PSILOGOS

Research Article

open-access

Medication-State Polarity and Item Stability in the Neuropsychiatric Fluctuation Scale: A Condition-Geometry Study

Liam Anderson

ABSTRACT

Neuropsychiatric symptoms in Parkinson's disease can reverse within a levodopa cycle, yet conventional psychometric summaries do not show whether the same questionnaire items retain a coherent directional organization across medication states. To determine whether the 20-item Neuropsychiatric Fluctuation Scale (NFS) forms a reproducible bipolar state axis and to identify items responsible for condition-specific instability. The analytic record contained a 20×3 matrix of principal-component loadings for Parkinson's disease OFF medication, Parkinson's disease ON medication, and healthy-control conditions, together with paired score summaries from 101 patients. A polarity-aware condition-geometry procedure quantified cosine similarity, angular separation, Pearson and Spearman association, exact sign concordance, item-bootstrap intervals, permutation probabilities, leave-one-item-out influence, loading-threshold sensitivity, and paired standardized effects. One hundred thousand item-bootstrap samples and 200,000 label permutations were generated from a fixed seed. Every item reversed sign from OFF to ON, producing a cosine similarity of -0.923 (95% item-bootstrap interval -0.980 to -0.843) and an angle of $157.4°$. ON and control loadings were closely aligned (cosine 0.960, 95% interval 0.932 to 0.982; angle $16.2°$), with identical signs for all 20 items. Items 5, 8, 13, and 19 were weak in one condition, but their joint removal strengthened all three pairwise similarities. The mean global NFS score moved from -15.0 in OFF to 16.8 in ON; the paired standardized transition was $d_z = 2.09$ (95% confidence interval 1.74–2.44). The plus-score increase and minus-score decrease contributed nearly equally to the global displacement. The NFS loading structure is best characterized as a stable bipolar axis whose orientation reverses with medication state. Its state discrimination is distributed across the item set rather than dependent on a small subset, although four items merit targeted investigation in future patient-level studies.

Keywords: Parkinson's disease; neuropsychiatric fluctuations; patient-reported outcome; levodopa; loading geometry; robustness analysis

Division of Psychiatry, Medical School, University of Western Australia, Perth, Western Australia, Australia
liam.anderson47@gmail.com

– **Published:** 15 December 2024

1. INTRODUCTION

Parkinson's disease is commonly recognized through bradykinesia, rigidity, tremor, and postural impairment, but its lived burden extends well beyond motor dysfunction. Mood, motivation, attention, fatigue, anxiety, autonomic symptoms, sleep disturbance, and impulse-control phenomena may vary over the day and may alter autonomy as strongly as the visible motor state. Population and clinic-based investigations have consequently moved non-motor symptoms from the periphery of assessment to a central position in Parkinson care.[1–3] The clinical difficulty is not simply that these manifestations are common. Their direction and intensity may change quickly in relation to dopaminergic exposure, creating short episodes that a retrospective inventory can average away.

Non-motor fluctuations describe within-person changes that accompany, precede, or follow medication-related motor variation. Neuropsychiatric fluctuations form an especially consequential part of this phenomenon. During an OFF period, a patient may experience anxiety, reduced initiative, mental slowing, sadness, or fatigue. During an ON period, the same person may regain interest and cognitive fluency, but some patients move further toward excessive well-being, overactivity, impulsivity, or hypomanic behavior. Detailed clinical series have shown that non-motor changes are frequent and that their severity is related, but not identical, to motor complications.[4–6] The partial dissociation matters: an apparently satisfactory motor response does not guarantee psychological equilibrium, and an assessment limited to movement may miss an important treatment cost.

The biological interpretation of these changes is consistent with medication-sensitive mesocorticolimbic signaling. Dopaminergic depletion can reduce reward responsiveness, vigor, and motivated behavior, whereas increasing dopaminergic stimulation can restore these functions or, in susceptible individuals, push them toward dysregulated reward seeking. Neuroimaging and clinical observations after subthalamic stimulation support an association between mesolimbic dopaminergic tone, apathy, withdrawal states, and hyperdopaminergic behavior.[7–10] This interpretation does not imply that every affective change is caused by dopamine or that symptom direction can be inferred from medication status alone. It does, however, make real-time measurement particularly valuable: the meaning of a symptom score depends on when it is obtained within the treatment cycle.

Several instruments identify wearing-off and non-motor change. The Non-Motor Symptoms Questionnaire broadens routine detection beyond movement;[11] the wearing-off questionnaires provide brief screening formats;[12–15] and dedicated inventories quantify non-motor fluctuation domains.[16–18] Behavioral measures also distinguish hypo- and hyperdopaminergic manifestations.[19,20] These tools answer related but different clinical questions. Some establish whether a symptom occurs, some ask whether it changes with medication, and others characterize behavior over a longer interval. A short scale completed at the moment of an OFF or ON examination serves a narrower purpose: it records the patient's current neuropsychiatric position before recall and temporal averaging obscure the transition.

The Neuropsychiatric Fluctuation Scale was created for this punctual purpose. Ten minus items represent anxiety, apathy, low mood, fatigue, attention difficulty, and related OFF experiences; ten plus items represent well-being, motivation, energy, communication, and related ON experiences. The minus and plus subscores each range from 0 to 30, while their difference creates a global score from -30 to $+30$. Initial work demonstrated feasibility and state-related associations with motor and behavioral measures.[21,22] This bipolar scoring rule is clinically intuitive, but its psychometric interpretation requires more than a reliability coefficient. A total can change because all items reverse coherently, because only a few highly weighted items dominate, or because condition-specific components use different subsets of the questionnaire. Those alternatives imply different degrees of interpretive stability.

Psychometric evaluation normally considers content validity, structural validity, internal consistency, reliability, measurement error, construct validity, responsiveness, and interpretability.[23–25] Internal consistency is useful only after the dimensional structure is plausible; a large coefficient can coexist with redundancy, multidimensionality, or condition-dependent item behavior.[26–28] Exploratory component or factor analysis addresses structure within a sample, but separate solutions are often summarized condition by condition. Such reporting can leave a clinically important question unanswered: do the loading patterns occupy the same direction in item space, the opposite direction, or unrelated directions?

Treating a vector of item loadings as a geometric object offers a direct answer. Cosine similarity measures directional agreement without being determined by the overall length of the vector. An angle near 0° denotes parallel organization, an angle near 180° denotes a common axis with reversed orientation, and an angle near 90° indicates little directional relation. Sign concordance complements the continuous measure by asking whether each item lies on the same side of the component. Leave-one-item-out analysis tests whether alignment depends on one influential item, while removal of condition-specific weak loadings tests whether uncertain elements conceal or create the geometric pattern. Item resampling provides an interval for the alignment itself. These procedures do not replace structural validity analysis; they interrogate cross-condition transportability of its loading solution.

The research question was therefore: *Does the NFS preserve a single, distributed bipolar item organization across OFF medication, ON medication, and non-*

Parkinsonian emotional conditions, and which items limit that stability? The primary expectation was complete or near-complete sign reversal between OFF and ON because the component orientation should follow the dominant neuropsychiatric state. A second expectation was positive alignment between ON and healthy-control loadings because both conditions place plus experiences on the positive side of the scale. A third expectation was that deletion of weak items would not materially reduce alignment, indicating that medication sensitivity is distributed across the questionnaire. Finally, the global, plus, and minus score transitions were standardized to connect loading geometry with the magnitude of observed within-patient change. This question, analytic record, and inferential emphasis differ from conventional validation work: the target is not whether the NFS has acceptable psychometric properties in one condition, but whether its item organization remains interpretable when the clinical state reverses.

2. MATERIALS AND METHODS

The numeric record comprised the 20 item loadings in three conditions, variance summaries for the first component, and paired NFS score summaries for 101 people with Parkinson's disease. This structured 60-loading dataset is designated NFS-CG60 in the accompanying files. Values were transcribed from Tables 1, 3, and 4 of Schmitt et al.[29] Participants had idiopathic Parkinson's disease defined by accepted clinical criteria;[30] motor state was quantified with the MDS-UPDRS part III.[31] The patient group had a mean disease duration of 9.6 years (SD 3.6), a mean levodopa-equivalent daily dose of 1307 mg (SD 467), an OFF-medication motor score of 45.2 (SD 16.08), and an ON-medication score of 16.7 (SD 8.9). The acute levodopa challenge averaged 282.9 mg (SD 76.7). Depression and apathy were characterized with inventories.[32,33] The accompanying control group contained 181 adults; its loading vector was calculated before an emotion-eliciting film procedure of the type validated by Schaefer et al.[34] The present analysis did not use participant-level records, item response distributions, or identifiers.

The three columns of the loading matrix were denoted $\mathbf{l}_{\mathrm{OFF}}$, $\mathbf{l}_{\mathrm{ON}}$, and $\mathbf{l}_{\mathrm{HC}}$, each with 20 ordered elements. The published item allocation defined items 1, 4, 5, 8, 9, 14, 15, 16, 19, and 20 as minus items and the remaining ten as plus items. Component loadings can be multiplied by -1 without changing the statistical solution. Accordingly, the analysis interpreted both direction and axis: negative similarity between OFF and ON represents an inverted orientation of a common symptom organization, not poor structural agreement.

The clinical encounter depicted in Figure 1 shows how medication state and self-reported neuropsychiatric experience can be recorded together. It is a synthetic illustration, contains no participant image, and is not observational evidence. Its purpose is to distinguish momentary administration from a long-interval retrospective questionnaire.

Figure 1. Momentary NFS assessment.

The quiet, face-to-face administration shown here emphasizes that the NFS records a current experience at a defined point in the levodopa cycle. The form is completed by the patient while the clinician independently preserves the timing and examination context.

Table 1. Clinical quantities in the analytic record.

Quantity	n	Mean (SD)	Range
Disease duration, years	101	9.6 (3.6)	4–21
Levodopa-equivalent daily dose, mg	101	1307.0 (467.0)	450–2715
MDS-UPDRS III, OFF	101	45.2 (16.08)	16–102
MDS-UPDRS III, ON	101	16.7 (8.9)	3–47
Acute levodopa challenge, mg	101	282.9 (76.7)	100–500
BDI-II	82	12.4 (7.5)	1–37
Starkstein apathy score	70	10.8 (4.9)	0–24

The quantities in Table 1 describe a clinically established fluctuating population rather than early or minimally treated disease. The 28.5-point mean motor improvement during the medication challenge verifies a large state contrast in the same assessment window. This context is essential for interpreting an inverted neuropsychiatric loading direction: the comparison concerns distinct dopaminergic conditions, not random administrations at an unspecified time.

2.1. Polarity-aware condition geometry

For conditions a and b, directional agreement was measured by cosine similarity,

$$C_{ab} = \frac{\mathbf{l}_a^{\mathsf{T}}\mathbf{l}_b}{\|\mathbf{l}_a\|_2\|\mathbf{l}_b\|_2}, \tag{1}$$

and converted to angular separation $\theta_{ab} = \cos^{-1}(C_{ab})$ in degrees. Equation 1 gives equal conceptual status to parallel and antiparallel structure: $C = 1$ denotes the same orientation, while $C = -1$ denotes the same axis with every weighted direction reversed. Pearson correlation assessed linear loading correspondence after centering, and Spearman correlation assessed ordinal correspondence. Correlations between absolute loading magnitudes were computed separately so that directional

agreement could not be mistaken for preservation of item salience.

Item signs were compared for all three condition pairs. Under a null model in which either sign is equally likely for every item, the two-sided probability of 20 concordant or 20 discordant signs is $2(0.5)^{20} = 1.91 \times 10^{-6}$. This exact result was considered descriptive evidence of collective polarity organization rather than an independence-based patient inference, because items within a questionnaire are correlated.

The loading map used a conventional absolute threshold of 0.30 to identify condition-specific weak elements. This threshold was not treated as a universal validity boundary. It served as a prespecified perturbation point for asking whether the overall axis survives removal of items that contribute little in at least one condition. The union of weak items was removed simultaneously, and the three cosine similarities were recomputed. The analysis also recorded Pearson correlation between absolute loading magnitudes. A low magnitude correlation alongside high signed alignment would mean that item importance changes even though polarity remains stable.

2.2. Resampling, influence, and permutation procedures

Uncertainty in each cosine similarity was estimated by resampling the 20 ordered item pairs with replacement 100,000 times. The 2.5th and 97.5th percentiles formed an item-bootstrap interval. The resampling unit was the item, not the patient, so the interval describes sensitivity to the composition of the questionnaire. Bootstrap logic is well suited to statistics with inconvenient analytic sampling distributions, provided that the resampling unit and inferential target are stated explicitly.[35] No claim of patient-population coverage is attached to these intervals.

Leave-one-item-out influence was assessed by recalculating C_{ab} after deleting item j,

$$\Delta_{j,ab} = C_{ab}^{(-j)} - C_{ab}. \tag{2}$$

The range of $C_{ab}^{(-j)}$ and the three largest values of $|\Delta_{j,ab}|$ were retained. A narrow range indicates that no single item controls the geometric conclusion. Label-permutation tests randomly reassigned one condition's 20 loading values to item positions while preserving its marginal distribution. Two hundred thousand permutations were generated for each pair, and the two-sided probability was calculated with the standard plus-one correction. Because a finite permutation run cannot resolve probabilities below approximately 5×10^{-6}, exact sign probabilities and Monte Carlo probabilities are reported separately.

2.3. Paired state-transition effects

Medication-associated score change was evaluated for the global score, plus subscore, and minus subscore. The available summaries provided $n = 101$, the OFF and ON means and standard deviations, and the paired t statistic. Cohen's paired standardized effect was therefore recovered as

$$d_z = \frac{|t|}{\sqrt{n}}. \tag{3}$$

The small-sample adjustment $J = 1 - 3/[4(n-1)-1]$ produced Hedges' $g_z = Jd_z$. Confidence limits for d_z were obtained by inverting the noncentral t distribution and dividing the noncentrality limits by $\sqrt{n}$, following recommendations to accompany standardized effects with interval estimates.[36,37] The standard deviation of the paired difference was recovered from $s_D = |\bar{D}|\sqrt{n}/|t|$ and used as an arithmetic consistency check.

The relative contribution of the plus and minus domains to the global movement was calculated from their absolute mean changes. Rounding in the tabulated subscore means produced a total of 31.7 points, compared with the 31.8-point global change; proportions were therefore based on 31.7. This decomposition is descriptive and does not assume that the two domains are statistically independent.

2.4. Reproducibility and reporting rules

All calculations were performed in Python using NumPy, pandas, and SciPy. The random-number seed was 20260814. The delivered project includes the comma-separated loading matrix, a machine-readable results file, the analysis script, and the figure-generation script. Numerical values in the text were rounded only after analysis. Similarities and correlations are shown to three decimals, angles to one decimal, and standardized effects to two decimals. Probability values from the finite permutation procedure are reported at their obtainable resolution. Emphasis was placed on effect magnitude, interval width, and sensitivity to item deletion rather than dichotomous interpretation of probability thresholds, consistent with contemporary statistical reporting guidance.[38]

3. RESULTS

3.1. Item-level polarity across conditions

The signed loading matrix displayed an unambiguous state pattern (Figure 2). All ten minus items loaded positively in OFF and negatively in ON. All ten plus items loaded negatively in OFF and positively in ON. The control vector used the same sign as ON for every item. Thus, OFF versus ON and OFF versus control showed 0 of 20 concordant signs, whereas ON versus control showed 20 of 20 concordant signs. The exact two-sided sign probability for either extreme was 1.91×10^{-6}.

The heat pattern in Figure 2 shows more than a difference in component sign convention. Item membership and clinical content agree with the observed reversal: minus experiences occupy one side during OFF, while plus experiences occupy that side during ON and in controls.

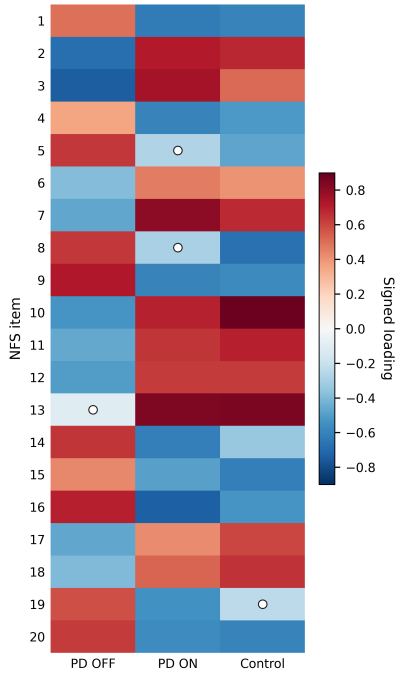

Figure 2. Polarity of NFS item loadings.

A purely arbitrary sign flip could be corrected by multiplying one vector by -1, but the clinical interpretation lies in which symptom domain becomes positive under each condition. Four hollow markers identify the only absolute loadings below 0.30: item 13 in OFF, items 5 and 8 in ON, and item 19 in controls. No item was weak in more than one condition.

The distribution also reveals condition-dependent salience. Item 13 had a loading of -0.108 in OFF but 0.837 in ON and 0.848 in controls, making it highly expressive outside OFF despite its weak OFF contribution. Items 5 and 8 were strong in OFF (0.639 and 0.634) but weak after medication (-0.269 and -0.293). Item 19 was appreciable in both patient conditions (0.582 and -0.543) but weak in controls (-0.236). Consequently, weakness in one condition cannot by itself justify item deletion; it may represent a symptom whose discriminative role is state dependent.

Table 2. NFS loading matrix used for condition geometry.

Item	OFF	ON	Control	Item	OFF	ON	Control
1	.499	-.626	-.604	11	-.458	.642	.701
2	-.682	.714	.680	12	-.505	.632	.626
3	-.740	.752	.513	13	-.108	.837	.848
4	.355	-.603	-.516	14	.640	-.615	-.338
5	.639	-.269	-.474	15	.434	-.486	-.618
6	-.387	.461	.402	16	.699	-.737	-.533
7	-.466	.802	.668	17	-.471	.426	.601
8	.634	-.293	-.668	18	-.398	.526	.649
9	.724	-.602	-.574	19	.582	-.543	-.236
10	-.531	.690	.881	20	.630	-.564	-.603

Direct verification of every sign and threshold classification is possible from Table 2. The largest OFF magnitude occurred for item 3 (-0.740), the largest ON magnitude for item 13 (0.837), and the largest control magnitude for item 10 (0.881). These maxima occur in different items, reinforcing the distinction between a stable axis and identical item weighting.

3.2. Directional similarity and angular separation

The OFF and ON vectors were strongly antiparallel: cosine similarity was -0.923, corresponding to an angle of $157.4°$. Pearson correlation was -0.941 and Spearman correlation was -0.785. The OFF and control vectors were similarly opposed (cosine -0.909, angle $155.3°$, Pearson $r = -0.931$, Spearman $\rho = -0.720$). ON and control were nearly parallel (cosine 0.960, angle $16.2°$, Pearson $r = 0.960$, Spearman $\rho = 0.798$). These three relations form a coherent geometry: OFF occupies the opposite end of a shared bipolar axis, while ON lies close to the control orientation.

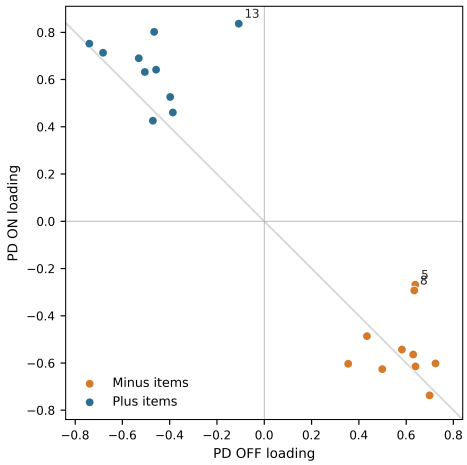

Figure 3. OFF-to-ON inversion of item loadings.

The points in Figure 3 cluster along the inverse-identity line, with plus and minus domains occupying opposite quadrants. Items 5, 8, and 13 deviate most visibly and were also among the most influential leave-one-item-out observations. Their deviations are not random sign failures: all remain in the expected quadrant. They instead indicate asymmetric changes in loading magnitude. This distinction explains why the rank correlation is weaker than the signed linear correlation. Medication preserves domain polarity more completely than it preserves the rank order of item importance.

Absolute loading magnitudes showed weak correspondence across conditions. The Pearson correlations for $|l|$ were -0.152 for OFF versus ON, -0.356 for OFF versus control, and 0.332 for ON versus control. A negative correlation between absolute values means that items weighted strongly in one condition tend, if anything, to be weaker in the other. The result rules out a simplistic interpretation in which a fixed group of high-loading items merely changes sign. The bipolar axis is distributed

across all items, but its local weighting adapts to the state.

The item-bootstrap intervals in Table 3 remained far from zero for every comparison. Their widths were modest despite resampling only 20 item pairs: 0.137 for OFF versus ON, 0.132 for OFF versus control, and 0.050 for ON versus control. The narrower ON-control interval indicates especially uniform alignment across the item set. Finite label-permutation probabilities were 1.0×10^{-5}, 2.0×10^{-5}, and 1.0×10^{-5}, respectively. These probabilities support the conclusion that the observed item correspondence is not reproduced by arbitrary reassignment of loading values to item labels.

3.3. Influence and weak-item sensitivity

Leave-one-item-out similarities varied from -0.957 to -0.916 for OFF versus ON, from -0.943 to -0.902 for OFF versus control, and from 0.956 to 0.969 for ON versus control. The complete-vector conclusions therefore survived every single-item deletion. Item 13 had the greatest influence on both comparisons involving OFF. Items 5 and 8 followed for OFF versus ON, while items 19 and 2 followed for OFF versus control. Items 8, 19, and 13 were most influential for ON versus control. Influence corresponded closely, but not perfectly, to the weak-item set because an item can affect cosine similarity through the contrast between two moderate loadings as well as through a single small loading.

Joint deletion of items 5, 8, 13, and 19 increased the magnitude of every condition relation. OFF versus ON changed from -0.923 to -0.979, OFF versus control from -0.909 to -0.950, and ON versus control from 0.960 to 0.975. Figure 4 places the full and reduced estimates on the same scale.

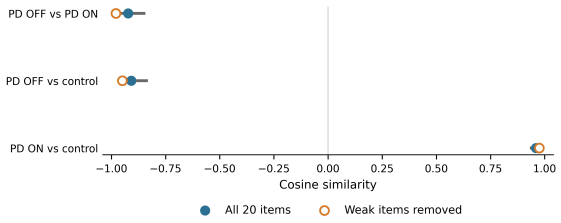

Figure 4. Stability of condition geometry.

The reduced estimates in Figure 4 lie beyond the full estimates in the expected direction, showing that weak items attenuate rather than generate the principal pattern. This finding answers the dependency part of the research question: no isolated question produces the state axis, and the four condition-specific weak items are not required for it. At the same time, stronger alignment after deletion is not evidence that a 16-item form is preferable. The weak items may capture clinically distinctive experiences that are expressed only in a particular state, and the available aggregate record cannot evaluate content loss, reliability of a shorter form, or individual classification.

3.4. Magnitude of within-patient score transitions

The mean global score rose by 31.8 points, from -15.0 (SD 10.6) in OFF to 16.8 (SD 9.4) in ON. The paired t value of 21.0 corresponded to $d_z = 2.09$ (95% confidence interval 1.74–2.44) and $g_z = 2.07$. The recovered standard deviation of the paired change was 15.22 points, consistent with the reported mean difference and test statistic. This effect is substantially larger than a conventional large standardized difference, confirming that the geometric reversal accompanies a major shift in patient-reported state.

The plus subscore increased from 4.6 (SD 4.4) to 21.0 (SD 6.3), a 16.4-point change with $d_z = 2.00$ (95% confidence interval 1.66–2.34). The minus subscore decreased from 19.5 (SD 7.2) to 4.2 (SD 4.4), a 15.3-point change with $d_z = 1.86$ (95% confidence interval 1.54–2.18). The absolute subscore changes accounted for 51.7% and 48.3% of their 31.7-point combined displacement. The global movement is therefore not produced mainly by recovery in one domain; increasing plus experiences and diminishing minus experiences contribute almost equally.

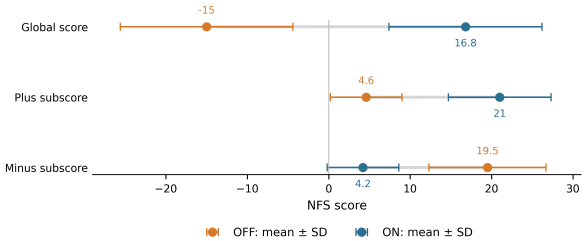

Figure 5. Medication-associated NFS score change.

The dumbbell positions in Figure 5 show that OFF and ON means exchange sides for the two subscores while the global score crosses zero. Standard-deviation whiskers are displayed to preserve the observed scale of heterogeneity, not as confidence intervals. The absence of raw paired observations prevents a distributional plot, but the paired standardized effects incorporate within-person change through the reported t statistics. Table 4 summarizes the effect estimates and makes their uncertainty explicit.

The effects in Table 4 should be interpreted as responsiveness to an acute medication contrast in a selected fluctuating population. They do not establish a minimally important difference for routine care. A large group mean transition can coexist with clinically important heterogeneity, including patients who improve only in minus symptoms, patients who develop marked plus symptoms, and patients with limited neuropsychiatric change despite motor improvement.

4. DISCUSSION

4.1. Principal interpretation

The analysis answers the paper's research question affirmatively. Across OFF, ON, and control conditions, the

Table 3. Pairwise condition geometry.

Comparison	Cosine	95% interval	Angle	Pearson r	Spearman ρ
OFF vs ON	-.923	[-.980, -.843]	157.4	-.941	-.785
OFF vs control	-.909	[-.964, -.832]	155.3	-.931	-.720
ON vs control	.960	[.932, .982]	16.2	.960	.798

Table 4. Paired NFS score transitions in 101 patients.

Score	OFF mean (SD)	ON mean (SD)	Change	d_z	95% CI
Global	-15.0 (10.6)	16.8 (9.4)	31.8	2.09	[1.74, 2.44]
Plus	4.6 (4.4)	21.0 (6.3)	16.4	2.00	[1.66, 2.34]
Minus	19.5 (7.2)	4.2 (4.4)	-15.3	1.86	[1.54, 2.18]

NFS preserves a distributed bipolar item organization. OFF is not geometrically unrelated to ON; it is positioned near the opposite end of the same axis. ON and control conditions are nearly parallel, while every OFF item reverses sign relative to both. No single item controls these relations, and deleting the four condition-specific weak items strengthens them. The score transition has a standardized magnitude greater than two within-person standard deviations, with nearly balanced contributions from increasing plus experiences and decreasing minus experiences.

This result refines the meaning of a one-component solution. A single component within each condition could, in principle, reflect different combinations of symptoms. The cross-condition geometry shows that the item signs retain a highly organized relation even as their magnitudes change. In practical terms, the NFS global score can be interpreted as position along a medication-sensitive neuropsychiatric continuum, provided that the administration condition is known. A value near the negative end represents dominance of minus experiences, and movement toward the positive end reflects both relief of those experiences and emergence of plus experiences.

The nearly parallel ON and control vectors require careful wording. They do not mean that treated patients and controls have equivalent symptom levels; group means differed in the parent clinical comparison, and the populations differed in age and depression. Parallel vectors mean that the items organize directionally in a similar way. The distinction between structure and level is fundamental. Two groups can use the same symptom axis while occupying different regions of it, just as two thermometers can preserve the same ordering while recording different temperatures.

The weak correlation of absolute loading magnitudes adds a second nuance. Medication reverses the direction of the overall organization, but it also redistributes which items are most expressive. Item 13 is the clearest example: nearly absent from the OFF component and prominent in ON and controls. Items 5 and 8 show the complementary pattern. A clinical instrument intended for repeated state assessment may legitimately include such conditional items. Requiring every item to load strongly in every state could remove the very content that distinguishes phases of the levodopa cycle.

4.2. Relation to Parkinson neuropsychiatry

The observed polarity accords with the clinical distinction between hypodopaminergic and hyperdopaminergic manifestations. OFF-related anxiety, sadness, reduced motivation, fatigue, and cognitive slowing have been described in levodopa-response studies and fluctuation cohorts, while ON periods can bring improved mood and attention or excessive activation.[5,6,39] These states are not simple mirror images at the biological level. Dopamine interacts with serotonin, noradrenaline, cognitive reserve, personality, disease distribution, medication history, and environmental context. Nevertheless, a scale can represent a clinically useful bipolar summary even when the underlying mechanisms are multifactorial.

The equal contribution of plus and minus subscore change is particularly relevant to treatment adjustment. If only the global score is inspected, a patient who moves from severe dysphoria to balanced well-being and a patient who moves from dysphoria to problematic overactivation might have similar positive shifts. Separate subscores retain the asymmetry needed for clinical interpretation. A treatment response should therefore be read in three coordinates: reduction in minus symptoms, increase in adaptive plus symptoms, and evidence that plus symptoms have crossed into behavior that is risky or distressing. The NFS can help identify the direction of change, but it cannot alone determine whether a positive state is optimal.

Current management recommendations for non-motor symptoms emphasize individualized review because evidence and treatment response differ across depression, anxiety, apathy, psychosis, sleep disturbance, and impulse-control disorders.[40] The geometric findings support repeated administration near clinically meaningful

time points: before a dose, near expected peak response, and during a patient-identified difficult period. Such timing is more informative than an unanchored daily score. The same principle applies around device-aided therapy or deep-brain stimulation, where motor benefit and motivational change may diverge.

4.3. Psychometric implications

The study illustrates why reliability, structure, and responsiveness should not be collapsed into a single claim. Internal consistency describes covariance among items, structural analysis describes latent organization, and responsiveness describes change under a relevant condition. A high alpha does not show that a scale retains meaning across states, and a large mean difference does not prove that the same items define the score in both states. The loading-geometry procedure connects these questions by directly comparing the item organization that accompanies the mean change.

The results also caution against treating component sign as disposable without documenting orientation. Algebraically, the signs of every loading can be reversed. Clinically, however, the chosen orientation should be tied to item content and score direction so that repeated analyses remain comparable. Reporting a fixed anchor, such as positive loading for plus items in ON, would prevent apparent contradictions across samples. Future confirmatory work could impose this orientation and test longitudinal measurement invariance at the ordinal item level.

The 0.30 threshold sensitivity analysis was deliberately subordinate to the continuous results. Loading cutoffs are conventions, not natural laws, and their meaning depends on sample size, item distribution, extraction method, and intended use.[41,42] Here the threshold identified four useful perturbations. Removing them made the geometry more extreme, but a shortening decision would require content-validity interviews, response-category analysis, reliability estimates suited to ordinal items, and independent replication. COSMIN criteria similarly place content validity and structural evidence before isolated numerical pruning.[24,43]

Coefficient choice will also matter in future patient-level work. The available analysis could not calculate alpha, omega, ordinal reliability, or conditional standard errors. Alpha can underestimate or overestimate reliability when tau-equivalence is violated, while omega coefficients require a defensible factor model.[27,28] Reporting several coefficients without checking their assumptions would not solve the problem. A stronger program would estimate an ordinal factor model for each state, evaluate invariance, calculate model-consistent reliability, and then quantify test-retest stability within a clinically stable medication window.

4.4. Clinical use and decision boundaries

At the bedside, the principal result supports interpreting the NFS as a state-position instrument rather than a trait diagnosis. A single administration documents the current balance of minus and plus experiences. Repeated administrations document direction and amplitude. The score should be timestamped relative to dose intake and interpreted alongside motor state, because a positive global value during an acute challenge has a different meaning from the same value during chronic overmedication.

The very large average transition should not be converted into a universal response threshold. The cohort consisted of people with established motor fluctuations undergoing a strong medication contrast, and the tails of individual change are unavailable. Clinical decision boundaries require patient-level receiver-operating analysis against an external state criterion, estimation of sensitivity and specificity in an independent cohort, and study of false-positive activation in people with depression, anxiety, bipolar-spectrum symptoms, or impulse-control disorders. Until those data exist, the NFS should complement, not replace, diagnostic interviewing and behavioral history.

The geometric stability nevertheless has immediate value for administration practice. Because no single item dominates the axis, occasional missing responses cannot be assumed harmless, but neither should one omitted item be expected to reverse the state interpretation. A formal missing-data rule still requires patient-level simulation. Likewise, the four weak items deserve attention during cognitive interviewing: respondents may interpret them differently across OFF, ON, and control contexts, or their conditional weakness may be exactly what the item was designed to capture.

5. CONCLUSION

The NFS does preserve a coherent medication-sensitive bipolar organization across the available conditions. All 20 items reverse direction from OFF to ON, ON aligns closely with the control orientation, and the relations remain strong after every single-item deletion and after joint removal of four condition-specific weak items. The accompanying 31.8-point global transition is very large and is divided almost equally between growth in plus experiences and reduction in minus experiences. The paper's research question is therefore answered by a distributed, robust state axis rather than by a small cluster of dominant questions. Clinically, this supports timed repeated administration of the full NFS to characterize the direction and amplitude of neuropsychiatric change, while separate plus and minus subscores remain necessary to distinguish recovery from excessive activation. Patient-level longitudinal validation is still required before individual decision thresholds or a shortened form can be justified.

GENERATIVE-IMAGE DISCLOSURE

The clinical illustration in Figure 1 was created with an image-generation model solely to depict a medication-state assessment encounter. It contains no participant, is not clinical documentation, and was not used as evidence. All quantitative figures were generated deterministically from the accompanying numeric record.

CONFLICT OF INTEREST

The author declares no conflict of interest.

BIBLIOGRAFIA / REFERENCES

1. Armstrong MJ, Rastgardani T, Gagliardi AR, Marras C. Impact of OFF periods on persons with Parkinson disease and care partners: a qualitative study. *Neurol Clin Pract*. 2021;11:e232–e238. doi:10.1212/CPJ.0000000000000921.
2. Breen KC, Drutyte G. Non-motor symptoms of Parkinson's disease: the patient's perspective. *J Neural Transm*. 2013;120:531–535. doi:10.1007/s00702-012-0928-2.
3. Schapira AHV, Chaudhuri KR, Jenner P. Non-motor features of Parkinson disease. *Nat Rev Neurosci*. 2017;18:435–450. doi:10.1038/nrn.2017.62.
4. Witjas T, Kaphan E, Azulay JP, et al. Nonmotor fluctuations in Parkinson's disease: frequent and disabling. *Neurology*. 2002;59:408–413. doi:10.1212/WNL.59.3.408.
5. Storch A, Schneider CB, Wolz M, et al. Nonmotor fluctuations in Parkinson disease: severity and correlation with motor complications. *Neurology*. 2013;80:800–809. doi:10.1212/WNL.0b013e318285c0ed.
6. Martínez-Fernández R, Schmitt E, Martinez-Martin P, Krack P. The hidden sister of motor fluctuations in Parkinson's disease: a review on nonmotor fluctuations. *Mov Disord*. 2016;31:1080–1094. doi:10.1002/mds.26731.
7. Thobois S, Ardouin C, Lhommée E, et al. Non-motor dopamine withdrawal syndrome after surgery for Parkinson's disease: predictors and underlying mesolimbic denervation. *Brain*. 2010;133:1111–1127. doi:10.1093/brain/awq032.
8. Lhommée E, Klinger H, Thobois S, et al. Subthalamic stimulation in Parkinson's disease: restoring the balance of motivated behaviours. *Brain*. 2012;135:1463–1477. doi:10.1093/brain/aws078.
9. Castrioto A, Thobois S, Carnicella S, Maillet A, Krack P. Emotional manifestations of PD: neurobiological basis. *Mov Disord*. 2016;31:1103–1113. doi:10.1002/mds.26587.
10. Béreau M, Fleury V, Bouthour W, Castrioto A, Lhommée E, Krack P. Hyperdopaminergic behavioral spectrum in Parkinson's disease: a review. *Rev Neurol*. 2018;174:653–663. doi:10.1016/j.neurol.2018.07.005.
11. Chaudhuri KR, Martinez-Martin P, Schapira AHV, et al. International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson's disease: the NMSQuest study. *Mov Disord*. 2006;21:916–923. doi:10.1002/mds.20844.
12. Stacy M, Bowron A, Guttman M, et al. Identification of motor and nonmotor wearing-off in Parkinson's disease: comparison of a patient questionnaire versus a clinician assessment. *Mov Disord*. 2005;20:726–733. doi:10.1002/mds.20383.
13. Stacy MA, Murphy JM, Greeley DR, et al. The sensitivity and specificity of the 9-item wearing-off questionnaire. *Parkinsonism Relat Disord*. 2008;14:205–212. doi:10.1016/j.parkreldis.2007.07.013.
14. Martinez-Martin P, Tolosa E, Hernandez B, Badia X. Validation of the "QUICK" questionnaire—a tool for diagnosis of wearing-off in patients with Parkinson's disease. *Mov Disord*. 2008;23:830–836. doi:10.1002/mds.21944.
15. Martinez-Martin P, Hernandez B. The Q10 questionnaire for detection of wearing-off phenomena in Parkinson's disease. *Parkinsonism Relat Disord*. 2012;18:382–385. doi:10.1016/j.parkreldis.2011.12.011.
16. Kleiner-Fisman G, Martine R, Lang AE, Stern MB. Development of a non-motor fluctuation assessment instrument for Parkinson disease. *Parkinsons Dis*. 2011;2011:292719. doi:10.4061/2011/292719.
17. Storch A, Schneider CB, Klingelhöfer L, et al. Quantitative assessment of non-motor fluctuations in Parkinson's disease using the Non-Motor Symptoms Scale. *J Neural Transm*. 2015;122:1673–1684. doi:10.1007/s00702-015-1437-x.
18. Faggianelli F, Loundou A, Baumstarck K, et al. Validation of a non-motor fluctuations questionnaire in Parkinson's disease. *Rev Neurol*. 2022;178:347–354. doi:10.1016/j.neurol.2021.06.013.
19. Ardouin C, Chéreau I, Llorca PM, et al. Assessment of hyper- and hypodopaminergic behaviors in Parkinson's disease. *Rev Neurol*. 2009;165:845–856. doi:10.1016/j.neurol.2009.06.003.
20. Rieu I, Martinez-Martin P, Pereira B, et al. International validation of a behavioral scale in Parkinson's disease without dementia. *Mov Disord*. 2015;30:705–713. doi:10.1002/mds.26223.
21. Schmitt E, Krack P, Castrioto A, et al. The Neuropsychiatric Fluctuations Scale for Parkinson's disease: a pilot study. *Mov Disord Clin Pract*. 2018;5:265–272. doi:10.1002/mdc3.12607.
22. Del Prete E, Schmitt E, Meoni S, et al. Do neuropsychiatric fluctuations temporally match motor fluctuations in Parkinson's disease? *Neurol Sci*. 2022;43:3641–3647. doi:10.1007/s10072-021-05833-8.
23. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539–549. doi:10.1007/s11136-010-9606-8.
24. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN

guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147–1157. doi:10.1007/s11136-018-1798-3.
25. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1171–1179. doi:10.1007/s11136-017-1765-4.
26. Sijtsma K. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*. 2009;74:107–120. doi:10.1007/s11336-008-9101-0.
27. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105:399–412. doi:10.1111/bjop.12046.
28. McNeish D. Thanks coefficient alpha, we'll take it from here. *Psychol Methods*. 2018;23:412–433. doi:10.1037/met0000144.
29. Schmitt E, Debu B, Castrioto A, et al. Fluctuations in Parkinson's disease and personalized medicine: bridging the gap with the Neuropsychiatric Fluctuation Scale. *Front Neurol*. 2023;14:1242484. doi:10.3389/fneur.2023.1242484.
30. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord*. 2015;30:1591–1601. doi:10.1002/mds.26424.
31. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale: scale presentation and clinimetric testing results. *Mov Disord*. 2008;23:2129–2170. doi:10.1002/mds.22340.
32. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories IA and II in psychiatric outpatients. *J Pers Assess*. 1996;67:588–597. doi:10.1207/s15327752jpa6703_13.
33. Starkstein SE, Mayberg HS, Preziosi TJ, et al. Reliability, validity, and clinical correlates of apathy in Parkinson's disease. *J Neuropsychiatry Clin Neurosci*. 1992;4:134–139. doi:10.1176/jnp.4.2.134.
34. Schaefer A, Nils F, Sanchez X, Philippot P. Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cogn Emot*. 2010;24:1153–1172. doi:10.1080/02699930903274322.
35. Efron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;7:1–26. doi:10.1214/aos/1176344552.
36. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods*. 2002;7:105–125. doi:10.1037/1082-989X.7.1.105.
37. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol*. 2013;4:863. doi:10.3389/fpsyg.2013.00863.
38. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 2016;70:129–133. doi:10.1080/00031305.2016.1154108.
39. Franke C, Storch A. Nonmotor fluctuations in Parkinson's disease. *Int Rev Neurobiol*. 2017;134:947–971. doi:10.1016/bs.irn.2017.05.021.
40. Seppi K, Ray Chaudhuri K, Coelho M, et al. Update on treatments for nonmotor symptoms of Parkinson's disease—an evidence-based medicine review. *Mov Disord*. 2019;34:180–198. doi:10.1002/mds.27602.
41. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods*. 1999;4:272–299. doi:10.1037/1082-989X.4.3.272.
42. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30:179–185. doi:10.1007/BF02289447.
43. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42. doi:10.1016/j.jclinepi.2006.03.012.