



Research paper

Reliability and reproducibility of visual e-norms plateau identification

Nicholas E. Earle^{a,b,c}, Joe F. Jabre^{b,c,*}^a Department of Neurology, Clínica Santa María, Santiago, Chile^b Corporación de Investigación de Neurología de Santiago (CINSAN), Santiago, Chile^c Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

ARTICLE INFO

Article history:

Received 27 June 2019

Received in revised form 17 March 2020

Accepted 26 March 2020

Available online 6 April 2020

Keywords:

E-norms

Normal values

Nerve conduction

EMG

ABSTRACT

Objective: To derive normal values from a lab's own diagnostic studies, the e-norms method relies on the proper identification of the e-norms plateau to derive descriptive statistics of the variable under study. This work was undertaken to compare the inter and intra-rater reliability of visual identification of the plateau by different raters analyzing laboratory nerve conduction study data.

Methods: Twenty raters were asked to visually identify the inflection points delineating an e-norms plateau to derive the Mean value of nerve conduction study laboratory data while blinded to the parameter they were analyzing. After a delay of 1–3 months, the same raters were asked to repeat some of the e-norms plateaus identification to assess delayed intra-rater reproducibility.

Results: Mean values derived from the identified plateau data were compared between raters (inter-rater) using a two factor ANOVA without replication. For the immediate inter-rater no statistically significant difference was found between the Means obtained by the different raters. For the delayed intra-rater, differences were found between raters.

Conclusions: This study suggests that visual identification of the e-norms plateau inflection point is reliable between raters but more research is needed to assess reproducibility for the same raters.

Significance: E-norms is a promising method for deriving reference values using data that is available in most electrophysiology laboratories.

© 2020 International Federation of Clinical Neurophysiology. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Despite great technical, genetic, and molecular strides witnessed by medicine in the past decades, to this day, the great majority of physicians still interpret their diagnostic studies using normal values collected by other labs; from different cohorts; using different equipment and techniques; and sometimes obtained decades earlier. This is not completely of their own choice however. The process of collecting one's normal values to appropriately interpret their diagnostic studies is arduous, time consuming, and requires navigating multiple ethical and legal hurdles well before getting started with the project. This arduous task becomes nearly impossible with all the challenges it poses in cohorts such as infants and children where, compounding all these hurdles, normal values change with age.

Indeed in a work entitled “Computing normative ranges without recruiting healthy subjects” (Yaar, 1997) Yaar argues that “if the existence or nonexistence of symptoms is the only criterion on which normative data are based, then the latter is not needed

at all; the diagnosis would have higher accuracy if based on the symptoms themselves.” The reader can find a brief discussion of normal versus abnormal in a recently published letter to the editor in Clinical Neurophysiology. (Jabre, 2018)

A technique one of the authors (JFJ) developed and refer to as the extrapolated norms or e-norms method (Jabre et al., 2015), allows a lab to extract its own normal values from diagnostic studies performed in their own laboratory, using their equipment and methods, on their own cohorts.

The technique does so by sorting a laboratory variable in ascending order, plotting a cumulative distribution curve of the data, and identifying a flat or plateau part of the curve where consecutive data points vary little from one value to the next.

Our work, and that of others to date (Nandedkar et al., 2015; Pitt and Jabre, 2017; Jabre et al., 2016; Zaccarini et al., 2016; Verma and Lin, 2017; Stålberg et al., 2019) has shown that data that lies in the plateau part of the curve represents that lab's own normal values and compares favorably with published normal data obtained by epidemiological studies. A detailed description of the e-norms method can be found in the 2015 published article on the method (Jabre et al., 2015).

* Corresponding author at: 167 N Poinsettia Pl, Los Angeles, CA 90036, USA.

E-mail address: info@teleemg.com (J.F. Jabre).

The purpose of this work was to compare the reliability and reproducibility of visual identification of the e-norms plateau when different raters from various backgrounds analyzed nerve conduction study data, while blinded to the parameters they were studying.

2. Methods

The study was conducted using an encrypted and secure e-norms web application (Jabre, 2018) allowing users to securely upload nerve conduction data utilizing an anonymous data import tool for e-norms analysis.

Twenty (20) raters were asked to identify the plateau part of an e-norms cumulative distribution curve by visually identifying the left and right inflection points of the curve that delineate the plateau; on the left, when the ascending part of the curve begins to flatten out; and on the right, when the flattened part of the curve begins to ascend. The visual e-norms evaluation method uses three criteria to identify the plateau as follows:

1. Visual identification of the inverted S curve (left Y Axis) left and right inflection points.
2. Identification of the lowest first order difference between successive data points (right Y Axis). The segments of the curve with the lowest differences correspond to the plateau.
3. Use a P3 (Polynomial 3 line) to help smooth out the curves in the inverted S curve to make these inflection points easier to identify.

The raters in this study were recruited from a diverse pool of hospital workers that included a secretary, two nurses, two EMG technicians, one radiologist, eight neurology residents, one neuromuscular fellow, and five neuromuscular specialists. All raters, except one of the authors (NE) were blinded for the nerve conduction parameter they were analyzing.

Prior to starting the study, raters who were not familiar with the e-norms method received a brief explanation of the method and the criteria for visual identification of the plateau. This introductory explanation was done on nerve conduction data that were not subsequently used in the study.

A total of 393 upper and 284 lower limb nerve conduction studies were included for analysis. During studies, skin temperature was recorded at the palm and lateral malleolus and efforts were made to maintain it at a minimum of 32 °C and 30 °C respectively. For inter-rater reliability raters identified the plateau for the 6 parameters described in Table 1. Intra-rater reproducibility was assessed after a delay between 1 and 3 months by asking the same

raters (without a repeat explanation of the e-norms plateau identification rules) to complete a new set of 7 plateaus, 4 of which were repeated from the first trial in a random order. This work was started after local ethics committee review and approval.

Fig. 1 shows a sample e-norms plot the raters used on the web application to identify the curve's inflection points that delineate the plateau between the cross marks. A first order derivative of the consecutive data points (value 2 - value 1, value 3 - value 2, value 4 - value 3 etc..) of the parameter under study is used to assist the user in identifying these inflection points.

3. Results

The Inter Rater two factor ANOVA without replication showed no significant difference between the nerve conduction Means derived by the raters from the e-norms plateau. Table 1 shows the means for each rater and nerve parameter. Table 2 demonstrates the results of this analysis. F value was below critical, rejecting the null hypothesis. 100% of participants performed the delayed intra-rater section of the study between 1 and 3 months after the first trial. None of the parameters achieved the primary outcome of showing no statistical difference.

4. Discussion

During our work with the e-norms method, we tested multiple algorithms in an attempt to automate the identification of the curve's inflection points and found that while some worked well, others often didn't. On the other hand, "visual" identification of the inflection points of the curve, aided by a first order derivative as can be seen in Fig. 1, proved on balance to be reliable and more easily reproducible with various types of data.

The present study was undertaken to determine whether or not our subjective impression of the visual plateau identification could be corroborated by statistical evidence. This study shows that immediate inter-rater reproducibility is good, even when raters are not health related, blinded to the parameters and identify the plateau with a simple set of rules. The delayed intra-rater reproducibility didn't show statistical significance and a number of factors could have contributed to this outcome. For one, an explanation of the set of rules was given only for the immediate part of the study, not in the delayed identification. For another, the raters included people who were not familiar with the technique or its use, and for whom a repeat explanation of the set of rules would probably have been helpful. This might have produced some difficulty evoking the rules when the raters performed the delayed inflection selection and was evident by inspecting the pla-

Table 1
Inter-rater reproducibility.

Parameter	Rater 1	Rater 2	Rater 3	Rater 4	Rater 5	Rater 6	Rater 7	Rater 8	Rater 9	Rater 10	
Median motor CV (m/s)	57.2	56.2	56.8	56.5	56.4	56.0	56.8	57.5	56.7	55.7	
Ulnar CV forearm (m/s)	59.6	60.0	60.0	60.2	59.7	59.7	59.4	59.7	59.7	59.4	
Peroneal motor CV fibular head (m/s)	46.6	47.3	47.8	47.0	44.0	47.6	48.2	46.6	43.9	47.6	
Median motor amp (mV)	8.2	8.2	8.4	8.4	7.6	8.1	8.3	7.6	7.1	8.4	
Sural amp (uV)	19.2	22.7	19.7	23.5	23.3	18.9	22.5	16.5	22.5	18.9	
Median motor DL 16–49y (ms)	3.6	3.5	3.6	3.7	3.6	3.8	3.8	3.8	3.8	3.8	
Parameter (cont.)	Rater 11	Rater 12	Rater 13	Rater 14	Rater 15	Rater 16	Rater 17	Rater 18	Rater 19	Rater 20	Mean (SD)
Median motor CV (m/s)	57.0	56.5	56.5	56.0	55.7	56.3	56.4	55.3	56.8	56.3	56.4 (0.62)
Ulnar CV forearm (m/s)	60.3	59.3	59.8	59.8	59.9	60.3	59.4	59.8	59.7	59.7	59.7 (0.29)
Peroneal motor CV fibular head (m/s)	49.2	48.2	44.9	46.8	46.8	48.4	47.9	47.6	47.8	47.6	47.0 (1.40)
Median motor amp (mV)	8.2	8.3	8.0	8.0	8.2	8.2	8.5	7.6	8.1	8.2	8.0 (0.36)
Sural amp (uV)	22.5	25.0	19.9	23.4	20.9	20.0	24.7	20.3	24.3	19.6	21.4 (2.33)
Median motor DL 16–49y (ms)	3.8	3.6	3.7	3.7	3.7	3.8	3.7	3.9	3.8	3.6	3.7 (0.11)

For inter-rater reproducibility, mean values derived from data within the plateau identified by each rater were then compared using a two factor ANOVA without replication. m/s: meters per second. DL: distal latency. uV: microvolt. mV: millivolt.

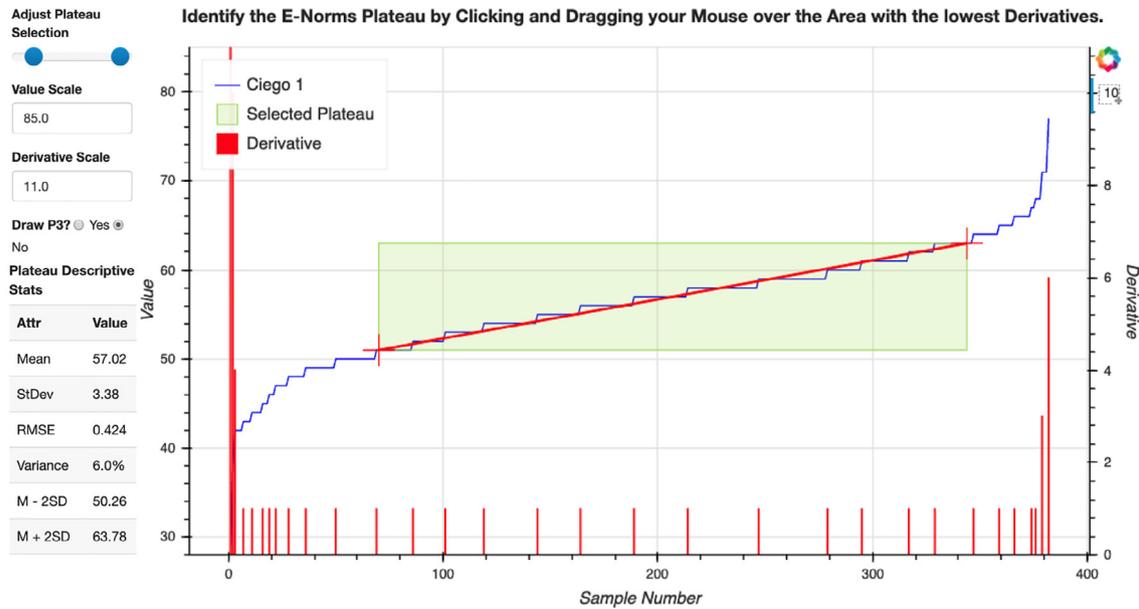


Fig. 1. Visual identification of the e-norms plateau. The rater logs on to the web app and securely uploads an Excel spreadsheet containing the nerve conduction data to be analyzed. The app automatically plots the e-norms curve (left Y axis) and the first derivative of the consecutive data points (right Y axis). The X axis represents the variable's rank. The rater then drags the mouse over the e-norms curve to delineate the plateau between the left and right inflection points. The plateau becomes highlighted by a rectangle seen in grey here. Once done, the program automatically calculates the descriptive statistics (seen to the left of the plot) of the data that lies within the plateau inside the rectangle.

Table 2

ANOVA: Two-Factor Without Replication Inter-Rater Mean Comparison.

Source of Variation	SS	df	MS	F	P-value	F critical
Raters (Rows)	9.8832425	19	0.52	1.36 *	0.17	1.70
E-norms Mean (Columns)	80803.3843	5	16160.68	42322.79	0.00	2.31
Error	36.2751225	95	0.38			
TOTAL	61705.94	119.00				

ANOVA: Two-Factor Without Replication Inter Rater Mean Comparison (Alpha 0.05). SS: Sum of Squares, df: degrees of freedom, MS: Mean Squares, F: F ratio.

* Note that the F value was below F critical, meaning no statistical differences was found between raters.

teau selection of some raters. Effort was taken to assess whether “real world” plateau identification was possible. Data with a wide variety of acceptable variances were selected and this resulted in inflection points that weren't obvious. This is one of the strengths of this study.

One can hardly argue that a method allowing providers to use normal values derived from their own patient referral pool cannot be more intrinsically appropriate for interpreting their diagnostic studies than normal values obtained from other labs, using different equipment and methods, and collected from different cohorts. To quote Kouri et al's., 1994 Seminal paper about normal values developed from data derived in this fashion (Kouri et al., 1994), “Healthy ambulatory individuals are not optimal references for hospitalized patients, because of differences in, e.g., body posture, physical activity, diet, with those prevailing in regular life. From that point of view, the best reference for a hospitalized patient is another patient not affected by the disease in question, but living under the same conditions as the patient whose laboratory result is being interpreted”. Deriving such normal values using the e-norms method has the added advantage of significantly reducing the time and effort required to collect them from months or years with traditional methods, to literally hours with the e-norms method.

But the idea of deriving normal values from one's own lab population is not new. Even though such concepts date back to at least 1963 by Hoffman (Hoffmann, 1963), and the eighties and nineties

(Kouri et al., 1994; Statland and Winkel, 1984; Sunderman, 1975) by the International Federation of Clinical Chemistry (IFCC), they still to this day face a great deal of challenge being adopted and universally accepted.

Since the publication of our work on the e-norms methodology in 2015, the method has been validated in over 20 Labs from 15 countries and numerous individual providers (Jabre 2018, unpublished observations from the e-norms web app) analyzing data ranging from nerve conduction studies, to motor unit potentials, neuromuscular jitter, brainstem auditory evoked potentials (BAEP), visual evoked potentials (VEP), somatosensory evoked potentials (SSEP), serum electrolytes, and acetylcholine receptor antibodies (AchRAb), with resulting values closely matching those obtained from epidemiological studies when available; and producing much needed normal values in cohorts for which no normal values were available. (Pitt and Jabre, 2017; Verma and Lin, 2017; Punga et al., 2019)

5. Conclusion

This work was undertaken to determine the reliability and reproducibility of visual pattern identification of the e-norms curve's inflection points delineating the plateau. The inter-rater ANOVA analysis revealed no statistically significant difference

between raters who were recruited from different backgrounds, and were blinded to the variable they were analyzing.

6. Ethical publication statement

The authors confirm that they have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Hoffmann, R.G., 1963. Statistics in the practice of medicine. *JAMA* 185, 864–873.
- Jabre, J., Kouyoumdjian, J., Moreira, Ferreira V., 2016. Carpal tunnel syndrome electrophysiological parameters acquired using the e-norms technique. *Muscle Nerve* 54, 546.
- Jabre J. enorms. <https://enorms.com/>. Accessed 24 October 2018.
- Jabre, J.F., Pitt, M.C., Deeb, J., Chui, K.K.H., 2015. E-norms: a method to extrapolate reference values from a laboratory population. *J. Clin. Neurophysiol.* 32, 265–270.
- Jabre, J.F., 2018b. Do you define the limits of normalcy from looking at the patient or the healthy subject? – An e-norms reply. *Clin. Neurophysiol.* 129, 1517–1518.
- Kouri, T., Kairisto, V., Virtanen, A., Uusipaikka, E., Rajamäki, A., Finneman, H., et al., 1994. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin. Chem.* 40, 2209–2215.
- Nandedkar, S., Sanders, D., Kouyoumdjian, L., Barkhaus, P., Stalberg, E., 2015. Confirming normal jitter limits using the extrapolated normal method. *Muscle Nerve* 52, S30.
- Pitt, M.C., Jabre, J.F., 2017. Determining jitter values in the very young by use of the e-norms methodology. *Muscle Nerve* 55, 51–54.
- Punga, A.R., Jabre, J.F., Amandusson, A., 2019. Facing the challenges of electrodiagnostic studies in the very elderly (>80 years) population. *Clin. Neurophysiol.* 130, 1091–1097.
- Stålberg, E., van Dijk, H., Falck, B., Kimura, J., Neuwirth, C., Pitt, M., et al., 2019. Standards for quantification of EMG and neurography. *Clin. Neurophysiol.* 130, 1688–1729.
- Statland, B.E., Winkel, P., 1984. Reference values: are they useful? *Clin. Lab. Med.* 4, 61–70.
- Sunderman, F.W., 1975. Current concepts of 'normal values', 'reference values', and 'discrimination values', in clinical chemistry. *Clin. Chem.* 21, 1873–1877.
- Verma, S., Lin, J., 2017. Jitter values in infants. *Muscle Nerve* 5, 140.
- Yaar, I., 1997. Computing normative ranges without recruiting normal subjects. *Muscle Nerve* 20, 1510–1514.
- Zaccarini, C., Zheng, C., Jabre, J., Jiang, J., Weber, R., Zhu, Y., 2016. Validation of the e-norms method to derive reference values of the Flexor Carpi Radialis H-Reflex latency. *Muscle Nerve* 54, 564.