Viewpoints

Estimating Change in Tremor Amplitude Using Clinical Ratings: Recommendations for Clinical Trials

Rodger J. Elble*

Department of Neurology, Southern Illinois University School of Medicine, Springfield, IL, USA

Abstract

Tremor rating scales are the standard method for assessing tremor severity and clinical change due to treatment or disease progression. However, ratings and their changes are difficult to interpret without knowing the relationship between ratings and tremor amplitude (displacement or angular rotation), and the computation of percentage change in ratings relative to baseline is misleading because of the ordinal nature of these scales. For example, a reduction in tremor from rating 2 to rating 1 (0–4 scale) should not be interpreted as a 50% reduction in tremor amplitude, nor should a reduction in rating 4 to rating 3 be interpreted as a 25% reduction in tremor. Studies from several laboratories have found a logarithmic relationship between tremor ratings R and tremor amplitude T, measured with a motion transducer: $\log T = \alpha \cdot R + \beta$, where $\alpha \approx 0.5$, $\beta \approx -2$, and log is base 10. This relationship is consistent with the Weber–Fechner law of psychophysics, and from this equation, the fractional change in tremor amplitude for a given change in clinical ratings is derived: $(T_f - T_i)/T_i = 10^{\alpha(R_f - R_i)} - 1$, where the subscripts i and f refer to the initial and final values. For a 0–4 scale and $\alpha = 0.5$, a 1-point reduction in tremor ratings is roughly a 68% reduction in tremor amplitude, regardless of the baseline tremor rating (e.g., 2 or 4). Similarly, a 2-point reduction is roughly a 90% reduction in tremor amplitude. These Weber–Fechner equations should be used in clinical trials for computing and interpreting change in tremor, assessed with clinical ratings.

*To whom correspondence should be addressed. E-mail: rodger.elble@gmail.com

Introduction

Modern transducers respond to energy from a physical system (i.e., stimulus) and produce an electrical signal (usually voltage) that is linearly proportional to the stimulus. Good transducers are not biased by the initial conditions. For example, a linear accelerometer can detect tremor fluctuations in inertial acceleration even though the transducer is subjected to the acceleration of gravity.[1] Similarly, a good force transducer is capable of measuring small forces (e.g., 10-g force) even if the initial force is much larger (e.g., 1-kg force). By contrast, human perception depends on the initial conditions, as shown by the German physiologist Ernst Heinrich Weber in the mid-1800s.[2] The addition of 10 g to an existing mass of 1 kg in a human hand is not perceived because human perception is strongly influenced by

the initial conditions and is therefore non-linear. The purpose of this Viewpoint is to review how the psychophysics of human perception affects the design and interpretation of clinical rating scales for tremor.

Weber–Fechner relationship for tremor

Weber found that the "just noticeable difference" or smallest discernible change ΔI in a sensory stimulus I is proportional to the initial stimulus intensity: $\Delta I = K \cdot I$, where K is a constant (i.e., Weber's constant).[3,4] Gustav Theodor Fechner, a student of Weber, reasoned that the increments in an ideal rating (i.e., perception) scale of stimulus magnitude would correspond to a series of just noticeable differences, starting at the threshold of perception I_0. Fechner derived a mathematical relationship between human perception P and stimulus intensity I: $P = C \cdot \log_{10}(I)$, where C is an empirically determined
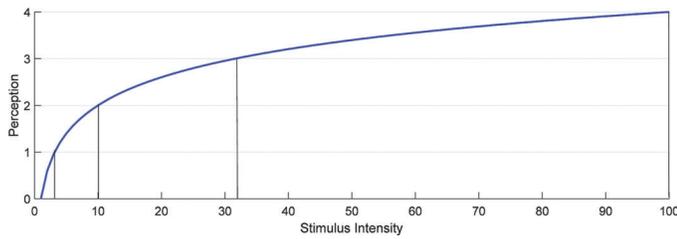
Figure 1. The Fechner Equation $P = C \cdot \log_{10} I$ Is Plotted for $C = 2$. Note the exponential increase in stimulus intensity (I) for each increment in stimulus perception (P).

constant or coefficient (Figure 1). The Fechner equation follows mathematically from Weber's law, and the logarithmic relationship between stimulus and perception is commonly referred to as the Weber–Fechner law of psychophysics. Exceptions to this law have been emphasized and debated extensively,[2] but data from many psychophysical studies have been consistent with this law.[3,4]

The Weber–Fechner law predicts that tremor ratings R will be proportional to the logarithm of tremor amplitude T (displacement or angular rotation), measured with a motion transducer. This relationship was found in early studies of tremor,[5,6] and subsequent studies from several laboratories confirmed a Weber–Fechner relationship for tremor, as expressed in equation 1.[7–11] This relationship also holds when tremor amplitude is derived quantitatively from pen-and-paper drawings of spirals that are scanned into a computer.[12]

$$\log T = \alpha \cdot R + \beta \qquad (1)$$

Values of slope α and intercept β in equation 1 are determined empirically.[7–11] The correlation between $\log T$ and R is best estimated when tremor rating and transducer measurement are performed simultaneously because tremor varies considerably over short intervals of time (i.e., minutes). For a 0–4 rating, estimates of α generally range from 0.4 to 0.6, and β from −1 to −3. These estimates came from studies of upper limb rest and action tremor and head action tremor, using accelerometers, gyroscopes, and digitizing tablets.[6,7,10] Estimates of α and β for tremor in other anatomical locations have not been computed. A value of 0.4 for α can be assumed when conservative estimates of tremor amplitude are desired, and higher values of alpha (e.g., 0.5 or 0.6) can be used for liberal estimates.[13,14]

Equation 1 is not limited to 5-point 0–4 ratings. It also applies to 0–3 ratings and to 0–10 ratings,[7,9] and it is theoretically applicable to any number of rating increments.[13,14] The value α_n for a 0–n rating can be estimated from α_4 for a 0–4 rating using equation 2.[13,14] For example, Elble and Ellenbogen[10] estimated α to be 0.6 for 0–4 ratings of tremor in Archimedes spirals. Haubenberger et al.[9] found α to be 0.19 to 0.24 for the 0–10 Bain and Findley scale. Using equation 2, one would have predicted a value of 0.6(4/10) = 0.24 for the Bain and Findley scale.

$$\alpha_n = \alpha_4 \left(\frac{4}{n} \right) \qquad (2)$$

Estimating change in tremor amplitude from tremor ratings

Tremor rating scales are now used in virtually all clinical treatment trials, and the Fahn–Tolosa–Marín Clinical Rating Scale has been used most commonly.[15] Tremor is rated 0 to 4 in each item or task of this scale and in most other tremor scales.[15] A problem arises when investigators attempt to compute change because the ordinal representations of perceived tremor amplitude are not linear measures of tremor amplitude, as would be obtained with a motion transducer. Consequently, computing percentage change in tremor ratings is misleading.

For example, suppose patients A and B have baseline right upper limb postural tremor ratings of 2 and 4, and both patients experience a 1-point improvement with treatment. It has been common practice to express improvement as a percentage of the baseline score, and the percentage improvements for patients A and B would be 50% and 25%, respectively. However, the actual percentage change in tremor amplitude (as recorded with a linear motion transducer) is the same for both patients because the fractional or percentage change in tremor amplitude is given in equation 3, derived from equation 1 (the indices i and f denote the initial and final tremor assessments). The percentage change is obtained by multiplying equation 3 by 100.

$$\frac{T_f - T_i}{T_i} = 10^{\alpha(R_f - R_i)} - 1 \qquad (3)$$

Thus, the fractional change in tremor amplitude T is simply a function of the change in tremor rating R, not in the fractional change $(R_f - R_i)/R_i$. This is why clinical change in clinical ratings should be reported, not the fractional or percentage change. One can see from equation 3 that the percentage or fractional improvement in tremor amplitude was the same for patients A and B: 68% reduction or improvement, assuming $\alpha = 0.5$.

It is often assumed that the total score of a scale with N items (each item with 0–n ratings) is more linear, and percentage changes in total scores are common in the clinical literature. However, this assumption is incorrect, as shown in equation 4 for a scale with items 1, 2,…, N.

$$R_{total} = R_1 + R_2 + \ldots + R_N = C \cdot \log T_1$$
$$+ C \cdot \log T_2 + \ldots + C \cdot \log T_N \qquad (4)$$

Similarly, the sum of all changes in the scale items is given in equation 5.

$$\Delta R_{total} = (R_{1f} - R_{1i}) + (R_{2f} - R_{2i}) + \ldots + (R_{Nf} - R_{Ni})$$

$$\Delta R_{total} = C \cdot \log\left(T_{1f}/T_{1i} \right) + C \cdot \log\left(T_{2f}/T_{2i} \right)$$
$$+ \ldots + C \cdot \log\left(T_{Nf}/T_{Ni} \right) \qquad (5)$$

The ratios T_f / T_i will be comparable for each scale item if the scale items are strongly correlated, and equation 5 can then be reduced to equation 6.

$$\Delta R_{total} = C \cdot N \cdot \log\left(T_f / T_i\right) \qquad (6)$$

Note that $1/C$ is α in equation 1 for a 0-n rating. The following equations 7 and 8 are derived from equation 6.

$$\frac{T_f}{T_i} = 10^{(\alpha/N)\Delta R_{total}} \qquad (7)$$

$$\frac{T_f - T_i}{T_i} = 10^{(\alpha/N)\Delta R_{total}} - 1 \qquad (8)$$

In equation 8, $\Delta R_{total}/N$ is simply the average change in N 0–n ratings, so equation 8 is simply equation 3 for the average change in ratings. These relationships illustrate how fractional or percentage clinical change can be estimated using change in a single rating or change in the total score of a scale with multiple strongly correlated clinical ratings.

Example

In the pivotal trial of focused ultrasound thalamotomy for essential tremor, a subscale of eight upper limb items (maximum total score 32 points) from parts A and B of the Fahn–Tolosa–Marín Clinical Rating Scale was used as the primary outcome measure. The percentage improvement in mean score was reported as 47% at 3 months, decreasing from 18.1 ± 4.8 to 9.6 ± 5.1 (mean arithmetic change of –8.5 points). Most patients and physicians would not be impressed with a 47% reduction in tremor amplitude after focused ultrasound thalamotomy or any other form of functional neurosurgery. However, this change in tremor rating cannot be interpreted as a 47% reduction in tremor amplitude. In fact, the percentage reduction in tremor amplitude is actually much greater than 47%. Assuming $\alpha = 0.5$, the actual reduction in tremor amplitude can be estimated using equation 8, as shown in equation 9.

$$\frac{T_f - T_i}{T_i} = 10^{(0.5/8)(-8.5)} - 1 = 0.706 \ or \ 70.6\% \qquad (9)$$

Subsequent analysis of the data from this study revealed that one of the eight scale items, rest tremor, was poorly correlated with the other items. Rest tremor was usually scored as 0, and test–retest reliability was very low.[16] Not surprisingly, the change score for rest tremor was statistically 0.[16] Thus, only seven of the eight items in the primary outcome subscale in the focused ultrasound study actually contributed to the total score, and the fractional change in tremor amplitude is more accurately given in equation 10 with $N = 7$, not 8, resulting in an improvement of 75.3%. Note that if a value of 0.6 were assumed for α, the estimated percentage change would be 81.3%.

$$\frac{T_f - T_i}{T_i} = 10^{(0.5/7)(-8.5)} - 1 = 0.753 \ or \ 75.3\% \qquad (10)$$

This example illustrates the important requirement that items of a scale or subscale be strongly correlated when using equation 8 to
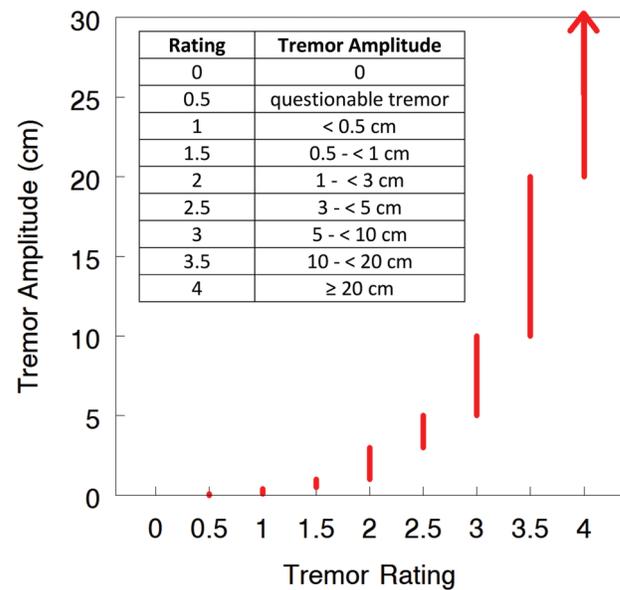

Rating	Tremor Amplitude
0	0
0.5	questionable tremor
1	< 0.5 cm
1.5	0.5 - < 1 cm
2	1 - < 3 cm
2.5	3 - < 5 cm
3	5 - < 10 cm
3.5	10 - < 20 cm
4	≥ 20 cm

Figure 2. Metric Anchors of the Essential Tremor Rating Assessment Scale for Upper Limb Tremor. The range of tremor amplitudes for each rating is illustrated in graphical and tabular formats.

estimate change in tremor amplitude. Poorly correlated or unreliable items, such as rest tremor in essential tremor, should be excluded. In the same study, postural tremor, wing-beating tremor, and finger–nose–finger tremor were rated using the Essential Tremor Rating Assessment Scale.[17] These three items were strongly correlated (Cronbach alpha = 0.83), and a mean reduction of 3.61 points occurred at 3 months. The fractional change in tremor estimated with this 12-point subscale is given in equation 11.

$$\frac{T_f - T_i}{T_i} = 10^{(0.5/3)(-3.61)} - 1 = 0.750 \ or \ 75.0\% \qquad (11)$$

Caveats

Many items of the Fahn–Tolosa–Marín Clinical Rating Scale[18] and the Essential Tremor Rating Assessment Scale[17] have metric anchors for ratings 0 to 4, and the defined range of amplitudes for each rating increases non-linearly (Figure 2). Therefore, one could argue that the Weber–Fechner relationship is by design rather than by psychophysics. However, the anchors for these scales were constructed with no attempt to fit R and T to a specific relationship. The fact that the ultimate relationship was Weber–Fechner speaks to the inherently logarithmic scaling of human perception in estimating tremor amplitude and in defining metric anchors for tremor ratings. The Bain and Findley spiral scale uses visual templates or examples to guide in the 0–10 rating of tremor amplitude,[19] but the relationship between R and T is still Weber–Fechner with a slope α_{10} that relates to the slope α_4 of 0–4 scales according to equation 2.[9,10] Moreover, the Fahn–Tolosa–Marín Clinical Rating Scale and the Essential Tremor Rating Assessment Scale spiral ratings have fairly crude descriptive anchors,

not metric anchors, and the relationship between R and T is still Weber–Fechner.[10] Thus, the Weber–Fechner relationship in equation 1 is clearly not by design.

Given the relatively simple physical quantity being assessed (tremor), one could reasonably consider the use of a visual analog scale instead of ordinal ratings. There is no published estimate of the mathematical relationship between a visual analog scale and transducer measures, but the data from Figure 1 of Knudsen et al.[20] suggest the relationship is logarithmic. Using a visual analog scale ranging from 0 to 30 cm, for example, it is easy to imagine the relative ease in distinguishing a 1-cm tremor from 2-cm tremor versus the difficulty of distinguishing 10-cm tremor from 11-cm tremor or 20-cm tremor from 21-cm tremor. Clearly, the use of a visual analog scale for tremor amplitude will be affected by Weber's law.

Conclusions

Linear measures of tremor with motion transducers correlate very well with clinical ratings; however, the relationship is logarithmic, not linear. The logarithmic relationship between tremor amplitude and tremor ratings is predicted by the Weber–Fechner law of psychophysics. Fractional or percentage change in tremor ratings is misleading because it does not reflect the true fractional change in tremor amplitude. Arithmetic differences in clinical ratings should be reported in clinical trials, not fractional or percentage changes relative to baseline. The fractional or percentage change in tremor amplitude should be estimated using the Weber–Fechner relationship between tremor ratings and amplitude.[13,14]

References

1. Elble RJ, McNames J. Using portable transducers to measure tremor severity. *Tremor Other Hyperkinet Mov* 2016;6. doi: 10.7916/D8DR2VCC

2. Gescheider GA. Psychophysics: the fundamentals. 3rd ed. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers; 1997: p 1–14.

3. Nieder A, Miller EK. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 2003;37:149–157. doi: 10.1016/S0896-6273(02)01144-3

4. Dehaene S. The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends Cogn Sci* 2003;7:145–147. doi: 10.1016/S1364-6613 (03)00055-X

5. Elble RJ, Brilliant M, Leffler K, Higgins C. Quantification of essential tremor in writing and drawing. *Mov Disord* 1996;11:70–78. doi: 10.1002/mds. 870110113

6. Matsumoto JY, Dodick DW, Stevens LN, Newman RC, Caskey PE, Fjerstad W. Three-dimensional measurement of essential tremor. *Mov Disord* 1999;14:288–294. doi: 10.1002/1531-8257(199903)14:2<288::AID-MDS1014>3.0.CO;2-M

7. Elble RJ, Pullman SL, Matsumoto JY, Raethjen J, Deuschl G, Tintner R. Tremor amplitude is logarithmically related to 4- and 5-point tremor rating scales. *Brain* 2006;129:2660–2666. doi: 10.1093/brain/awl190

8. Lin PC, Chen KH, Yang BS, Chen YJ. A digital assessment system for evaluating kinetic tremor in essential tremor and Parkinson's disease. *BMC Neurol* 2018;18:25. doi: 10.1186/s12883-018-1027-2

9. Haubenberger D, Kalowitz D, Nahab FB, Toro C, Ippolito D, Luckenbaugh DA, et al. Validation of digital spiral analysis as outcome parameter for clinical trials in essential tremor. *Mov Disord* 2011;26:2073–2080. doi: 10.1002/mds.23808

10. Elble RJ, Ellenbogen A. Digitizing tablet and Fahn-Tolosa-Marin ratings of Archimedes spirals have comparable minimum detectable change in essential tremor. *Tremor Other Hyperkinet Mov* 2017;7. doi: 10.7916/D89S20H7

11. Giuffrida JP, Riley DE, Maddux BN, Heldman DA. Clinically deployable Kinesia technology for automated tremor assessment. *Mov Disord* 2009;24:723–730. doi: 10.1002/mds.22445

12. Kraus PH, Hoffmann A. Spiralometry: computerized assessment of tremor amplitude on the basis of spiral drawing. *Mov Disord* 2010;25:2164–2170. doi: 10.1002/mds.23193

13. Deuschl G, Raethjen J, Hellriegel H, Elble R. Treatment of patients with essential tremor. *Lancet Neurol* 2011;10:148–161. doi: 10.1016/S1474-4422 (10)70322-7

14. Elble RJ, Shih L, Cozzens JW. Surgical treatments for essential tremor. *Expert Rev Neurother* 2018:1–19.

15. Elble R, Bain P, Forjaz MJ, Haubenberger D, Testa C, Goetz CG, et al. Task force report: scales for screening and evaluating tremor: critique and recommendations. *Mov Disord* 2013;28:1793–1800. doi: 10.1002/mds.25648

16. Ondo W, Hashem V, LeWitt PA, Pahwa R, Shih L, Tarsy D, et al. Comparison of the Fahn-Tolosa-Marin Clinical Rating Scale and the Essential Tremor Rating Assessment Scale. *Mov Disord Clin Pract* 2018;5:60–65. doi: 10. 1002/mdc3.12560

17. Elble R, Comella C, Fahn S, Hallett M, Jankovic J, Juncos JL, et al. Reliability of a new scale for essential tremor. *Mov Disord* 2012;27:1567–1569. doi: 10.1002/mds.25162

18. Fahn S, Tolosa E, Marín C. Clinical rating scale for tremor. In: Jankovic J, Tolosa E, editors. Parkinson's disease and movement disorders. 2nd ed. Baltimore: Williams & Wilkins 1993; p 225–234.

19. Bain PG, Findley LJ. Assessing tremor severity: a clinical handbook. London: Smith-Gordon 1993.

20. Knudsen K, Lorenz D, Deuschl G. A clinical test for the alcohol sensitivity of essential tremor. *Mov Disord* 2011;26:2291–2295. doi: 10.1002/ mds.23846