V1 V2 Pilot Study Data Analysis

Last Update: Dec 2024

TABLE OF CONTENTS

Table of Contents 1

Introduction 2

Method 2

- Experimental Setup 2
- Test Exercises 2
- Participants 3
- Statistical Analysis 4
  - Test-retest Reliability Analysis 4
  - Normal-Adverse Tests: Group Statistical Differences 5

V1 Results 5

- Visualizations 5
- Test-Retest Reliability 6
- Adverse-Normal Statistical Analysis 7

V2 Results 9

- Visualizations 9
- Test-Retest Reliability 12
- Adverse-Normal Statistical Analysis 14

V1-V2 Results Comparison 16

Conclusion 19

References 19

INTRODUCTION

This document contains the analysis of the data collected from the Test-Retest NeuroFlex-Local Pilot Study over the week of November 4, 2024.

The purpose of this study is to evaluate the test-retest reliability of the legacy test exercises (NeuroFlex L1.1.0 – release) and to compare the results with the test-retest reliability of newly developed and improved version of test exercises targeting the same vestibulo-ocular functions (NeuroFlex L2.0.0 – development version).

In the current document, for ease of use, we will distinguish versions of test exercises as V1 (L1.1.10) and V2 (L2.0.0).

METHOD

EXPERIMENTAL SETUP

As described in Clinical Trial Plan Oct 2024.docx, the experiment consists of four test protocols with each version of the NeuroFlex Local software: L1.1.10 and L2.0.0 (dev). A test protocol consists in a set of multiple test exercises. They were recorded over two days, one day apart. Two test protocols with L1.1.10 and two test protocols with L2.0.0 were recorded per day. Each two consecutive tests during a day are between 1.5-3 hours apart.

All the test protocols are performed in normal conditions, i.e. the participant is sitting in a quiet environment without distraction, except for the last tests of the second day for each version conducted while the participant answers medium-level mathematical questions. The condition of this situation is called “adverse” and is used to evaluate the validity of the biomarkers of NeuroFlex vestibular and oculomotor tests.

TEST EXERCISES

The below table describes the test exercises that were assessed during the experiment, in the same order that they were presented to the users. We also added the type of vestibulo-ocular function that each of them assesses.

Table 1: Test exercises tested by each NeuroFlex Local version

#	L1.1.10 (V1)	L2.0.0 (dev) (V2)
1	Smooth Pursuit (Head Free) *(Pursuit)*	Spontaneous Nystagmus v2 *(Spontaneous Nystagmus)*
2	Active Visual VOR (Horizontal) *(VOR)*	VOMS *(VOR)*
3	Active Visual VOR (Vertical) *(VOR)*	Smooth Pursuit H (Head Fixed) v2 *(Pursuit)*
4	Smooth Pursuit (Head Fixed) *(Pursuit)*	Saccades 2D v2 *(Saccades)*

5	Saccades 2D *(Saccades)*	Anti-saccades v2 *(Anti-saccades)*
6	Antisaccades *(Anti-saccades)*	Active VOR in the Dark *(VOR)*
7	Optokinetic Response (OKR/OKN) *(OKN)*	Optokinetic Response (OKR/OKN) v2 *(OKN)*
8	Spontaneous Nystagmus *(Nystagmus)*	Stare Optokinetic Response *(OKN)*
9	-	Smooth Pursuit (Head Fixed) v2 *(Pursuit)*
10	-	Active Visual VOR (Horizontal) v2 *(VOR)*
11	-	Gaze Evoked Nystagmus v2 *(Nystagmus)*

PARTICIPANTS

Details of the participants' information and approximate testing times are displayed in Table 2. Participants comprised of 8 males and 2 females.

Table 2. Full test schedule of the pilot study.

	v1 patient ID	**v2 patient ID**	**Sex**	**Day 1 Date**	**Day 1 - v2 (1) start time**	**Day 1 - v1 (1) start time**	**Day 1 - v2 (2) start time**	**Day 1 - v1 (2) start time**	**Day 2 Date**	**Day 2 - v2 (3) start time**	**Day 2 - v1 (3) start time**	**Day 2 - v2 (distraction) start time**	**Day 2 - v1 (distraction) start time**
1	2013-2259-59	2055-5740-92	M	Mon, Nov 4	9:20 AM	11:33 AM	13:45 PM	15:55 PM	Wed, Nov 6		11:30 AM	13:40 PM	15:42 PM
2	8061-2868-40	3055-2747-92	F	Mon, Nov 4	9:20 AM	11:33 AM	13:45 PM	15:55 PM	Wed, Nov 6	9:20 AM	11:30 AM	13:40 PM	16:10 PM
3	1264-3387-82	4057-0749-72	M	Mon, Nov 4	10:04 AM	11:55 AM	14:50 PM	16:37 PM	Wed, Nov 6	9:50 AM	11:40 PM	14:45 PM	16:30 PM
4	6890-2944-23	0056-6740-63	M	Mon, Nov 4	10:04 AM	11:55 AM	14:20 PM	16:08 PM	Wed, Nov 6	9:50 AM	11:40 PM	15:00 PM	16:42 PM
5	7001-7431-60	8054-6749-79	M	Mon, Nov 4	10:30 AM	12:23 PM	14:20 PM	16:08 PM	Wed, Nov 6	10:30 AM	11:54 AM	14:16 PM	15:55 PM
6	7067-1265-47	7057-3745-71	M	Mon, Nov 4	10:30 AM	12:23 PM	14:50 PM	16:37 PM	Wed, Nov 6	10:30 AM	11:54 AM	14:30 PM	16:15 PM
7	8985-8363-07	5051-1733-85	F	Mon, Nov 4	11:10 AM	11:30 AM	12:56 PM	15:25 PM	Wed, Nov 6	11:00 AM	12:50 PM	15:43 PM	17:10 PM
8	3221-1335-16	3050-9739-92	M	Mon, Nov 4	11:10 AM	11:30 AM	12:56 PM	15:25 PM	Wed, Nov 6	11:00 AM	12:50 PM	15:18 PM	16:58 PM
9	8796-6233-93	0058-2735-95	M	Mon, Nov 4	13:20 PM	15:15 PM	absent	absent	Wed, Nov 6	13:10 PM	15:15 PM	absent	absent
10	1871-0910-16	8457-2542-33	M	Tue, Nov 5	9:45 AM	11:40 AM	13:30 PM	16:00 PM	Thu, Nov 7	9:43 AM	11:30 AM	13:38 PM	16:00 PM

STATISTICAL ANALYSIS

The test-retest reliability is measured by the intraclass correlation coefficient (ICC) metrics [1, 2], and the differences between normal and adverse conditions using the Whitney U test (MNU). Detailed instructions, adopted from Clinical Trial Plan Oct 2024.docx, are explained below.

Test-retest Reliability Analysis

Once the metrics have been computed, the data can be analyzed. A holistic approach for a reliability analysis is proposed in [1, 2], which is as follows:

For each metric, the intraclass correlation coefficient (ICC) will be calculated. There are many different types of intraclass correlation coefficients that can be computed depending on the conditions and the experiment characteristics. Compute the following three single-score ICC formulas, i.e.,

i. **Classical ICC**, denoted by ICC(1) or ICC(1,1): Since we have only one device, the "*one-way, random effects*, single measurement" could be an appropriate choice. Here, there is one systematic source of variance associated with differences among measured subjects, *without a bias*. The raters here are considered random exam tests performed by the device (the randomness comes from variabilities in the device usage, time of the day, individual differences, etc.). ICC(1,1) assesses the variability due to the subjects versus the variability due to measurement error. Here, measurements are independent, and you are not interested in comparing different raters [1].

ii. ICC(2,1): it is a "*two-way random effect*, signal measurement", assuming a *random bias*, i.e., a random sampling of measurement biases is introduced. The ICC assumptions can lead to the definitions of Agreement ICC and Consistency ICC, similar to those computed from the "two-way mixed/fixed effect, single measurement" model which will be described below.

iii. ICC(3,1): since the device (here considered as the rater) is fixed, the "*two-way mixed/fixed effect*, single measurement", denoted by ICC(3,1), is a suitable choice. Here, the set of both subjects and raters are fixed variables (*fixed bias*) rather than randomly selected, and we would like to take all measurements into account rather than their average. Note that the formula for getting the ICC between two-way mixed effects and two-way random effects absolute agreement, single rater/measurement is the same. However, the main difference is in the interpretation of the results (See [4] under ICC). Here, the same set of subjects is measured by the same rater or device, but the device or rater is considered a fixed effect (i.e., you are only interested in the reliability of this specific rater/device and not generalizing to others). The assumptions of this model and the previous one (two-way models) lead to identical formulas for the:
   - **Agreement ICC**, or ICC(A,1), and
   - **Consistency ICC,** or ICC(C,1).

If ICC(1,1) ≈ ICC(A,1) ≈ ICC(C,1), we can infer the absence of bias (systematic error) and hence report either the classical (one-way random) ICC(1) or ICC(3,1).

Otherwise, especially if the ICC(C,1) > ICC(A,1), there is a likely non-negligible bias present. Here, the classical ICC(1,1) is invalid and misleading. Instead, we report both ICC(A,1) and ICC(C,1) as they provide complementary information [4]. In this scenario, it is suggested that the Standard Error of Measurement (SEM) and associated SEs be reported since they provide a sense of absolute error in the measurements' unit, complementary to the unitless ICC. The SEs help with the interpretation of participants' scores as they provide confidence intervals (Cis) for true scores [3]. For instance, the 95% CI for the measured score compared with the true score is ±1.96×SEM.

Normal-Adverse Tests: Group Statistical Differences

To quantify how adverse conditions affect the vestibular oculomotor functions, we aim to show how the new revised or old protocols can distinguish between tests in normal and adverse conditions (for the full study, between healthy and those with neurological conditions).

To this end, for each metric and between the two exam groups (all baselines and retests in the normal condition in one group and all adverse in the other), we will conduct a two-tailed (or two-sided) MWU test for adverse vs normal, and return the "statistic" and "p-value" outputs

V1 RESULTS

In this section, we will visualize and analyze all V1 collected data. The analyses are two-fold:

1. Determining the test-retest reliability
2. Determining the validity, i.e., how much and in which directions the metrics (biomarkers computed based on each NeuroFlex protocol) change when the participant performs them in adverse versus normal conditions.

VISUALIZATIONS

Figure 1 depicts box plots for all 61 V1 metrics, differently colored for the first, second, third, and distraction (fourth) sessions. Note that to allow for easier comparison, some metrics are scaled to fit in one plot. We observe that changes between normal and distraction conditions are visible in certain metrics. For instance, the below metrics are worse when collected in adverse conditions versus in normal:

- Vergence standard deviation in AVVOR H & V
- Mean latency and directional accuracy in Antisaccades
- Acquisition error in Saccades and antisaccades
- Mean error of Smooth in pursuit head fixed & free
- Number of saccades in Smooth pursuit head fixed & free

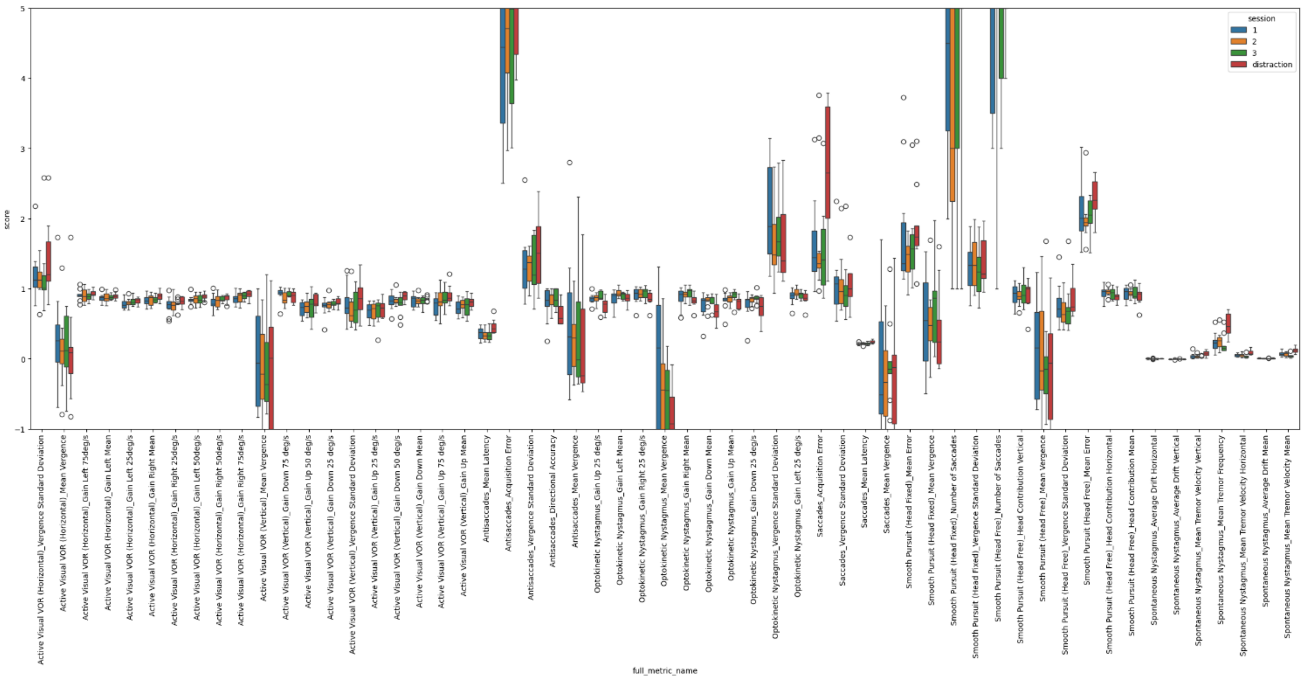

Figure 1. Box plot of all 61 V1 metrics, differently colored for the first, second, third, and distraction sessions.

In the next sections, we statistically quantify how reliable the test retests are and then how different are adverse tests from the normal ones.

TEST-RETEST RELIABILITY

We use three reliability categories [3]:

1. Excellent reliability for ICC > 0.9
2. Good reliability for 0.75 < ICC ≤ 0.9
3. Medium reliability for 0.5 < ICC ≤ 0.75

Table 3 displays the number of V1 metrics per each protocol with ICC(3,1) > 0.5. This is also visualized in Figure 2. Here, we observe that 26 V1 metrics have medium reliability and higher, where the protocols are ordered in Table 3 from having the most reliable metrics to the least.

Table 3. The distribution of V1 metrics counts with ICC(3,1) above 0.5.

Protocol Name	Count 0.5<ICC(3,1)≤0.75	Count 0.75<ICC(3,1)≤0.9	Count ICC(3,1)>0.9
Active Visual VOR (Horizontal)	7	1	0
Antisaccades	2	1	1
Active Visual VOR (Vertical)	3	1	0
Optokinetic Nystagmus	4	0	0
Saccades	1	2	0

Smooth Pursuit (Head Free)	2	0	0
Spontaneous Nystagmus	1	0	0
Smooth Pursuit (Head Fixed)	0	0	0
SUM	20	5	1

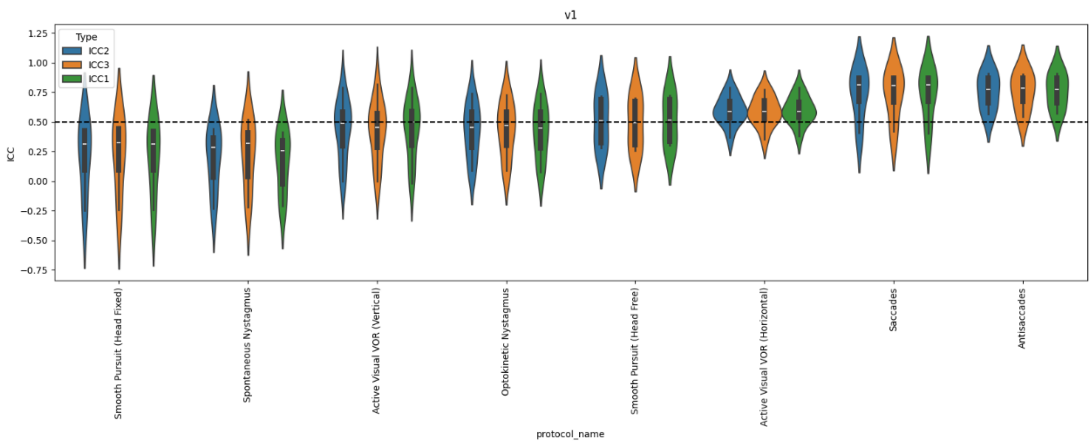

Figure 2. Violin plot of the ICC values for all V1 metrics.

ADVERSE-NORMAL STATISTICAL ANALYSIS

Next, we evaluate the statistical differences between metrics collected during normal versus adverse conditions where participants were cognitively engaged through solving mathematical questions. Visualized in Figure 3, we observe that the 10 V1 metrics, also listed in Table 4, with p-value < 0.05 show significant differences between normal versus adverse tests.

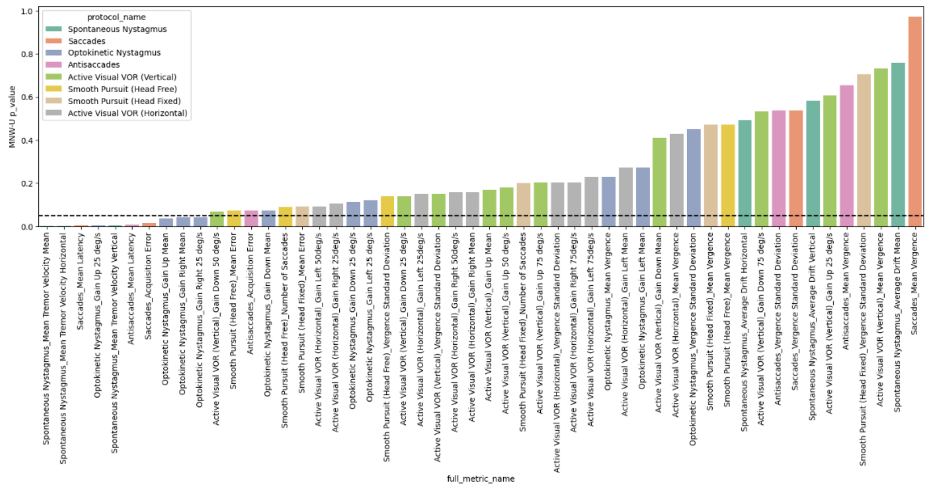

Figure 3. Sorted p-values using MWU test on V1 metrics of normal versus adversely collected data.

Table 4. V1 metric list with MNU p-value < 0.05.

	protocol_name	metric_name	MNU p_value	MNU statistic
1	Antisaccades	Mean Latency	0.008	208
2	Optokinetic Nystagmus	Gain Up 25 deg/s	0.005	48
3	Optokinetic Nystagmus	Gain Up Mean	0.036	69
4	Optokinetic Nystagmus	Gain Right Mean	0.043	71
5	Optokinetic Nystagmus	Gain Right 25 deg/s	0.043	71
6	Saccades	Mean Latency	0.004	216
7	Saccades	Acquisition Error	0.015	202
8	Spontaneous Nystagmus	Mean Tremor Velocity Mean	0.001	226
9	Spontaneous Nystagmus	Mean Tremor Velocity Horizontal	0.002	223
10	Spontaneous Nystagmus	Mean Tremor Velocity Vertical	0.005	212

Among the differentiable metrics listed in Table 4, four of them also have medium-good reliability, as shown in Table 5.

Table 5. V1 metric list with both ICC>0.5 and MNU p-value < 0.05

protocol_name	metric_category	metric_name	MNU p_value	MNU statistic	ICC
Saccades	latency	Mean Latency	0.003517	216	0.886141
	error	Acquisition Error	0.014777	202	0.749682
Optokinetic Nystagmus	gain	Gain Right 25 deg/s	0.042788	71	0.529455
Antisaccades	latency	Mean Latency	0.008198	208	0.854496

V2 RESULTS

Visualizations

Similar to V1 plots, the large box plots in Figure 4 and Figure 5, we display box plots of V2 metrics. Note that some metrics are scaled to help visualize better in a plot. Since V2 includes many additional metrics (a total of 3692 in v2 versus 57 in v1 as displayed in all_v2_metrics.csv)_than V1, we only plot those with good and excellent reliability scores with Normal-Adverse distinguishable powers (i.e., ICC(3,1) > 0.75, MNU test p value<0.05). For instance, the rate of fast phases is significantly higher when in distraction, while their duration drops down, and the asymmetry between leftward and rightward movements increases. Another example is the significant increase in saccade and antisaccade MCS (main centrifugal saccade) latency.

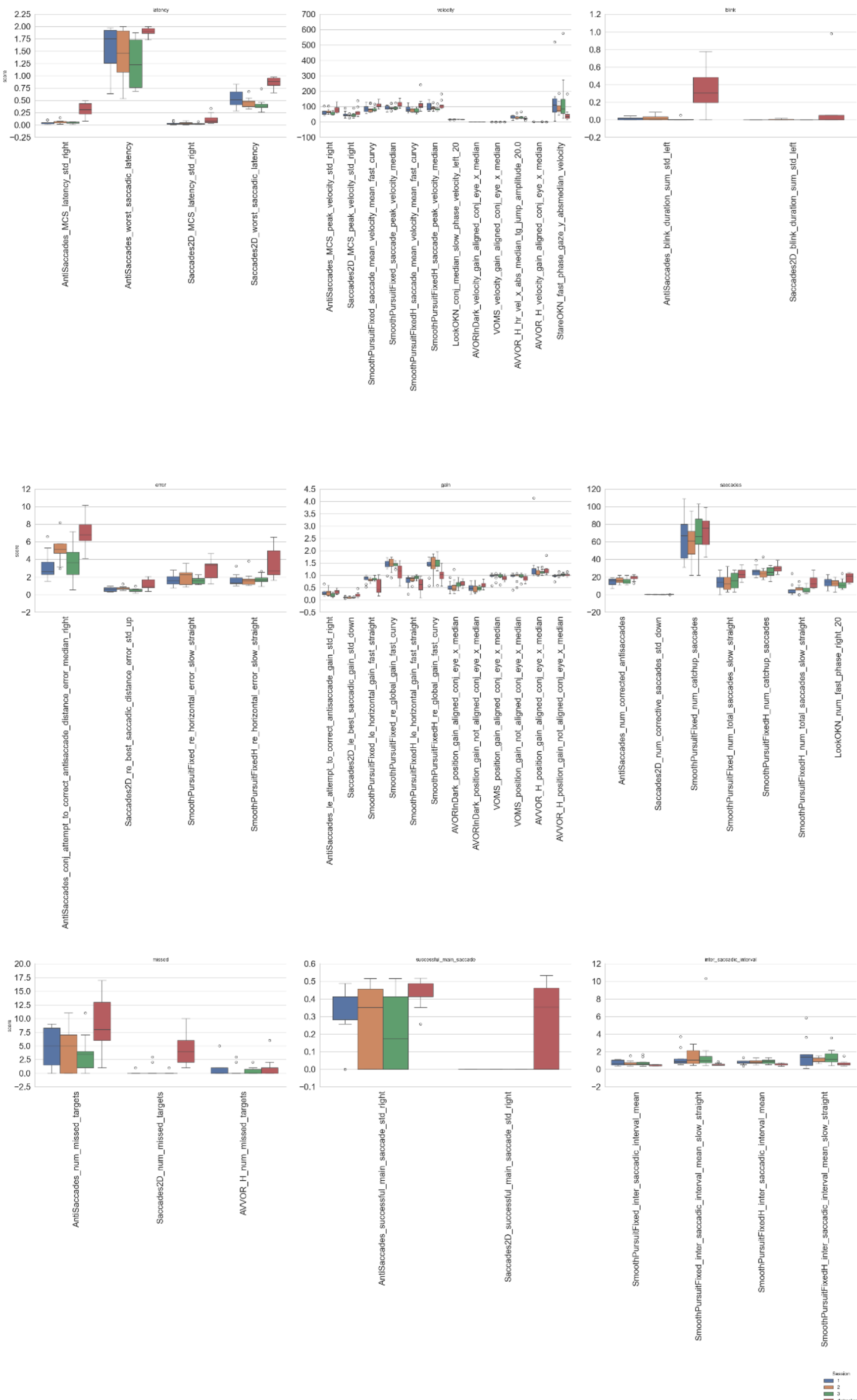

Figure 4. Scatter of V2 metrics with ICC>0.75 and MNU p value<0.05, computed based on all normal and adverse sessions.

Figure 5. Continued Scatter of V2 metrics with ICC>0.75 and MNU p value<0.05, computed based on all normal and adverse sessions.

TEST-RETEST RELIABILITY

Similar to V1, Figure 7 show boxplots of the protocol metrics count in V2 tests with separately with >medium, >good, and >high reliability.

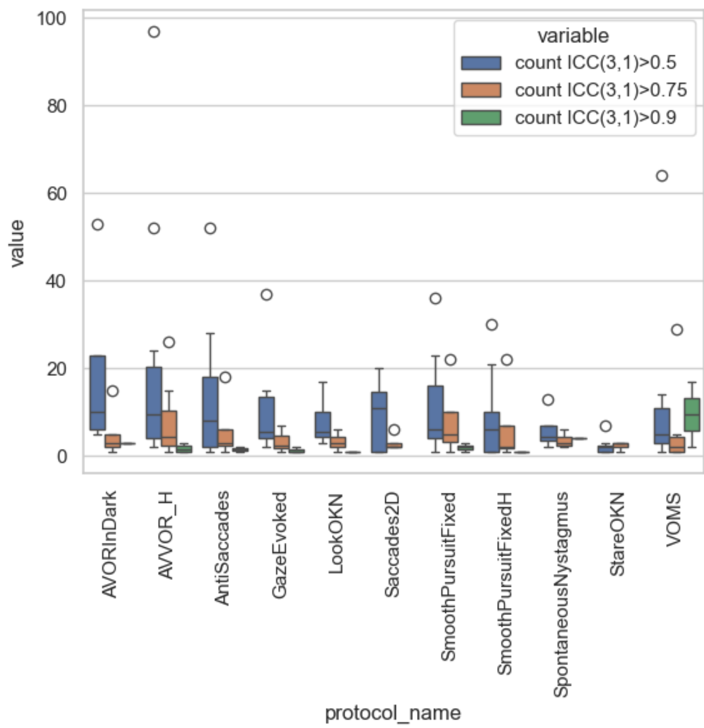

Figure 6. Number of V2 metrics per protocol with ICC(3, 1)>0.5, 0.75, 0.9.

In order to ensure that the reliable metric counts are diverse (i.e., not all of them are errors or gains), Table 2 displays metric diversity per protocol which is defined as the number of metric categories (i.e., error, gain, count, etc.) the metrics to which each protocol with ICC>0.5 belongs. The total number of metrics per protocol and per category are displayed "metric_category" column of Table 8. The categories are defined based on finding a substring of the category name, from the following list (if two are found in one metric name, the category that comes last in the list is assigned):

'blink', 'error', 'saccades', 'auc', 'slip', 'slow_phase_amp', 'fast_phase_amp', 'jump_amp', 'vergence', 'velocity', 'latency', 'okan', 'asymmetry', 'movements', 'proportion', 'rate', 'overshoot', '_vel_', 'pso', 'drift', 'successful_main_saccade', 'missed', 'distance_to_tg', 'inter_saccadic_interval', 'saccade_duration', 'consecutive', 'single_nystagmus', 'phase_duration', 'gaze_x_std', 'gaze_y_std', 'offset', 'num_fast_phase', 'std_of_gaze', 'peak_power', 'gain'

Table 6. The diversity of V2 metrics (number of non-zero metric categories) counts with ICC(3,1)>0.5.

Protocol Name	Metric Diversity
AntiSaccades	10
LookOKN	10
AVVOR_H	10
SmoothPursuitFixedH	9
GazeEvoked	8
Saccades2D	8
StareOKN	7
SmoothPursuitFixed	7
VOMS	7
AVORInDark	5
SpontaneousNystagmus	4

To have a final visualization of the reliabilities of V2 metrics Figure 7 displays the distribution of ICC values across metrics with ICC>0.5. Wider bars are indicative of more metrics.

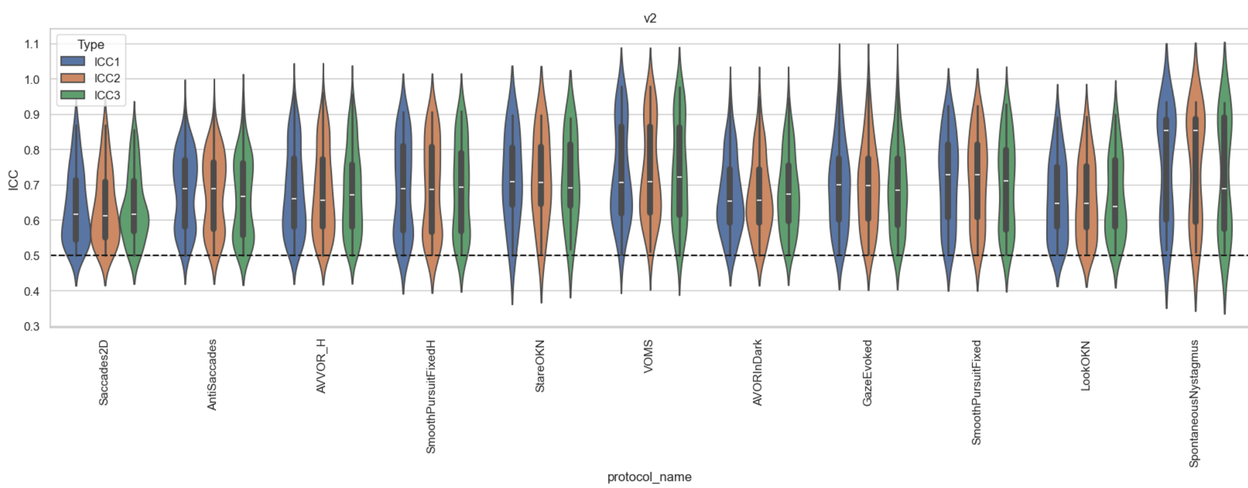

Figure 7. Violin plot of the ICC values for all v2 metrics with ICC(3,1) > 0.5.

ADVERSE-NORMAL STATISTICAL ANALYSIS

Similar to V1, Figure 8 and Figure 9 list the V2 metrics that have significant differences between normal versus adverse tests (i.e., with MNU p-value < 0.05).

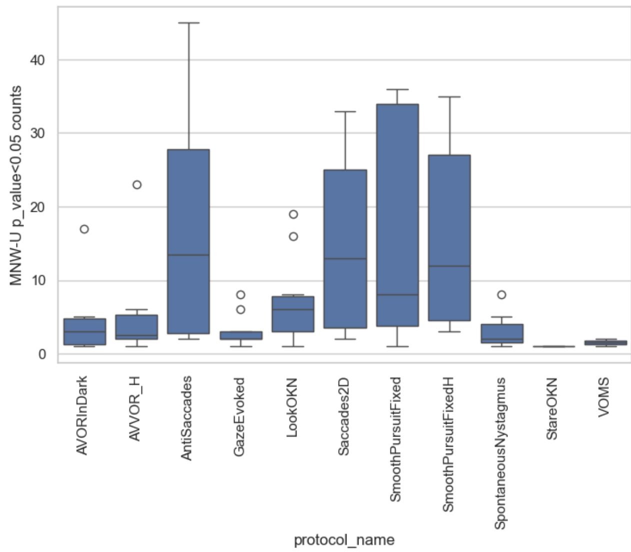

Figure 8. The number of metrics in each V2 protocol shows significant differences between normal and adverse conditions.

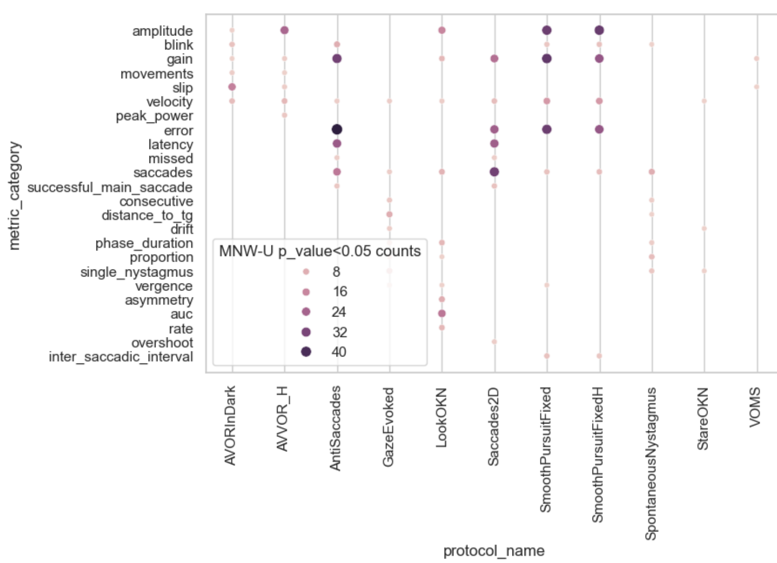

Figure 9. The circle sizes represent the number of metrics categories in each V2 protocol that show significant differences between normal and adverse conditions.

Next, Table 7 lists the 46 metrics that are both reliable (ICC>0.5) and normal-adverse differentiable (MNU p value<0.05). We observe that 16 of these metrics are based on monocular eye movements.

Table 7. V2 metric list with both medium reliability (ICC>0.5) and high differentiability (MNU p-value<0.05).

protocol_name	metric_category	metric_name	MNW-U p_value	MNW-U statistic	ICC
VOMS	slip	Group_slip_leftwards_100-150deg/s_re_x_mean	0.04	61	0.90
	gain	position_gain_aligned_conj_eye_x_median	0.02	55	0.62
StareOKN	velocity	fast_phase_gaze_y_absmedian_velocity	0.01	41	0.62
SpontaneousNystagmus	proportion	fixation_proportion_UNguided	0.04	71	0.58
SmoothPursuitFixedH	inter_saccadic_interval	inter_saccadic_interval_mean	0.00	44	0.68
	saccades	num_catchup_saccades	0.03	193.5	0.66
	amplitude	re_gaze_x_jump_amp_mean_slow_straight	0.04	172	0.56
	error	re_horizontal_error_slow_straight	0.00	227	0.52
	velocity	saccade_mean_velocity_mean_fast_curvy	0.00	215	0.51
	gain	le_horizontal_gain_fast_straight	0.03	67	0.50
SmoothPursuitFixed	saccades	num_total_saccades_slow_straight	0.01	204.5	0.80
	error	re_horizontal_error_slow_straight	0.02	199	0.57
	amplitude	re_gaze_x_jump_amp_mean_fast_straight	0.01	211	0.56
	velocity	saccade_peak_velocity_median	0.00	226	0.55
	inter_saccadic_interval	inter_saccadic_interval_mean_slow_straight	0.00	45	0.52
	gain	re_global_gain_fast_curvy	0.01	51	0.51
Saccades2D	latency	worst_saccadic_latency	0.00	251.5	0.56
	saccades	num_corrective_saccades_std_down	0.02	194	0.54
	velocity	MCS_peak_velocity_std_right	0.01	207	0.54
	gain	le_best_saccadic_gain_std_down	0.00	231	0.53
	error	re_best_saccadic_distance_error_std_up	0.01	204	0.52
LookOKN	saccades	num_fast_phase_right_20	0.01	205	0.73
	rate	rate_fast_phase_right_20	0.02	201.5	0.73
	proportion	proportion_fast_phase	0.02	63	0.68
	phase_duration	median_fast_phase_duration_right_60	0.01	55	0.63
	auc	le_auc_left_60	0.02	201	0.57
	velocity	conj_median_slow_phase_velocity_left_20	0.04	71	0.52
	amplitude	re_median_slow_phase_amplitude	0.04	69	0.50
GazeEvoked	consecutive	max_num_consecutive_nystagmus_UNguided_up	0.05	176	0.73
	single_nystagmus	freq_total_single_nystagmus_UNguided_up	0.04	180	0.69
	distance_to_tg	conj_distance_to_tg_std_Guided_up	0.02	198	0.55
	phase_duration	nystagmus_slowPhase_duration_median_Guided_right	0.04	39	0.52
AntiSaccades	missed	num_missed_targets	0.01	204.5	0.68
	successful_main_saccade	successful_main_saccade_std_right	0.04	189	0.61
	blink	blink_duration_sum_std_left	0.01	93.5	0.57
	gain	le_attempt_to_correct_antisaccade_gain_std_right	0.02	198	0.54
	error	conj_attempt_to_correct_antisaccade_distance_e...	0.00	223	0.54
	saccades	num_corrected_antisaccades	0.01	203	0.51
	latency	MCS_latency_std_right	0.00	257	0.50
AVVOR_H	amplitude	linearFit_deviation_rightwards_100-150deg/s_re...	0.01	17	0.73
	velocity	hr_vel_x_abs_median_tg_jump_amplitude_20.0	0.01	53	0.65
	slip	Group_slip_leftwards_100-150deg/s_le_x_median	0.04	80	0.50
AVORInDark	gain	position_gain_not_aligned_conj_eye_x_median	0.02	200	0.60
	velocity	velocity_gain_aligned_conj_eye_x_median	0.00	213	0.60
	amplitude	saccade_overt_re_y_jump_amp_max	0.02	37	0.55
	slip	Group_slip_leftwards_100-150deg/s_re_x_median	0.03	193	0.51

V1-V2 RESULTS COMPARISON

Figure 10, Figure 11, and Table 8 depict the analysis result of the V1 and V2 studies in terms of the number and diversity of metrics that are deemed reliable, and/or the difference between normal and adverse conditions.

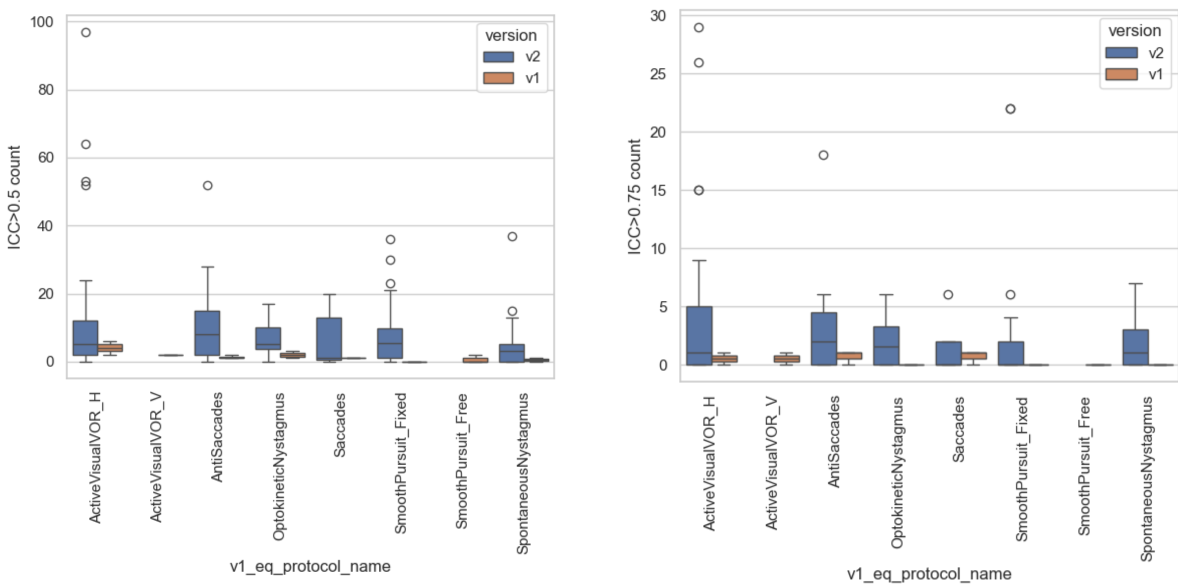

Figure 10. Comparison of the number of metrics for v1 versus v2 protocols with >medium reliability (left) and >good reliability scores (right).

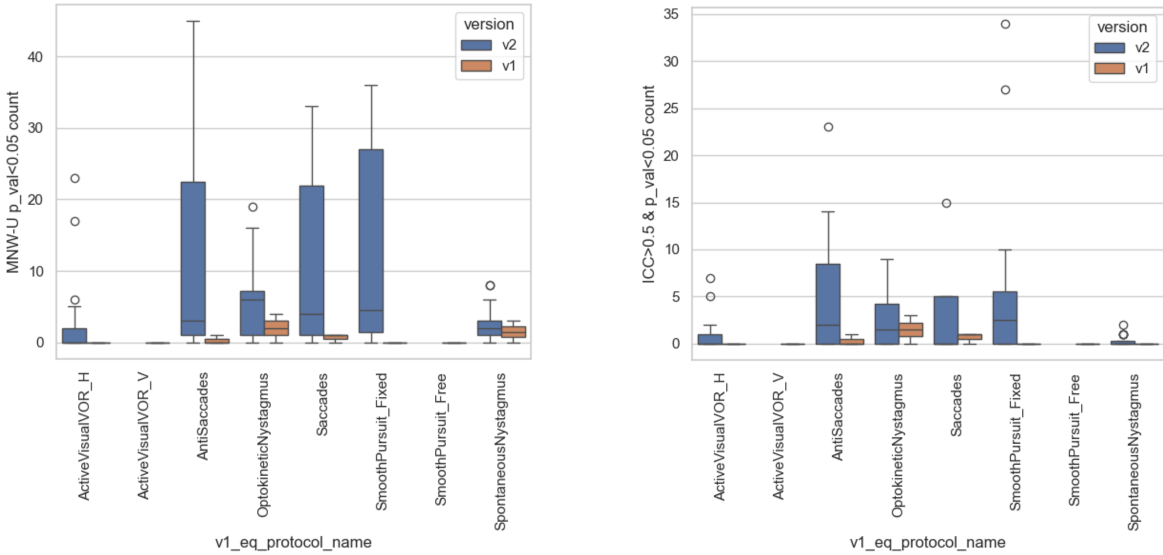

Figure 11. Comparison of the number of metrics for v1 versus v2 protocols that significantly differentiate Adverse from Normal (left) and those which on top of that, have >medium reliability (right).

Table 8. Summary table for v1 and v2 metrics including the high ICC and MNU p_value<0.05 counts.

v1_eq_protocol_name	protocol_name	version	metric_category	metric_category_count	ICC>0.5 count	ICC>0.75 count	MNW-U p_val<0.05 count	ICC>0.5 & p_val<0.05 count	ICC>0.75 & p_val<0.05 count	ICC median	varInter median	varIntra median	MNW-U statistic median
ActiveVisualVOR_H	AVORInDark	v2	amplitude	54	10	1	1	1	0	0.586996	0.000147	5.84E-05	37
		v2	blink	5	0	0	2	0	0				204.75
		v2	gain	12	5	2	1	1	0	0.654003	2.74E-07	9.17E-08	200
		v2	movements	8	0	0	2	0	0				188.75
		v2	peak_power	4	0	0	0	0	0				
		v2	slip	99	53	15	17	7	0	0.680534	0.020025	0.006379	193
		v2	velocity	51	23	5	5	5	2	0.654202	0.008744	0.000846	211
		v2	vergence	6	6	3	0	0	0	0.753493	9.00E-06	4.24E-06	
	AVVOR_H	v2	amplitude	549	52	15	23	1	0	0.693395	0.000223	9.91E-07	114.5
		v2	blink	20	2	1	0	0	0	0.696984	4.85E-06	1.22E-07	
		v2	gain	48	10	3	2	0	0	0.710168	0.000159	9.38E-05	204.5
		v2	missed	8	0	0	0	0	0				
		v2	movements	32	10	6	0	0	0	0.751162	1.68E-05	5.37E-06	
		v2	offset	4	4	0	0	0	0	0.547038	3.72E-06	2.02E-06	
		v2	peak_power	16	2	1	3	0	0	0.678766	5.2E+11	5.98E+10	13
		v2	slip	396	97	26	2	1	0	0.653362	0.000394	6.69E-05	76.5
		v2	std_of_gaze	31	9	0	0	0	0	0.596867	2.87E-06	8.53E-07	
		v2	velocity	204	24	9	6	1	0	0.66141	0.002943	0.000559	149.5
		v2	vergence	6	4	3	0	0	0	0.807422	7.86E-06	1.17E-06	
	Active Visual VOR (Horizontal)	v1	gain	8	6	0	0	0	0	0.594196	0.043522	0.016763	
		v1	vergence	2	2	1	0	0	0	0.685172	5.90E-05	4.56E-05	
	VOMS	v2	amplitude	54	5	1	0	0	0	0.705016	0.007174	0.0003	
		v2	blink	5	0	0	0	0	0				
		v2	gain	12	8	3	2	2	1	0.674763	9.13E-08	3.04E-08	56.5
		v2	movements	8	2	0	0	0	0	0.527196	3.14E-05	2.69E-06	
		v2	peak_power	4	1	1	0	0	0	0.85189	7.31E+11	2.5E+10	
		v2	slip	99	64	29	1	1	1	0.74013	0.007302	0.000735	61
		v2	velocity	51	14	5	0	0	0	0.710919	0.003077	0.001504	
		v2	vergence	6	4	1	0	0	0	0.719805	9.59E-06	5.65E-06	
	Active Visual VOR (Vertical)	v1	gain	8	2	0	0	0	0	0.570488	0.150664	0.041272	
ActiveVisualVOR_V		v1	vergence	2	2	1	0	0	0	0.700903	0.000313	5.24E-05	

AntiSaccades	AntiSaccades	v2	blink	24	8	2	8	1	0	0.674101	1.90E-08	2.95E-09	94.75
		v2	error	123	28	3	45	14	3	0.583609	1.88E-05	1.05E-05	214
		v2	gain	123	52	18	33	23	4	0.680425	2.60E-07	1.16E-07	70
		v2	latency	41	9	3	26	8	3	0.657491	1.95E-08	3.15E-09	220.25
		v2	missed	2	2	0	2	2	0	0.678926	8.70E-05	3.09E-05	204.75
		v2	overshoot	4	2	0	0	0	0	0.659998	2.24E-07	6.63E-08	
		v2	pso	10	0	0	0	0	0				
		v2	saccades	60	21	6	19	9	0	0.666568	7.67E-07	2.30E-07	223
		v2	successful_main_saccade	4	2	1	3	2	1	0.69848	5.13E-07	4.23E-08	189
		v2	velocity	13	8	6	2	0	0	0.760292	0.021427	0.004723	200.5
		v2	vergence	2	1	0	0	0	0	0.676024	2.91E-06	1.27E-06	
	Antisaccades	v1	error	1	1	0	0	0	0	0.543086	0.002828	0.000473	
		v1	latency	1	1	1	1	1	1	0.854496	4.697532	0.959553	208
		v1	vergence	2	2	1	0	0	0	0.809543	0.00111	9.96E-05	
OptokineticNystagmus	LookOKN	v2	amplitude	42	17	6	16	9	2	0.60653	0.002137	0.00024	69
		v2	asymmetry	21	0	0	8	0	0				50
		v2	auc	28	10	2	19	8	2	0.6255	138.787	85.93014	200
		v2	blink	29	4	0	0	0	0	0.670803	6.74E-05	4.02E-05	
		v2	gain	42	14	1	6	0	0	0.603276	9.74E-08	6.53E-08	57.5
		v2	okan	6	0	0	0	0	0				
		v2	phase_duration	14	5	3	6	4	2	0.802919	4.09E-08	9.46E-09	58.25
		v2	proportion	7	4	0	1	1	0	0.706335	1.15E-08	7.40E-09	63
		v2	rate	7	5	4	6	4	3	0.778081	6.84E-06	1.58E-06	202.75
		v2	saccades	8	6	5	7	5	4	0.773579	0.000641	0.000208	205
		v2	velocity	24	10	2	2	2	0	0.581005	2.89E-05	2.71E-05	71
		v2	vergence	6	3	0	1	0	0	0.643936	3.60E-06	6.10E-06	64.5
	Optokinetic Nystagmus	v1	gain	8	3	0	4	3	0	0.603496	0.03591	0.019026	70
		v1	vergence	2	1	0	0	0	0	0.710177	0.001297	0.000369	
Saccades	Saccades	v1	error	1	1	0	1	1	0	0.749682	0.000337	2.22E-05	202
		v1	latency	1	1	1	1	1	1	0.886141	0.190288	0.019921	216
		v1	vergence	2	1	1	0	0	0	0.866318	9.35E-05	1.06E-05	
	Saccades2D	v2	blink	44	1	0	0	0	0	0.731032	9.53E-09	9.62E-10	
		v2	error	105	12	0	25	5	0	0.588797	8.89E-07	5.65E-07	200
		v2	gain	105	20	2	20	5	0	0.602864	3.42E-08	2.33E-08	200.5
		v2	latency	35	17	6	24	15	6	0.708958	3.95E-09	1.22E-09	216
		v2	missed	2	0	0	2	0	0				254
		v2	overshoot	8	1	0	2	0	0	0.507074	3.65E-07	4.01E-08	188.25
		v2	pso	18	1	0	0	0	0	0.5	1.44E-08	0	
		v2	saccades	72	14	2	33	5	1	0.612558	0.001193	0.00058	203
		v2	successful_main_saccade	8	0	0	4	0	0				197.5
		v2	velocity	21	10	2	5	2	1	0.709023	0.009071	0.001756	61
		v2	vergence	2	0	0	0	0	0				
SmoothPursuit_Fixed	Smooth Pursuit (Head Fixed)	v1	error	1	0	0	0	0	0				
		v1	saccades	1	0	0	0	0	0				
		v1	vergence	2	0	0	0	0	0				
	SmoothPursuitFixed	v2	amplitude	90	23	4	34	7	2	0.612271	1.13E-05	3.08E-06	204.5
		v2	blink	20	0	0	1	0	0				178
		v2	error	45	1	0	34	1	0	0.566881	6.15E-06	1.92E-06	204
		v2	gain	45	36	22	36	34	22	0.785886	3.00E-07	5.75E-08	51
		v2	inter_saccadic_interval	10	6	1	4	2	0	0.694732	6.99E-07	1.02E-07	53.5
		v2	phase_duration	15	3	0	0	0	0	0.566609	2.07E-10	2.19E-10	
		v2	saccades	10	9	6	4	4	4	0.767612	0.001141	0.000247	203.75
		v2	velocity	20	5	0	12	4	0	0.636222	0.002592	0.000802	210.5
		v2	vergence	2	0	0	1	0	0				67.5
	SmoothPursuitFixedH	v2	amplitude	90	21	2	35	10	2	0.605349	1.08E-05	2.82E-06	204
		v2	blink	20	1	0	4	0	0	0.733333	4.59E-07	0	179.25
		v2	error	45	6	1	27	3	0	0.657827	1.35E-06	1.43E-06	206
		v2	gain	45	30	22	27	27	20	0.830517	2.48E-07	7.23E-08	53
		v2	inter_saccadic_interval	10	3	0	3	1	0	0.545402	6.45E-07	2.58E-07	48
		v2	phase_duration	15	1	0	0	0	0	0.508468	4.27E-10	2.56E-10	
		v2	saccades	10	8	2	5	3	2	0.620823	0.00014	5.63E-05	213.5
		v2	velocity	20	10	0	12	6	0	0.552738	0.0083	0.001789	197.5
		v2	vergence	2	1	0	0	0	0	0.654019	8.05E-06	1.43E-06	
SmoothPursuit_Free	Smooth Pursuit (Head Free)	v1	error	1	0	0	0	0	0				
		v1	saccades	1	0	0	0	0	0				
		v1	vergence	2	2	0	0	0	0	0.68134	0.000291	6.31E-05	
SpontaneousNystagmus	GazeEvoked	v2	blink	24	0	0	0	0	0				
		v2	consecutive	24	4	2	3	1	0	0.75376	4.14E-07	0	193
		v2	distance_to_tg	72	15	2	8	1	0	0.6413	2.76E-05	8.98E-06	203.5
		v2	drift	32	0	0	2	0	0				116.75
		v2	phase_duration	24	5	1	2	1	0	0.586479	4.30E-05	3.58E-06	37
		v2	proportion	24	13	7	1	0	0	0.77073	3.62E-07	8.54E-08	201
		v2	saccades	32	4	4	2	0	0	0.869764	0.000227	7.38E-06	196.5
		v2	single_nystagmus	48	6	3	6	2	0	0.729686	5.63E-07	0	187.25
		v2	velocity	104	37	7	2	0	0	0.650672	1.12E-05	3.14E-06	123.5

		v2	vergence	22	2	1	1	0	0	0.792437	3.01E-05	7.46E-06	206
	Spontaneous Nystagmus	v1	drift	3	1	0	0	0	0	0.5225	4.11E-08	2.06E-08	
		v1	velocity	3	0	0	3	0	0				223
	SpontaneousNystagmus	v2	blink	6	0	0	1	0	0				186
		v2	consecutive	6	0	0	2	0	0				175.75
		v2	distance_to_tg	18	4	2	1	0	0	0.79926	1.45E-05	1.63E-06	198
		v2	drift	8	0	0	0	0	0				
		v2	phase_duration	6	0	0	2	0	0				38.5
		v2	proportion	6	2	0	5	1	0	0.606962	5.38E-08	9.62E-08	206.5
		v2	saccades	8	0	0	8	0	0				184.75
		v2	single_nystagmus	12	0	0	3	0	0				178.5
		v2	velocity	26	13	6	0	0	0	0.642005	3.63E-06	2.49E-06	
		v2	vergence	10	5	3	0	0	0	0.838571	3.68E-05	4.85E-06	

CONCLUSION

- V2 protocols show an overall significant improvement when compared to V1, in terms of reliability and validity scores. More specifically, V2 has several additional metrics with medium-excellent reliability, as well as the metrics that significantly differed during the Adverse condition from the Normal (see Figure 10, Figure 11, and Table 8).
- In V2, Antisaccades, Saccades, and Smooth Pursuit protocols had a greater impact than the others in differentiating the Normal condition from the Adverse (see Figure 11).
- In V2, VOR-related protocols (AVVOR, VOMS, and VORinDark) as well as antisaccades and Smooth Pursuit, had the most “good-excellent” reliable metrics, in the mentioned order (see Figure 6).
- The highly reliable metrics can guide us in choosing which metrics, among the many generated in V2, should be chosen to be displayed to the user.

REFERENCES

[1] Weir, J. P. (2005). Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. The Journal of Strength & Conditioning Research, 19(1), 231-240.

[2] Liljequist, D., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation–A discussion and demonstration of basic features. PloS one, 14(7), e0219854.

[3] Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of chiropractic medicine, 15(2), 155-163.