Automated Quantitative Assessment of Archimedes Spirals in Essential Tremor: Development and Validation of a Regulatory-Grade Digital Biomarker Pipeline for the ESSENTIAL3 Program

Steven Petrou, PhD, on behalf of the ESSENTIAL3 Study Team

Cofounder and President of Research and Development, Praxis Precision Medicines

Disclosure: Steven Petrou is a current employee of Praxis Precision Medicines and is a Praxis stockholder

© 2026 American Academy of Neurology

Essential Tremor: High burden, characterized by significant functional disability, loss of independence and social isolation

7M

Americans affected

One of the most common movement disorders



>85%

Report functional disability

impacting daily activities like writing, eating, dressing, and social engagement



41%

Receive any treatment

Vast majority untreated or undertreated



<20%

Stay on therapy 2+ years

High discontinuation due to poor tolerability

- Propranolol, FDA-approved in 1986, remains the only approved therapy
- Limited efficacy, multiple contraindications in the elderly, and high discontinuation rates

Louis ED, Ottman R. *Tremor Other Hyperkinet Mov (N Y)*. 2014;4:259; Louis ED, et al. *Movement Disorders*. 2001;16(5):914-20; Elble RJ. *Curr Neurol Neurosci Rep*. 2013 Jun;13(6):353; Putzke JD, et al. *J Neurol Neurosurg Psychiatry*. 2006 Nov;77(11):1235-7; Vetterick C, et al. *Adv Ther*. 2022;39(12):5546-5567; Praxis Data on File

© 2026 American Academy of Neurology

American Academy of Neurology™

The Archimedes Spiral: 30 years of use in ET

A standardized fine motor task intended to approximate functional hand use and patient disability

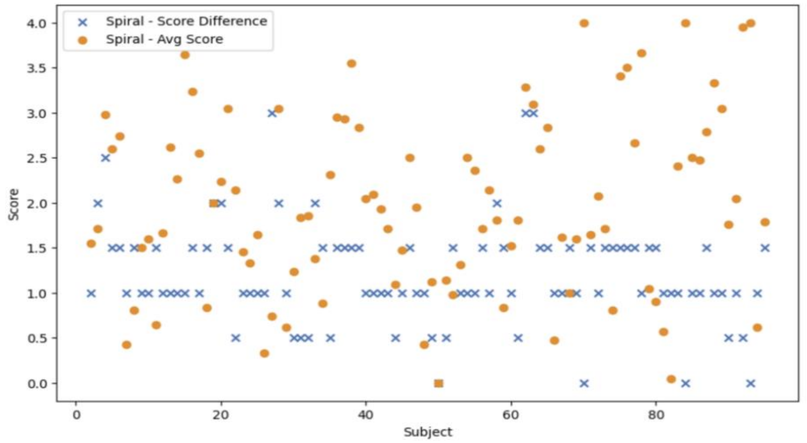
- Standardized visual rating scales since 1993 (Bain & Findley)
- Incorporated into clinical scales (FTM, TETRAS)
- Meant to capture kinetic tremor during a structured motor task
- Quick and believed to be repeatable; widely used in clinical trials



But… Visual ratings are sensitive to methodology, introducing variability beyond disease signal

- Ignores inconsistent nature of the pathology
- Multiple distinct scores assigned to the same spiral
- Small procedural differences can significantly affect scores
- Complete lack of agreement in scoring (6% of 64 samples assessed by 21 TRG specialists) and only at extremes
- Scoring varies depending on observation method
- Patient compensation can alter apparent severity

Same spiral drawings receive variable scores across raters



Avg Score – Average of scores assigned by all raters for each measure

Score Difference – Numerical difference between lowest and highest score amongst raters

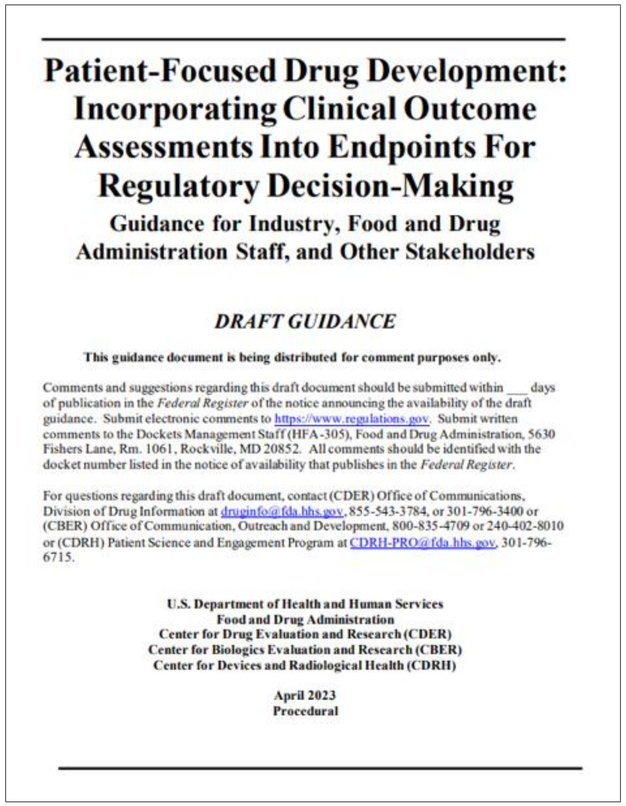
Bain et al. *J Neurol Neurosurg Psychiatry*. 1993; 56(8): 868–873; Ondo et al. *Tremor Other Hyperkinet Mov (N Y)*. 2020; 10: 36; Ondo et al. *Tremor Other Hyperkinet Mov (N Y)*. 2021; 11: 50; Praxis Data on File

FTM=Fahn–Tolosa–Marin Clinical Rating Scale for Tremor; TETRAS=The Essential Tremor Rating Assessment Scale; TRG=Tremor Research Group

© 2026 American Academy of Neurology

How should functional disability in ET be assessed?

Tremor is not constant, a spot measurement cannot capture what varies continuously. This is a sampling theorem problem, not a sensitivity problem



Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints For Regulatory Decision-Making

Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within ___ days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Office of Communications, Division of Drug Information at druginfo@fda.hhs.gov, 855-543-3784, or 301-796-3400 or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010 or (CDRH) Patient Science and Engagement Program at CDRH-PRO@fda.hhs.gov, 301-796-6715.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

April 2023
Procedural

MOTOR PERFORMANCE TASKS

e.g. Spirals, handwriting

- Brief; clinic-based tools
- Single 90-second sample of a fluctuating signal , violates basic sampling requirements
- Artificial conditions: time of day, anxiety, caffeine, sleep
- Cannot capture compensation, avoidance, or patient's lived experience over time
- Decades of use as endpoints have yielded zero new approved therapies

≠

FUNCTIONAL DISABILITY SCALES

e.g. TETRAS-ADL

- Patient-focused COA quantifying lived experience over time
- Retrospective measure of daily impairment; evaluable across entire trial treatment period
- Multiple functional domains capture meaningful impact: eg speaking, feeding, writing, hygiene
- Aligned to regulatory guidelines for clinically meaningful endpoints

FDA Guidance Document. 2023 FDA-2023-D-0026; Giroux et al. MDS 2023; Giroux et al. IAPRD 2023

American Academy of Neurology

© 2026 American Academy of Neurology

Patients know what matters, and it is exactly what ADL captures

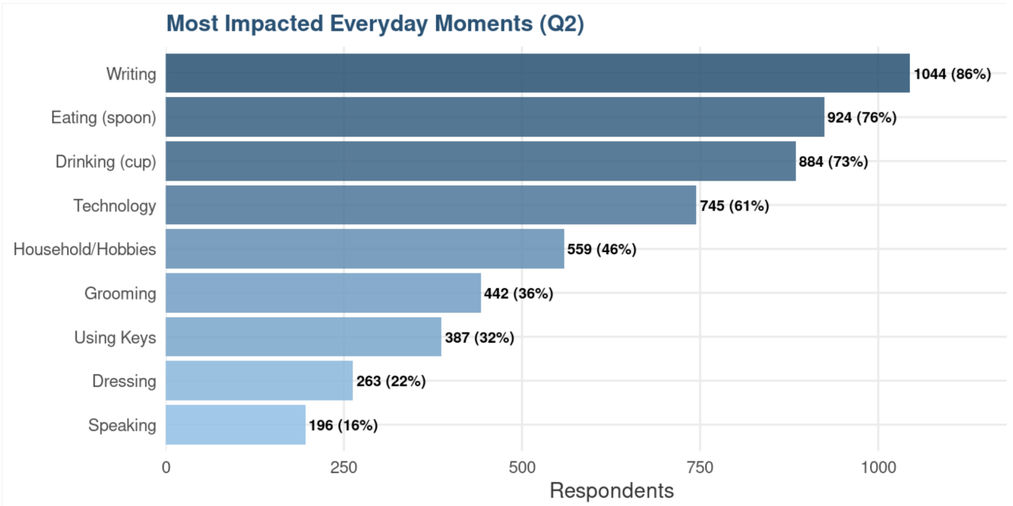
1,213 ET patients independently identified the daily activities most impacted by their tremor, they describe the ADL construct, not a spirals or performance measures…

The Sampling Problem

- Tremor amplitude fluctuates continuously, hour to hour, day to day
- No amount of analytical sophistication overcomes insufficient sampling, this is Nyquist, not noise
- The field's reliance on performance snapshots as endpoints has inadvertently thwarted drug advancement for 40 years
- Yet what patients say matters most – writing (86%), eating (76%), drinking (73%), technology (61%) – maps directly to the ADL construct, not to a spiral or performance measure
- A spiral captures a 90–second snapshot under artificial conditions. mADL captures a week of lived experience across 11 functional domains. One aligns with what patients say matters. The other does not

1,213
Total Responses

64%
See a Neurologist



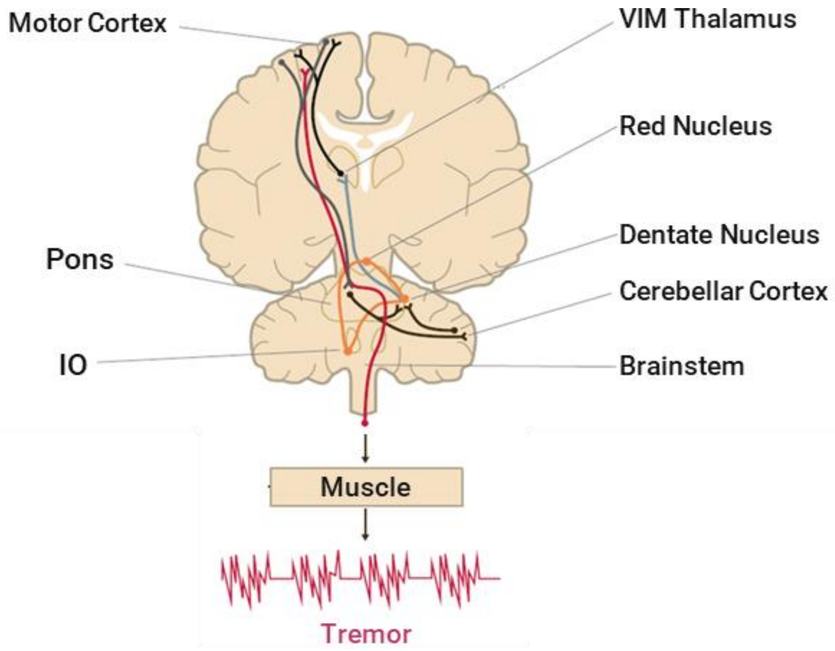
Patient Survey conducted by Praxis 2026

© 2026 American Academy of Neurology

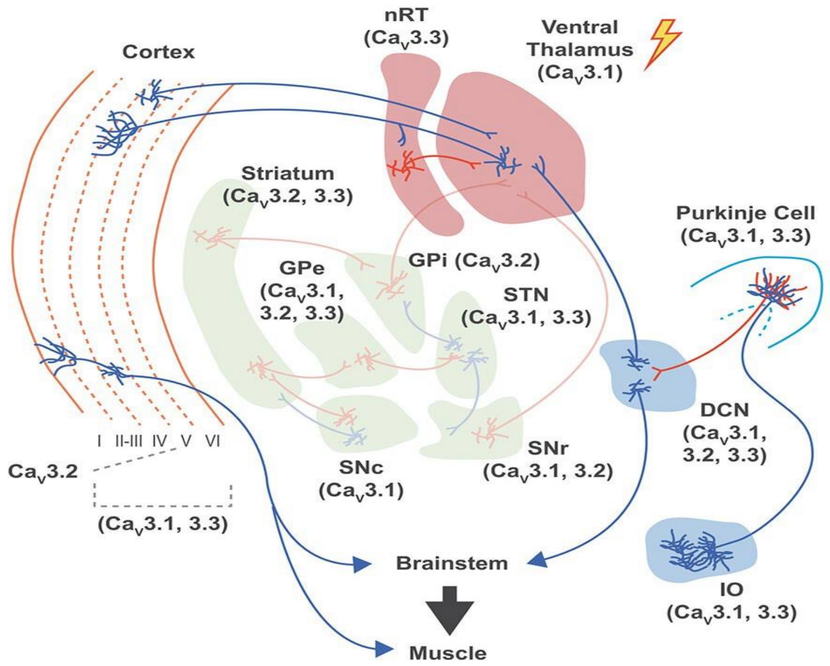
American Academy of Neurology™

Ulixacaltamide HCl: The first mechanism-based approach to treating ET at its neurophysiological source

Aberrant T-type calcium channel activity in the cerebello-thalamo-cortical circuit drives essential tremor



Targeting T-type calcium channels offers circuit-level normalization



Circuit images from Matthews et al. *Ann Clin Transl Neurol*. 2023; 10: 462–483

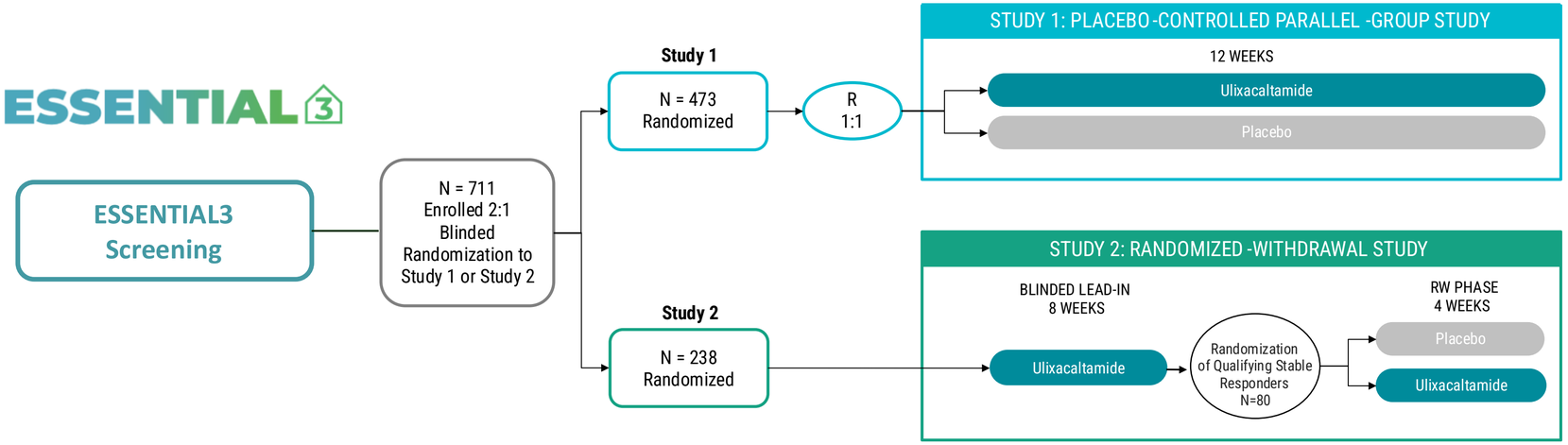
IO=inferior olive; vim=ventralis intermediate nucleus

American Academy of Neurology™

© 2026 American Academy of Neurology

ESSENTIAL3: The first positive Phase 3 program in Essential Tremor

- MONDAY POSTERS 17-002 & 17-003
- TUESDAY PLENARY PL5-001



Unified Recruitment, Screening and Visit Procedures

- Patients had comprehensive secondary pre-screening calls and medical records reviewed by MDs with movement disorder experience prior to referral to Central Coordinating Site
- 3 fellowship-trained movement disorder specialist PIs and >40 sub-investigators coordinated execution under standardized procedures, reducing inter-site variability
- Assessments (TETRAS-ADL, C/PGI) were conducted with participants at home, minimizing potential bias from clinic-based assessment

Eligibility Confirmation

- Comprehensive intake and screening (ET/medical history, medical records) performed across 4 visits in screening period
- Standardized neuro exams recorded and adjudicated by MD Specialists on the Eligibility Review Committee to review ET history and confirm ET diagnosis

Blinding Protections

- Blinded study allocation and treatment arm allocation (in both studies)
- Investigators, site staff and participants were blinded to enrollment criteria for threshold of disability and mADL11 stability, and threshold to qualify as a responder for RW phase in Study 2
- Contemporaneous, centralized data entry at all visits, with blinded transformation from TETRAS-ADL to mADL11 Total Score

Primary Endpoint:

- mADL11 change from baseline to Day 56

Key Secondary Endpoints:

- Rate of disease improvement (slope of mADL11 change) through Day 84
- PGI-C at Day 56
- CGI-S at Day 56

Exploratory Endpoints:

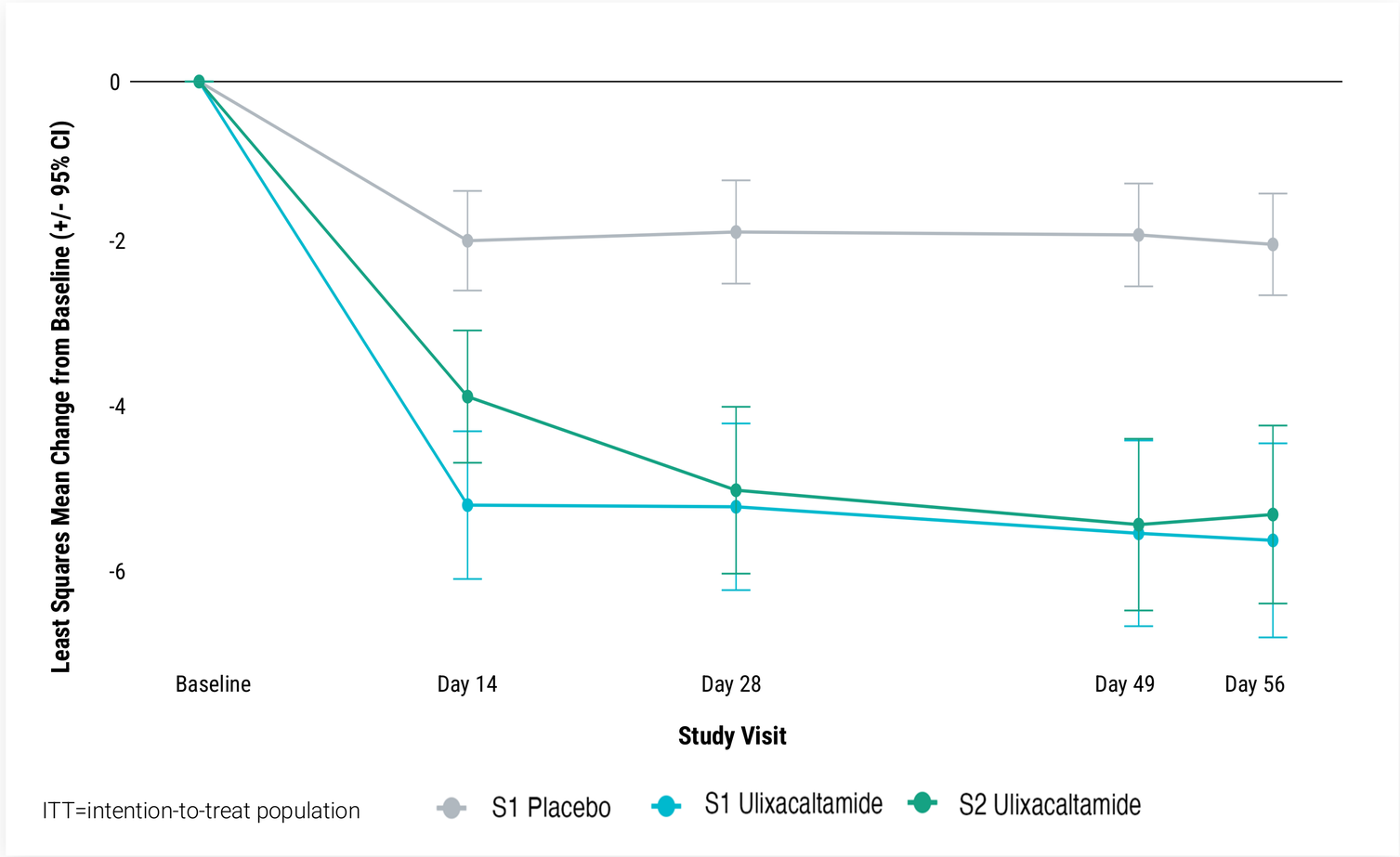
- mADL11 change across timepoints
- TETRAS-ADL and PGI/CGI-S across timepoints
- Sensitivity and subgroup analyses
- **Change in dominant-hand Archimedes spiral assessments**

C/PGI=clinical and patient global impression; ET=essential tremor; TETRAS-ADL=The Essential Tremor Rating Assessment Scale – Activities of Daily Living subscale

© 2026 American Academy of Neurology

Rapid onset and durable clinically meaningful and statistically significant effect observed in TETRAS–ADL (ITT) for both studies

1. *Speaking*
2. *Feeding with a spoon*
3. *Drinking from a glass*
4. *Hygiene*
5. *Dressing*
6. *Pouring*
7. *Carrying food trays, plates, or similar items*
8. *Using keys*
9. *Writing*
10. *Working*
11. *Overall disability with most affected task*
12. *Social impact*



American Academy of Neurology™

© 2026 American Academy of Neurology

Pipeline inputs and extraction of 13 objective spiral features

Paper-drawn Archimedes spirals are digitized and converted into 13 interpretable metrics spanning four tremor-driven spiral distortion domains

ACQUISITION & FEATURE EXTRACTION

Sample & protocol

- **459 Study 1 subjects** (225 ulixacaltamide, 234 placebo); essential tremor population
- Dominant hand, large and small Archimedes spirals, standardized instructions
- Drawn at baseline, Day 56 and Day 84; all features computed as change from baseline (CFB)

Digitization pipeline

- Paper spirals scanned, template removed, stroke skeletonized and parameterized
- 13 objective metrics extracted per spiral, per visit, no subjective rater scoring

13 METRICS ACROSS 4 DOMAINS

Spatial containment
Area, BBox Area (%), Conv Hull Area, Conv Hull (%)
How far does tremor push the stroke outside its expected boundary?

Stroke regularity
Curv Std, RMS Deviation, Width Std
How consistent is curvature, template fit and stroke width?

Temporal / frequency
HF Power Ratio, Line Length Ratio, Length
How much high-frequency oscillatory content is in the stroke path?

Complexity / disorder
Entropy, Fractal Dim, Segments
How fragmented and disordered is the overall trace?

ESSENTIAL 3

American Academy of Neurology

© 2026 American Academy of Neurology

Automated pipeline with five sequential analyses on the feature set

Staged framework. *ROC AUC is the common yardstick throughout, outcome swaps between arm, responder and spiral-type as the question changes*

FEATURE DISCRIMINATION & SEARCH

1 Individual feature discrimination

- **Per-feature, no composite.** Cohen's d + ROC AUC for each of the 13 features, ulixacaltamide vs placebo, at Day 56 and Day 84

2 Composite construction & exhaustive search

- **Full 13-feature composite first, not discriminatory,** motivating a subset search
- **377 subsets** (13 singles + 78 pairs + 286 triplets; capped at 3 for interpretability) scored by ROC AUC with treatment arm as the binary outcome

3 Spiral–mADL correlation

- **Spearman ρ** between each feature's CFB and mADL11 total CFB, does spiral change track real-world disability, not just arm?

RESPONDER PREDICTION & VALIDATION

4 Responder analysis

- **Responder =** mADL total CFB ≤ −3 points (Day 56, Day 84 separately)
- **Same 377-subset ROC AUC framework, outcome swapped** from treatment arm to responder status, does the spiral track clinical benefit, not just drug exposure?
- **k-means (k=2)** on the 13-feature space; ARI (adjusted rand index) vs arm and responder labels
- **Cross-validated logistic regression** on top features, held-out ROC AUC as a bias-corrected check on the exhaustive search's optimism

5 Big vs Small spiral head-to-head

- **Full pipeline run independently on big and small spirals,** then compared on ROC AUC, top-10 feature frequency and direction agreement
- If both spiral types pick the same features in the same direction, the signal is biological rather than an artifact of one drawing mode

ESSENTIAL 3

American Academy of Neurology™

© 2026 American Academy of Neurology

Spiral feature analysis struggles to detect pharmacodynamic signals
Coupling with mADL response required to show marginal predictive power

DAY 56 FEATURES (n = 313)

4 significant features *(p < 0.05)*

- HF Power Ratio d = 0.26
- Line Length Ratio d = 0.26
- Segments d = 0.25
- Curv Std d = 0.24

➢ *Significant but small effect sizes in features reflecting tremor-related irregularity, consistent with T-type calcium channel block*

DAY 84 FEATURES (n=267)

No significant features

- HF Power Ratio d = 0.25
- RMS Deviation d = 0.22
- Curv Std d = 0.21

➢ *Same directional trends as Day 56 → hints at a persistent signal*

PREDICTIVE PERFORMANCE

Best DAY 56 Composite: Conv Hull (%) + Curv Std + Width Std
- AUC = 0.61

Best DAY 84 Composite: Area + Conv Hull(%) + HF Power Ratio
- AUC = 0.60

➢ *Best composite models show modest discriminative ability*

CLINICAL RELEVANCE

Minimal Spiral–mADL11 correlation

- Best correlated feature: Entropy (ρ = 0.24) at Day 84
- Accounts for ~5.7% of variance in mADL11 change

➢ *Suggests partial overlap with limited clinical alignment*

RESPONDER ANALYSIS

mADL11 responder definition: CFB ≤−3 at Day 56 or Day 84

Best Day 84 composite: Conv Hull(%) + Entropy
- AUC = 0.64 (supervised)

Dominant single feature: Conv Hull (%)
- AUC = 0.62

Improved discrimination over time (treatment duration):
- Day 56 AUC = 0.59 → Day 84 AUC = 0.70

0.70
AUC
(Day 84, logistic regression)

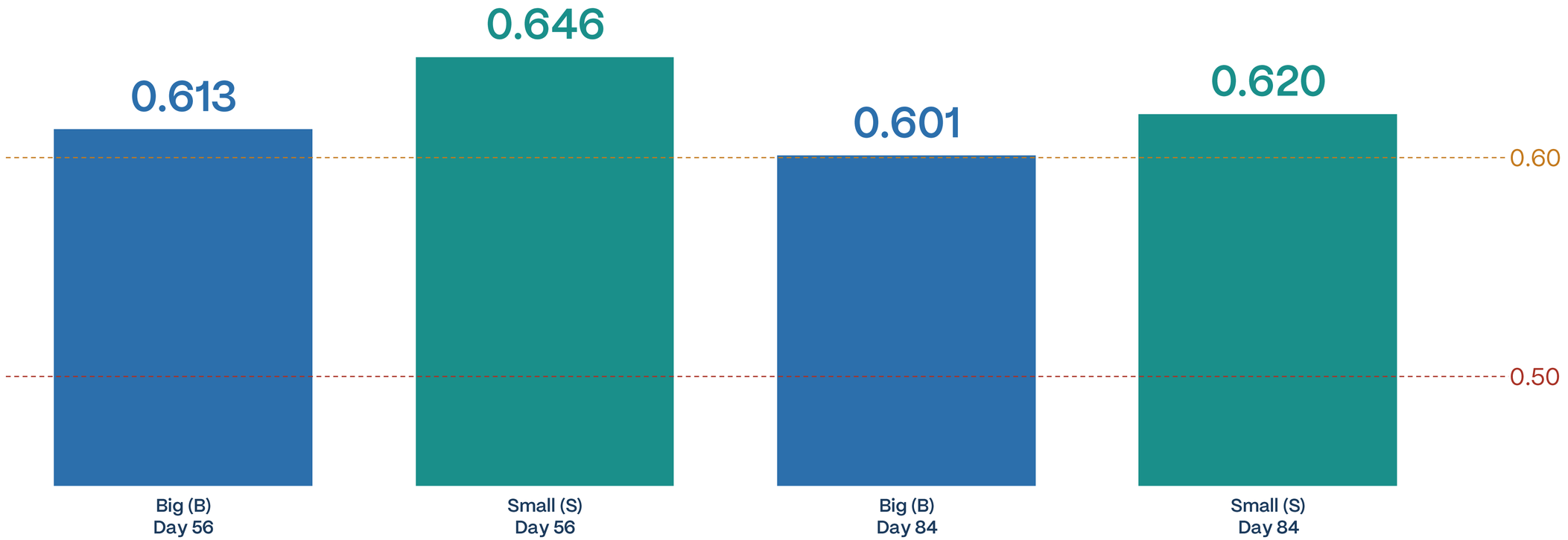
➢ *k-means ARI ≈ 0 at Day 56 and 84 → substantial overlap limits unsupervised separation, consistent with a diffuse signal*

Feature-level signal is modest and diffuse → becomes more detectable when aligned with clinically meaningful response

© 2026 American Academy of Neurology

Big vs. Small Spirals: Replication

Identical analysis pipeline applied independently to large and small spiral drawings shows weak but replicable pharmacodynamic signal across both spiral types. Not only replicates but validates analysis pipeline.



▲ **Small spiral marginally outperforms big at both timepoints** (+0.033 AUC at Day 56, +0.019 at Day 84)
BBox Area (%) dominant in small spiral (appears in 9 of 10 top Day 56 subsets) — tremor excursions amplified by smaller template
All 13 features agree in direction at Day 84 (0/13 disagreements) —minor PD signal is biologically replicated across spiral types

American Academy of Neurology™

© 2026 American Academy of Neurology

Across multiple studies, spirals do not significantly inform functional outcomes

70%

ADL variance explained by TETRAS performance score (PS) alone

Hollý et al. 2024 (n = 40 ET patients)

- TETRAS PS explains 68.6% of ADL variance
- Adding spirals does not improve ADL prediction

Rudzinska et al. 2007 (n = 52 ET patients)

- Automated spiral measures show only moderate correlation with ADL
- r ≈ 0.56; explains ~31% of variance
- Moderate signal but substantially lower than clinical scales

Wang et al. 2025 (n = 1130 ET patients)

- Correlation with functional outcomes is weak and inconsistent
- Functional domains show the weakest alignment with spiral measures

American Academy of Neurology™

Hollý et al. *Mov Disord Clin Pract.* 2024; 11: 1582–1586; Rudzinksa et al. *Neurol Neurochir Pol.* 2007; 41(6):510–6; 1742–1755

© 2026 American Academy of Neurology

Conclusions

The first Phase 3 test of digitized spirals as an exploratory drug endpoint and why we should move beyond performance snapshots as outcome assessments…

- **Spirals and performance scale assessments have thwarted drug advancement, this is a sampling problem, not a sensitivity problem**
  - **Tremor is not constant: a single brief sample of a fluctuating signal cannot capture disability**
- **Spiral analysis is not the limiting factor**
  - **State-of-the-art analysis, while better than humans, does not overcome the core sampling problem**
- **Patients independently identify writing, eating, drinking, technology as most impacted, these are ADL domains assessed over a multiday look back, not spiral features measured over a minute or two period.**
- **The field must align endpoints with patient-relevant functional impact**

American Academy of Neurology™

© 2026 American Academy of Neurology

Acknowledgements

- Monique Giroux
- Sahaj Patel
- Kimberly Crimin
- Tyler Mackovic
- Kurtis Cruz
- Brian Hannigan
- Rachel Smith
- Marcio Souza
- Arie Nakhmani
- Matt Might
- On behalf of the ESSENTIAL3 Study Team

© 2026 American Academy of Neurology