6

Distribution of the five data categories for both eyes and both moving directions

min = minimum; max = maximum; SD = standard deviation; OD = right eye; OS = left eye; SPV = slow phase velocity

^a Distribution of the five data categories for both eyes and both moving directions.^b Values represent the accuracy of tracking, where 100% = perfect correlation of eye and target position.