Digitizing Tablet and Fahn–Tolosa–Marin Ratings of Archimedes Spirals have Comparable Minimum Detectable Change in Essential Tremor

Rodger J Elble1* & Aaron Ellenbogen2

1 Department of Neurology, Southern Illinois University School of Medicine, Springfield, IL, USA, 2 Michigan Institute for Neurological Disorders, Farmington Hills, MI, USA

Abstract

Background: Drawing Archimedes spirals is a popular and valid method of assessing action tremor in the upper limbs. We performed the first blinded comparison of Fahn–Tolosa–Marin (FTM) ratings and tablet measures of essential tremor to determine if a digitizing tablet is better than 0–4 ratings in detecting changes in essential tremor that exceed random variability in tremor amplitude.

Methods: The large and small spirals of FTM were drawn with each hand on two consecutive days by 14 men and four women (age 60 ± 8.7 years [mean ± SD]) with mild to severe essential tremor. The drawings were simultaneously digitized with a digitizing tablet. Tremor in each digitized drawing was computed with spectral analysis in an independent laboratory, blinded to the clinical ratings. The mean peak-to-peak tremor displacement (cm) in the four spirals and mean FTM ratings were compared statistically.

Results: Test–retest intraclass correlations (ICCs) (two-way random single measures, absolute agreement) were excellent for the FTM ratings (ICC 0.90, 95% CI 0.76–0.96) and tablet (ICC 0.97, 95% CI 0.91–0.99). Log10 tremor amplitude (T) and FTM were strongly correlated (logT=aFTM + β, a=0.6, β=−1.27, r=0.94). The minimum detectable change for the tablet and FTM were 51% and 67% of the initial assessment.

Discussion: Digitizing tablets are much more precise than clinical ratings, but this advantage is mitigated by the natural variability in tremor. Nevertheless, the digitizing tablet is a robust method of quantifying tremor that can be used in lieu of or in combination with clinical ratings.

Keywords: Essential tremor, rating scale, spirography, minimum detectable change

Introduction

Tremor rating scales provide crude, nonlinear, subjective assessments of tremor severity.1 The Fahn–Tolosa–Marín (FTM) tremor rating scale2 uses 0–4 anchors to assess tremor in drawings of Archimedes spirals. The Bain and Findley scale uses 0–10 anchors.3 Both scales have a strong logarithmic relationship with tremor amplitude measured with a digitizing tablet, consistent with the Weber–Fechner law of psychophysics.4–7

Digitizing tablets are capable of providing linear objective measures of tremor in writing and drawings.6,8–13 The Wacom Intuos 3 digitizing tablet (www.wacom.com) has been used most commonly and has an accuracy of ±0.25 mm and a sampling frequency of 100 samples/s, which are adequate for measuring the amplitude and frequency of a tremor that is visible to the unaided eye.8,9 Digitizing tablets are unable to detect pen motion when the pen tip is greater than 1 cm above the tablet surface and lack sufficient sensitivity to measure physiologic...
tremor. Thus, tablets, like clinical ratings, have ceiling and floor effects at the extremes of tremor amplitude.

The greater precision of tablets, relative to rating scales, enables one to detect much smaller changes in tremor amplitude. However, this advantage of tablets is diminished when random variability in tremor is large. Tablets measure random variability precisely, but a change in tremor must exceed random variability to be recognizable as a statistically significant change resulting from treatment or disease progression (minimum detectable change). Therefore, we sought to determine if a digitizing tablet is better than FTM part B spiral ratings in detecting changes in essential tremor that exceed random variability in tremor amplitude.

Methods

Twenty patients were enrolled in an unpublished open-label pharmacokinetic-pharmacodynamic study of sodium oxybate for the treatment of essential tremor, conducted by Jazz Pharmaceuticals. Details of the study design can be found on ClinicalTrials.gov (https://clinicaltrials.gov/ct2/show/study/NCT00596078). All patients participated after giving their informed written consent, approved by a local human subjects committee. The patients stopped all drugs for tremor at least five half-lives before the study. They also abstained from alcohol and caffeine for 48 hours. Fourteen men and four women (age 60 ± 8.7 years [mean ± SD]) with mild to severe essential tremor were examined by the same neurologist (A.L.E.). The paper and baseline 1 refers to the baseline data from study day 2, and baseline 2 refers to the data from study day 3. Baseline assessments from these two days were used to compute test–retest reliability (two-way random single measures intraclass correlations [ICCs], absolute agreement) and minimum detectable change (MDC) for the FTM spiral ratings and digitizing tablet measurements. MDC was computed using the formula MDC = SDd 1.96, where SDd is the standard deviation of the differences for the two measurements. For the grand average of the four FTM spiral ratings, MDC was expressed as a percentage of the baseline 1 mean (MDC%). The tablet data were positively skewed, so log10 transformation was performed to normalize these data. Note that SDd of log-transformed data is a ratio, and the MDC is therefore also a ratio; they are not log SDd and log MDC of the non-transformed data. MDC% of the log-transformed data is expressed as a percentage of the baseline geometric mean, using the equation MDC% = \(1 - \frac{1}{\text{MDC}}\) × 100. All statistical analyses were performed with MedCalc® statistical software (www.medcalc.org).

Results

The mean spiral ratings did not differ statistically from a normal distribution (D’Agostino–Pearson test: p=0.16 for baseline 1 data and p=0.13 for baseline 2 data). The tablet data were positively skewed and deviated significantly from a normal distribution (D’Agostino–Pearson test: p<0.0001 for baseline 1 and 2 data), so log10 transformation was performed to normalize these data, producing data that did not deviate significantly from a normal distribution (D’Agostino–Pearson test: p=0.25 for baseline 1 data and p=0.18 for baseline 2 data). The FTM ratings exhibited a floor effect in this patient population (Figure 1).

Figure 1. Distributions of Mean FTM Spiral Ratings and Log Tremor Amplitudes. Notched box and whisker plots of mean FTM spiral ratings and tablet measures are shown for the two baseline assessments. The baseline medians did not differ significantly for either measure of tremor severity. The FTM data exhibit a floor effect at 0. FTM, Fahn–Tolosa–Marin.
Regression analysis revealed a very strong linear Weber–Fechner relationship \( \log(T) = \alpha \cdot \text{FTM} + \beta \) between mean FTM spiral ratings and log mean tablet tremor amplitudes \( T \) (cm) for baseline 1 and baseline 2 measurements (Figure 2). Test–retest ICC was excellent for the FTM ratings (ICC 0.90, 95% CI 0.76 to 0.96) and log-transformed tablet measures (ICC 0.97, 95% CI 0.91 to 0.99).

The MDC for the digitizing tablet was 51% of the baseline geometric mean tremor amplitude (Table 1). The MDC for FTM was 90% of mean baseline spiral rating. However, FTM is a non-linear ordinal scale, so computing % change is not valid.\(^5\) Therefore, we converted FTM to actual tremor amplitude using the average slope and intercept in Figure 2 for the two regression equations relating FTM and log tremor amplitude (average slope \( \alpha = 0.6 \) and intercept \( \beta = -1.26 \)), and we computed the MDC% using the following equations derived by Elble and colleagues.\(^5\)

\[
\log T = \alpha \cdot \text{FTM} + \beta \\
\left( \frac{T_2 - T_1}{T_1} \right)_{100} = \left[ 10^{\alpha \cdot \text{FTM}_2 + \beta} - 1 \right]_{100} \\
\alpha = 0.6 \\
\text{FTM}_2 - \text{FTM}_1 = \text{MDC} = SDd \cdot 1.96 = -0.41 \cdot 1.96 = -0.80 \\
\left( \frac{T_2 - T_1}{T_1} \right)_{100} = \text{MDC}\% = (10^{0.6 \cdot \text{MDC}} - 1)_{100} = -67\%
\]

In the above calculations, SDd (-0.41) is the standard deviation of the differences between the baseline 1 and baseline 2 FTM scores. This estimate of MDC\% (67%) is similar to that found for the tablet.

Our estimates of MDC\% appear to be very robust and not dependent on normalization of the data. We computed the MDC\% of the tablet data without log transformation, using the baseline 1 mean (0.62 cm) and the SDd of the two baselines (0.20 cm). Using these values, the MDC\% is as follows:

\[
\text{MDC}\% = 0.20 \cdot 1.96 \cdot 100 / 0.62 = 63\%
\]

### Discussion

This is the first blinded study demonstrating a strong correlation between tablet and FTM spiral ratings, and this study provides much-needed estimates of test-retest reliability and MDC\% for tablet and FTM spiral ratings. We have shown that tablet measures are highly correlated with FTM tremor ratings. The test-retest ICC for the tablet was only marginally better than the FTM ICC. However, the FTM ICC probably would have been lower if different raters had been used to assess the two baselines because intra-rater reliability is much better than inter-rater reliability for tremor rating scales.\(^1\) Also, we compared the average ratings and amplitudes of 4 spirals, and this is known to reduce test–retest variability.\(^17\)
Haubenberger and colleagues\(^6\) found a strong (\(r>0.9\)) logarithmic relationship between tablet measures of tremor amplitude and the Bain and Findley 0–10 ratings of tremor in Archimedes spirals, and the slope of this relationship was 0.2436. From this relationship for 0–10 ratings, the slope for a 0–4 scale can be estimated as 0.2436\(\cdot(10/4)=0.601\).\(^{16}\) which is what we found in this study. Thus, the logarithmic relationship between tablet measures and tremor ratings is robust, regardless of the scale used.

There is no published evidence that the Bain and Findley 0–10 ratings are more sensitive to change than FTM 0–4 ratings.\(^1\) Hopfner and coworkers\(^10\) estimated the minimum detectable change of the Bain and Findley scale to be 2 points, or 20% of the maximum rating 10. We found the MDC of the mean FTM spiral rating to be 0.8 points, which is 20% of the maximum rating 4.

Detectable change in essential tremor is limited by the considerable natural variability of tremor amplitude over time. The variability in tremor amplitude is so great that the MDC (the smallest detectable change exceeding random variability) of the digitizing tablet is similar to the MDC of the FTM 0–4 ratings and the Bain and Findley 0–10 ratings. Digitizing tablets are much more precise than clinical ratings, but this advantage is mitigated by the natural variability in tremor.

Digitizing tablets have potential floor and ceiling effects. They cannot measure tremor that is not visible because their accuracy is roughly \(\pm 0.25\) mm. They also cannot record tremor that is so severe that the pen tip does not remain within 1 cm of the tablet surface. However, FTM ratings had an obvious floor effect in our patient population, but the tablet exhibited no floor effect for these patients. Tremor severity was not great enough in our patient cohort to examine a ceiling effect for the tablet vs. FTM.

Nevertheless, the digitizing tablet is clearly a valid and robust method of quantifying tremor. It can be used in lieu of, or in combination with, clinical ratings of tremor in Archimedes spirals. The tablet provides an accurate, clinically meaningful assessment of tremor amplitude. These devices cost a few hundred dollars, and free software for tremor analysis is available on the internet.\(^5,9\)

Our study has limitations. Our estimates of test–retest reliability and MDC% were computed using two baseline assessments at the same time of the day on two consecutive days, while controlling for tremor medications, caffeine, and alcohol. Random test–retest variability might be greater if the interval between assessments was longer and if the other controls were less stringent. Our results need to be confirmed using baseline assessments at intervals of 1 week and 1 month, which are common intervals of assessment in clinical trials.

References