RESEARCH ARTICLE

Using kinematics to re-define the pull test as a quantitative biomarker of the postural response in normal pressure hydrocephalus patients

Samuel Daly[1] · Jacob T. Hanson[1] · Vibha Mavanji[2] · Amy Gravely[3] · James Jean[1] · Alec Jonason[1] · Scott Lewis[4,5] · James Ashe[4,5] · John M. Looft[2] · Robert A. McGovern[1,6]

Abstract

Quantitative biomarkers are needed for the diagnosis, monitoring and therapeutic assessment of postural instability in movement disorder patients. The goal of this study was to create a practical, objective measure of postural instability using kinematic measurements of the pull test. Twenty-one patients with normal pressure hydrocephalus and 20 age-matched control subjects were fitted with inertial measurement units and underwent 10–20 pull tests of varying intensities performed by a trained clinician. Kinematic data were extracted for each pull test and aggregated. Patients participated in 103 sessions for a total of 1555 trials while controls participated in 20 sessions for a total of 299 trials. Patients were separated into groups by MDS-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) pull test score. The center of mass velocity profile easily distinguished between patient groups such that score increases correlated with decreases in peak velocity and later peak velocity onset. All patients except those scored as "3" demonstrated an increase in step length and decrease in reaction time with increasing pull intensity. Groups were distinguished by differences in the relationship of step length to pull intensity (slope) and their overall step length or reaction time regardless of pull intensity (*y*-intercept). NPH patients scored as "normal" on the MDS-UPDRS scale were kinematically indistinguishable from age-matched control subjects during a standardized perturbation, but could be distinguished from controls by their response to a range of pull intensities. An instrumented, purposefully varied pull test produces kinematic metrics useful for distinguishing clinically meaningful differences within hydrocephalus patients as well as distinguishing these patients from healthy, control subjects.

Keywords Postural instability · Reaction time · Perturbation · Step response · Hydrocephalus

✉ Robert A. McGovern
rmcgover@umn.edu

1. Department of Neurosurgery, University of Minnesota Medical School, University of Minnesota, 420 Delaware St. SE, MMC 96, Room D-429, Minneapolis, MN 55455, USA

2. Division of Prosthetics, Motion Capture Analysis Laboratory, Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, USA

3. Department of Statistics, Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, USA

4. Department of Neurology, University of Minnesota Medical School, Minneapolis, MN, USA

5. Department of Neurology, Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, USA

6. Division of Neurosurgery, Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, USA

Introduction

Balance impairments are common in older adults and lead to falls with subsequent injuries, reduced mobility, and decreased quality of life (Burns et al. 2016). In neurodegenerative disorders such as Parkinson's disease (PD) and normal pressure hydrocephalus (NPH), balance impairment—or postural instability—is a key feature of the disease process and can be defined as an inability to maintain an upright posture in response to perturbations or changes in the environment. Neurologists and neurosurgeons commonly assess postural instability using the pull test (UPDRS$_{PT}$) (Goetz et al. 2008) or push and release test (Valkovič et al. 2008), while physiatrists and physical therapists may also use the Berg Balance Scale (Berg et al. 1989), Mini-BEST(Yingyongyudha et al. 2016) or a combination of all these assessments. While useful to individual providers, these assessments are based on

semi-objective ordinal ratings and are, therefore, subject to variability in both execution and interpretation (Munhoz et al. 2004; Nonnekes et al. 2015; Smith et al. 2016). The variety of testing and subjective scoring can make effective communication between providers and specialties difficult when discussing a particular patient's condition.

Quantitative biomarkers for postural instability have historically required a large, dedicated setup and are time-consuming, costly, and not widely available to clinicians. Thus, identifying easily implemented biomarkers for the diagnosis, monitoring and therapeutic assessment of postural instability in the clinical setting is a priority (Horak and Mancini 2013). Such a biomarker should be able to accurately communicate a patient's condition between providers and standardize clinical trial outcomes by negating any dependence on inter-rater reliability. Improvements in wearable technology have led to the emergence of these devices as potential tools in the development of quantitative biomarkers for gait and postural instability (Erb et al. 2020; Dorsey et al. 2020). Metrics from wearable sensors have been developed for almost all aspects of balance including quiet standing (Mancini et al. 2012; Rucco et al. 2018), anticipatory postural adjustments (Mancini et al. 2009), and balance during gait (Zampieri et al. 2010; Weiss et al. 2010, 2013). The major challenges with wearable technology are relating these developed metrics to clinically meaningful results such as fall risk and identifying which metrics are specific to different disease processes. For example, patients with PD will likely have a different pattern of deficits on quantitative wearable metrics compared to patients with NPH or cerebellar ataxia. This will depend on both the underlying disease process and the metrics chosen to evaluate patients.

In this study, we present a method of using kinematic measurements from the pull test in NPH patients to create a quantitative, reliable, and clinically useful measure of postural instability. We have chosen to quantify the $UPDRS_{PT}$ because it is the neurological gold standard and most widely used clinical test available for postural instability. The $UPDRS_{PT}$ can be simply described kinematically as a pull backwards at the shoulders manifested by an increase in the patient's center of mass (COM) acceleration backwards, followed by a latency period prior to taking a step backwards in recovery. During one pull test trial, the patient's COM velocity begins at zero, rises to a peak value and falls back to zero as he recovers. We hypothesized that when separated into groups by their pull test score, COM velocity profiles would significantly differ within hydrocephalus patients and that patients would significantly differ in their step length and reaction times by pull test score. We further hypothesized that within a subset of patients with "normal" pull test scores, patient COM velocity profiles would significantly differ from healthy, age-matched controls.

Methods

Participants

Twenty-one NPH patients and 20 age-matched healthy control participants were consecutively prospectively enrolled over a period of 2 years from the Minneapolis VA Health Care System (MVAHCS) and University of Minnesota. Control participants had unimpaired gait and balance and no prior clinical diagnoses of a neurological disorder affecting movement or perception of movement. NPH participants first see a movement disorders neurologist and are referred for neurosurgical consultation if they have clinical symptoms typical of NPH (gait/balance dysfunction, cognitive difficulties + urinary incontinence) along with ventriculomegaly out of proportion to cortical atrophy. They were excluded if they did not have the capacity to consent as identified by the University of San Diego Brief Assessment of Capacity to Consent (UBACC). We collected relevant demographic data from each patient as well as their responses to a number of questionnaires related to their balance and falls. Patients underwent kinematic, neuropsychological and physical therapy assessment prior to placement of a lumbar drain. They then underwent CSF drainage for 3 days and were re-assessed immediately after removal on the morning of the 4th day. The treating neurosurgeon (RM) used these assessments to decide whether to offer treatment with ventriculoperitoneal shunt (VPS) placement. Regardless of the surgical decision, all patients were prospectively followed with continued assessments. This study was approved by the MVAHCS and University of Minnesota Institutional Review Boards, and all participants provided informed consent for participation.

The NPH patients participated in 103 sessions with a mean of 15.1 pull test trials per session, for a total of 1555 trials. The age-matched controls participated in 20 sessions with a mean of 15.0 trials per session for a total of 299 trials. Demographic data can be found in Table 1. There were no statistically significant differences between groups in terms of age, as expected, but there were differences in sex and weight as most of the NPH patients were veterans, and therefore men. Control participants walked faster than NPH patients and had better subjective and semi-objective measures of balance (ABC, $UPDRS_{PT}$ scores, respectively). We did not find any differences in kinematic outcomes based on sex within $UPDRS_{PT}$ groups (Supplemental Fig. 1). Within NPH participants, stride velocity decreased with increasing $UPDRS_{PT}$ score due to decreases in step length (Supplemental Table 1).

Table 1 Participant demographics

Variable	NPH, $n=21$	Controls, $n=20$	p value
Age, mean years (SD)	72.6 (7.6)	70.0 (4.0)	0.354
Sex	3 F, 18 M	16 F, 4 M	n/a
Height	171.8 (10.5)	168.3 (9.3)	0.512
Weight	89.0 (21)	69.4 (12)	< 0.001
VP shunt	6N:15Y	n/a	n/a
Number of sessions	103	20	n/a
Number of trials	1555	299	n/a
Average number of trials per session	15.1	15	n/a
UPDRS$_{PT}$ score	1.4 (0.9)	0.3 (0.5)	< 0.001
ABC score	68.0 (24)	96.4 (4.5)	< 0.001
Stride velocity (m/s)	1.15 (0.3)	1.7 (0.3)	< 0.001
Stride length (m)	1.3 (0.5)	1.0 (0.2)	0.01
Cadence (steps/s)	0.9 (0.2)	1.7 (0.3)	< 0.001

UPDRS$_{PT}$ Unified Parkinson's Disease Rating Scale for Pull Test, *VP* ventriculoperitoneal, *ABC* Activities-Specific Balance Confidence Scale

Task details

Participants were equipped with a set of 15 inertial measurement units (IMUs; Xsens, Enschede, The Netherlands) recording at 60 Hz during each pull test session (Supplemental Fig. 2). Body limb segments for each patient were measured at their baseline testing session. The pull test was executed for each patient by a trained clinical examiner. The examiner followed the instructions on the MDS-UPDRS form for conducting the pull test (Goetz et al. 2008) and conducted between 10 and 20 pull test trials for each patient. The examiner used clinical discretion to determine the force of the induced perturbation during the recorded trials, but they were instructed to use a variety of intensities throughout the trials as able. The first trial after the instructional trial was scored in the standard manner on the MDS-UPDRS scale. This is scored as "0—normal": recovers with 1–2 steps; "1—slight": 3–5 steps but recovers unaided; "2—mild": More than 5 steps but recovers unaided; "3—moderate": stands safely but with absence of postural response, falls if not caught by examiner; "4—severe": very unstable, tends to lose balance spontaneously or with just a gentle pull on the shoulders. Although there were 10–20 trials during each session, the patient's session was scored solely on the first trial after the instructional trial as required by the MDS-UPDRS scale (Goetz et al. 2008).

Data extraction

A three-dimensional biomechanical model is created for each patient based on limb segment inputs and sensor orientation (Schepers et al. 2018). Center of mass (COM) and foot position data were exported from the recorded motion capture analysis file and imported and analyzed within Igor Pro 6.00 (Wavemetrics, Oregon, USA) to calculate velocity and acceleration for the COM, feet and other body segments. Custom functions were then used to identify the relevant points of interest (Supplemental Methods, Supplemental Fig. 1). Once relevant points were extracted from the trials, they were exported into R (v 3.6.2) (R Core Team 2018) and further analyzed. Plots were created using the R package "ggplot2." Supplemental Table 2 contains a description and interpretation of the relevant kinematic variables while Supplemental Fig. 3 demonstrates the relevant points of interest for each individual trial (McGovern et al. 2020).

Statistical analysis

The purpose of this study was to identify kinematic outcomes specific to the scored UPDRS$_{PT}$ group and to look for differences between patients scored as "normal" or "0" on the MDS-UPDRS scale and age-matched controls also scored as "0." As a result, we did not group NPH patients by treatment status, visit or any other surgically related variable as the only a priori factors we were interested in were the UPDRS$_{PT}$ score and disease vs. control. Therefore, once the relevant kinematic points were identified and variables calculated as described above, we grouped patients by UPDRS$_{PT}$ score and performed the statistical testing described below. We also took a subset of NPH patients scored as "0" and compared them to age-matched controls using the same set of statistical testing.

Statistically significant differences in center of mass velocity (V_{COM}) plots were considered to be non-overlapping 95% confidence intervals of timeframes 50 ms or greater. Linear models were created to analyze the interaction between step length and pull intensity as well as reaction time and pull intensity for each UPDRS$_{PT}$ group. To test for an overall effect, we first performed ANCOVA. If the effect was significant, we then compared differences between the slope of each of these groups using the estimated marginal means of linear trends (package "emmeans" in R) with a Tukey HSD correction. To compare the y-intercept of each of these groups, we used the estimated marginal means of the intercept values, again using a Tukey HSD correction. To assess whether there was a learning effect associated with repeated pull test measurements, we included trial number and its interaction with pull intensity in the linear models described above. Trial number did not show any statistically significant effects on step length or reaction time and so it was not considered in the analyses (Supplemental Table 3). For standardized perturbations, differences in step length, reaction time or pull intensity were evaluated with two-sided t tests. p values less than 0.05 were considered statistically significant.
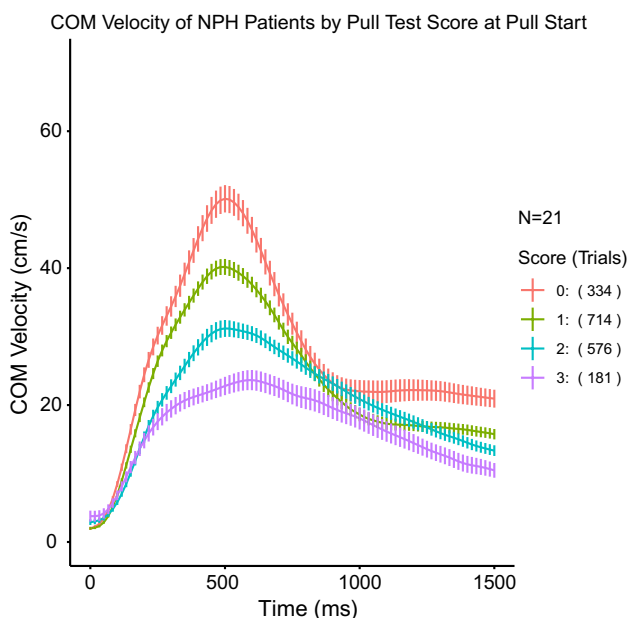
Results

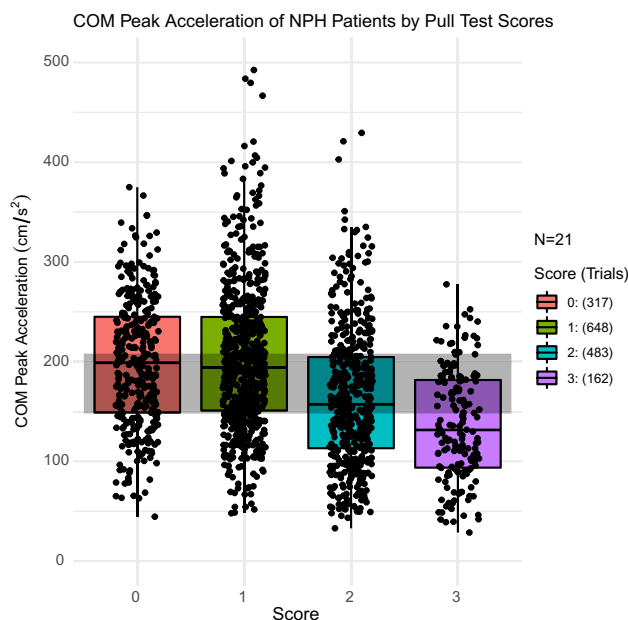## NPH patient kinematic profile

V_{COM} differentiates between UPDRS$_{PT}$ scores within NPH patients

We first sought to describe an objective, kinematic profile for each pull test trial using the patient's center of mass velocity (V_{COM}). In its simplest terms, each pull test trial can be described by the patient's V_{COM} as it begins at zero when the patient is at rest, rises as the patient is pulled backwards, continues to rise as the patient steps backwards and then declines and approaches zero as the patient recovers from the perturbation (middle panel, Supplemental Fig. 3). When we stratified each patient session by UPDRS$_{PT}$ score and averaged V_{COM} for all trials, we found statistically significant differences in V_{COM} between all groups of patients within the first 500 ms after pull onset. Each group's V_{COM} diverged within the first 300 ms of pull test initiation as increasing UPDRS$_{PT}$ score correlated with decreases in peak V_{COM} values (Fig. 1).

We considered, however, that there are both clinician and patient-related factors that could lead to the kinematic differences seen between groups. First, differences in how hard patients in each group were pulled could lead to some of these differences as harder pulls would lead to higher peak V_{COM} values. Figure 2 shows that patients scored as "0" (198.8 + 66 cm/s) or "1" (202.8 + 74 cm/s) were pulled at
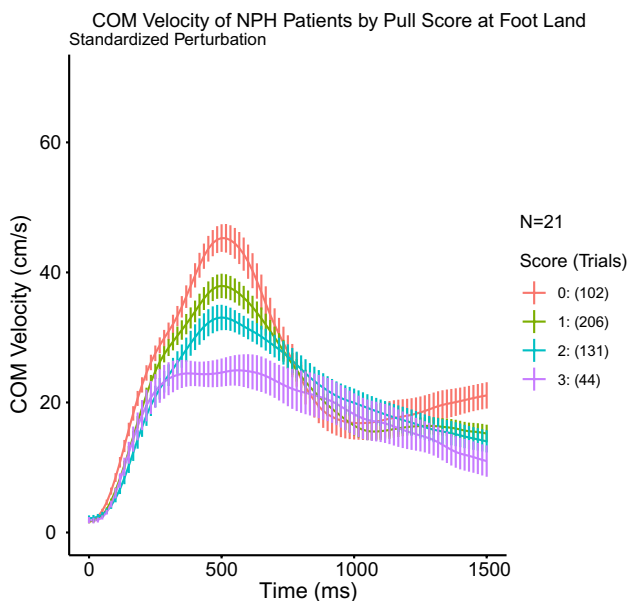

Fig. 2 Mean pull intensity grouped by UPDRS$_{PT}$ score. Shaded area represents the 33rd to 66th percentile of all pulls

similar intensity to each other overall but were pulled harder on average than patients scored as "2" (161.6 + 69 cm/s; 0 vs. 2, $p < 0.001$, 1 vs. 2, $p < 0.001$ Tukey HSD test) or "3" (137.3 + 56 cm/s; 0 vs. 3, $p < 0.001$, 1 vs. 3, $p < 0.001$, Tukey HSD test).

NPH patients' response to a "Standardized Perturbation"

Therefore, we next restricted our analysis to a narrow range of pull intensities (33rd to 66th percentile, shaded area in Fig. 2), as measured by peak A_{COM} prior to step onset (see Supplemental Methods and Supplemental Table 2 for details). Since A_{COM} is measured solely prior to step onset, it should purely reflect the kinematic effect of the pull on the COM. When we examined the pull intensity of each group after restricting the data, we saw that our intended effect occurred with almost identical mean pull intensities (Supplemental Fig. 4) between each group.

Within the "standardized" perturbations, a very similar pattern of V_{COM} outcomes emerged for all groups with a few key differences when compared to all perturbations. First, because there were fewer trials included, all groups demonstrated wider confidence intervals with more overlap between groups (Fig. 3). Patients scored as "1", "2" and "3" were all statistically indistinguishable within the first 300–400 ms until the patients scored as "3" diverged at 300 ms to a lower peak V_{COM} value. Patients scored as "1" and "2" had very similar V_{COM} responses, only separating from each other at peak V_{COM} values at 500 ms. Statistically


Fig. 1 Kinematic pull test outcomes in NPH patients. Mean V_{COM} vs. time grouped by UPDRS$_{PT}$ score

Fig. 3 "Standardized" kinematic pull test outcomes in NPH patients. Mean V_{COM} vs. time for a "standardized" perturbation grouped by $UPDRS_{PT}$ score

significant decreases in peak V_{COM} remained associated with increasing $UPDRS_{PT}$ scores for all groups. Since statistically significant differences remained between patient groups even after standardizing the perturbation, we considered that patient-related factors must also be important in their distinct outcomes. As V_{COM} of all patients except those

considered "normal" ("0") were all quite similar within the first 350 ms of the "standardized" perturbation, we hypothesized that reaction time (time from pull onset to step initiation; Supplemental Fig. 3) may be faster for the patients scored as "0" but similar for all other groups. Since V_{COM} distinguished between all groups at 350–650 ms, we also hypothesized that the patient's initial step length backwards would differ between all groups as this was the when the first step backwards was typically taken (Supplemental Fig. 3).

Reaction time of NPH patients in a "Standardized" perturbation

To examine this, we first looked at reaction time over the standardized range of perturbations (Fig. 4A). Our original hypothesis was confirmed in that reaction time was faster in patients scored as "0" (227 ms + 74 ms) compared to all other groups. However, reaction time was also faster in patients scored as "1" (297 ms + 130 ms) compared to the "2" (380 ms + 180 ms) and "3" (358 ms + 170 ms) groups despite similar appearing V_{COM} in the initial 400 ms time period between these three groups. We then restricted V_{COM} exclusively to the time from step onset until the end of the trial to eliminate the effect of the patient's reaction time on V_{COM} (Fig. 4B). All groups started with a nearly identical V_{COM}, indicating that the initial divergence seen between "0" and all other groups were likely related to differences in reaction time. From the initial step onset, V_{COM} quickly diverged within the first 300–400 ms (approximately the
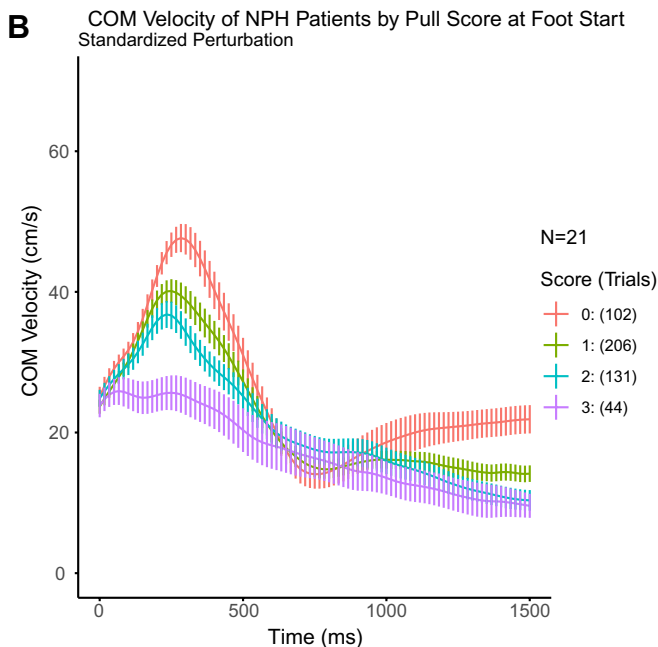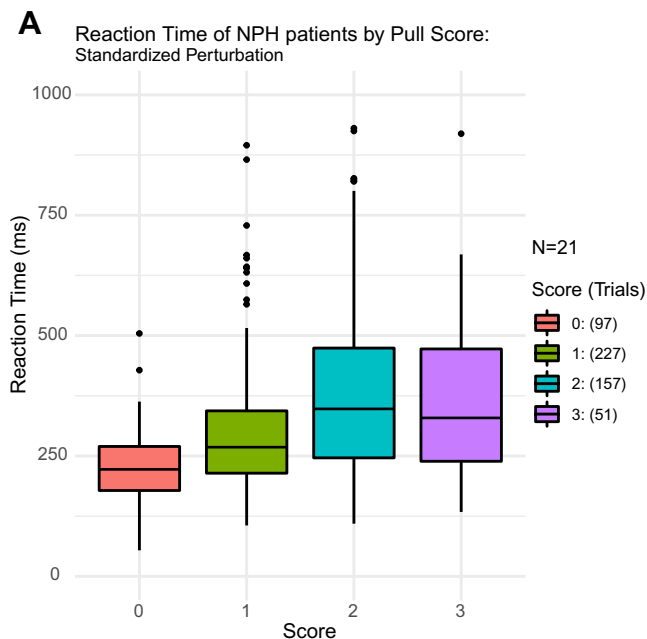
Fig. 4 Reactive postural response to a "standardized" perturbation of NPH patients. **A** Reaction time to a standard perturbation grouped by $UPDRS_{PT}$ score. **B** Mean V_{COM} vs. time for a standardized perturbation starting at the beginning of the first step, grouped by $UPDRS_{PT}$ score

duration of the first step) and then converged again at the end of the second step at 600–700 ms. Statistically significant differences between similarly scored groups (i.e., "0" vs "1", "1" vs "2") occurred at or near peak V_{COM} values while profiles for dissimilar groups (i.e., "0" vs. "2", "1" vs. "3") remained easily differentiated throughout the duration of the first step. This indicated to us that differences in the first step backwards were likely responsible for this subsequent divergence in V_{COM}.

Initial step length of NPH patients in a "Standardized" perturbation

We, therefore, looked at the first step length over the standardized range of perturbations. Step length decreased with increasing $UPDRS_{PT}$ score with statistically significant differences between each group (Fig. 5A). When we restricted V_{COM} exclusively to the time from the end of the first step until the end of the trial to eliminate the effect of the patient's first step, we saw no further differences between groups (Fig. 5B). This indicated to us that the vast majority of differences in V_{COM} between groups for a "standardized" perturbation was the result of reaction time and first step differences.

While a standardized perturbation provides insight into how patients with varying postural instability respond differently, in the real-world setting, patients frequently experience many different unexpected perturbations of a variety of intensities. To simulate this, we pulled patients backwards

at a variety of intensities and examined the same parameters discussed above to examine the patients' responses throughout the entirety of pull intensities.

Reaction time and step length scaling to perturbation intensity

Given the differences in reaction time and step length between groups for a standardized perturbation, we hypothesized that there may be differences in the relationship of reaction time and/or step length to pull intensity between groups as well. When examining reaction time, all groups except patients scored as "3" could scale their reaction time to increases in pull intensity (Fig. 6A). Thus, patients scored as "0", "1", and "2" all demonstrated quicker reaction times for increasing pull intensities (slope < 0) while patients scored as "3" demonstrated no significant change in reaction time with increasing pull intensity (slope = 0). Groups were mostly statistically indistinguishable when comparing reaction time slope and intercept values with the sole exception of the comparison between patients scored as "2" and all other groups (Supplemental Table 4). In general, the relationship between reaction time and pull intensity did not differentiate groups as well as the relationship between step length and pull intensity.

Similar to reaction time, all groups except patients classified as "3" increased their initial step length for increases in pull intensity (non-zero positive slope, Fig. 6B, Supplemental Table 5). Groups could be stratified according
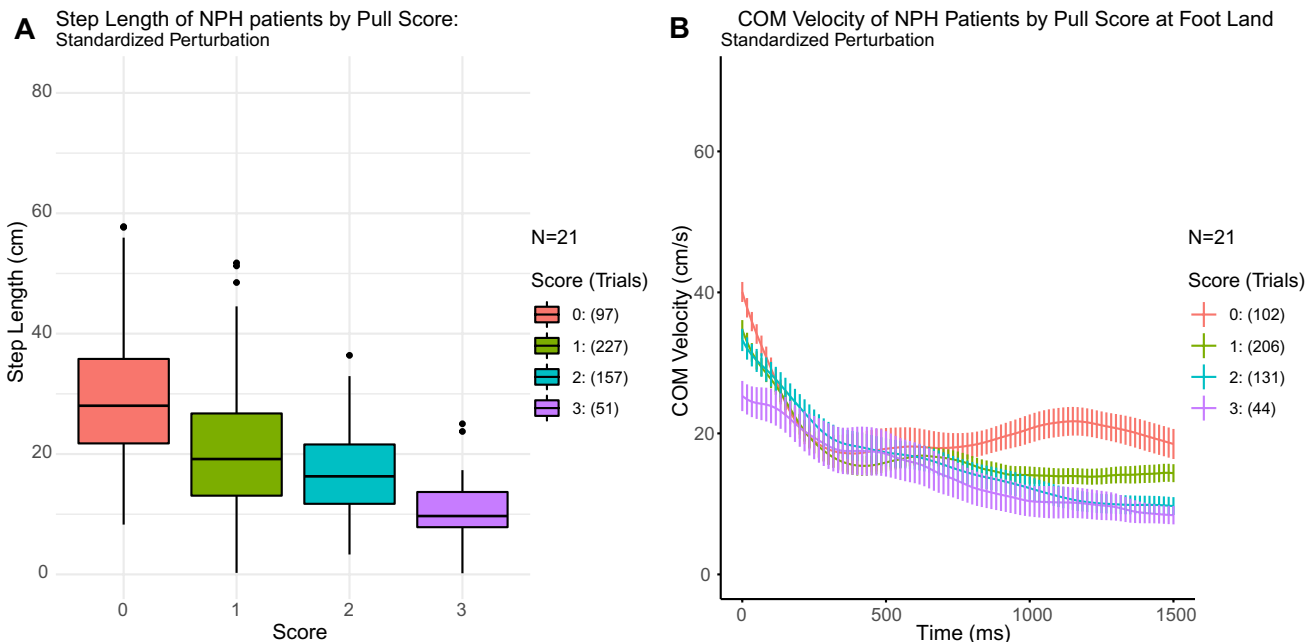
Fig. 5 Reactive postural response to a "standardized" perturbation of NPH patients. **A** Initial step length to a standard perturbation grouped by $UPDRS_{PT}$ score. **B** Mean V_{COM} vs. time for a standardized perturbation starting at the end of the first step, grouped by $UPDRS_{PT}$ score
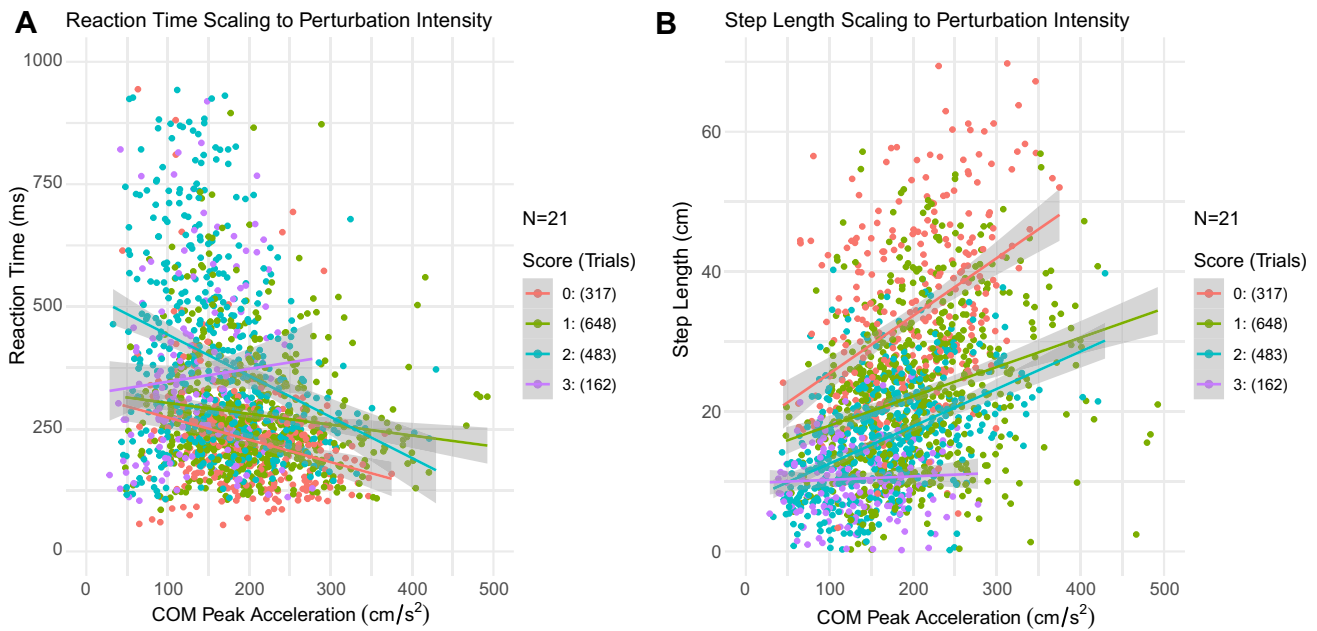
Fig. 6 Reactive postural response scaling of NPH patients. **A** Reaction time as a function of pull intensity grouped by UPDRS$_{PT}$ score. **B** Initial step length as a function of pull intensity grouped by UPDRS$_{PT}$ score

to both their overall step length for a given pull intensity (*y*-intercept) and their ability to scale the initial step length to pull intensity (slope). Patients scored as "0" had the highest intercept and slope values with statistically significantly larger slope than all other patient groups and statistically significant larger intercept values than all other groups except "1" patients. Patients scored as "1" and "2" had similar slope values but could be statistically distinguished by their *y*-intercepts (smaller for "2" patients). Patients scored as "3" had similar intercept values to the other groups, but could be easily statistically differentiated by their slope values as they were the only group unable to scale their step length to increases in pull intensity (i.e., slope=0).

## Kinematic outcomes in "Normal" NPH patients and age-matched control participants

We wanted to compare whether NPH patients with a "normal" score on the UPDRS$_{PT}$ scale could be differentiated from age-matched control participants using V_{COM}. Overall, both NPH patients and age-matched controls had very similar V_{COM} profiles with control participants diverging from NPH patients at approximately 180 ms and reaching a slightly higher peak V_{COM} (59.1 + 2 vs 54.6 + 2.1 cm/s) with slightly later peak V_{COM} onset (533 vs. 500 ms; Fig. 7A). Control participants were pulled harder on average with mean peak A_{COM} values statistically significantly higher (268.4 + 83 cm/s^2 vs. 198.8 + 66 cm/s^2, $p < 0.001$ Tukey HSD test; Fig. 7B). To normalize for pull intensity, we next

examined only responses to a "standardized" perturbation in a similar process to that described above for NPH patients. Control and "normal" NPH patients demonstrated an almost identical V_{COM} response to a "standardized" perturbation with no statistical differences throughout the entire response (Fig. 7C). There were no significant differences in either step length (37.9 + 7 vs. 37.2 + 12 cm, control vs. "normal" NPH, $p = 0.62$) or reaction time (202 + 50 vs. 213 + 90 ms, control vs. "normal" NPH, $p = 0.31$) responses to the "standardized" perturbation between NPH patients and controls (Fig. 7D, E). This indicated to us that the small differences seen in V_{COM} were likely due to the differences in pull intensity as controls were pulled harder.

Reaction time and step length scaling distinguishes NPH patients from controls

Interestingly, although control and "normal" NPH patients had nearly identical kinematic outcomes in response to a standardized perturbation, their responses to the entire pull intensity range differed significantly (Fig. 8). Control participants had faster reaction times for a given pull intensity (lower *y*-intercept, 244 + 14 ms vs 323 + 21 ms, $p = 0.006$, Tukey HSD; Fig. 8A) but demonstrated less ability to scale their reaction time to pull intensity (lower slope -1.33×10^{-4} vs. -4.6×10^{-4}, $p = 0.006$, Tukey HSD). Similarly, for step length, control participants demonstrated longer step lengths for a given pull intensity (increased *y*-intercept, 25.6 + 1.7 cm vs. 17.4 + 2.2 cm, $p = 0.009$) but less ability to
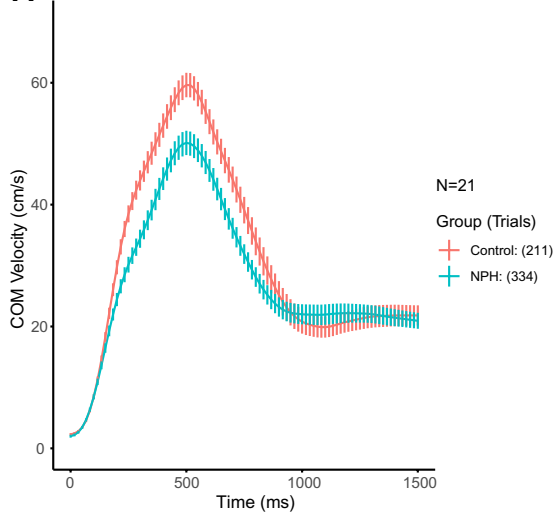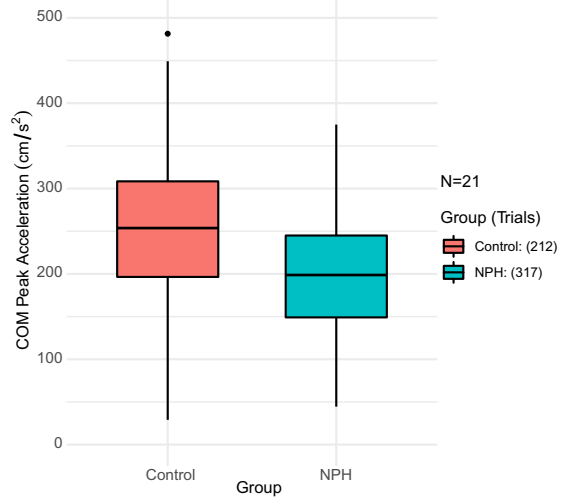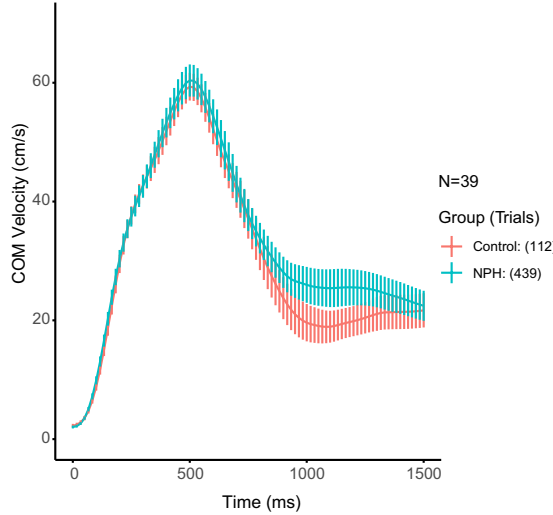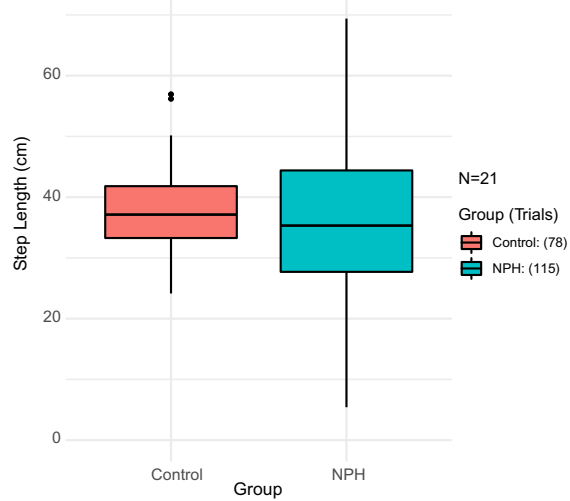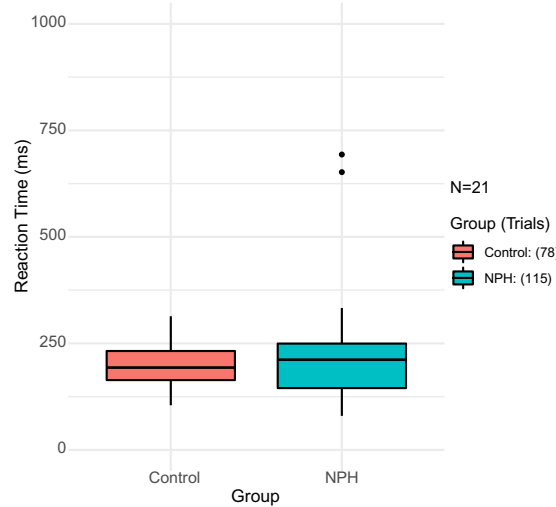
◄**Fig. 7** Comparison of reactive postural response in control and "normal" NPH patients. **A** Mean V_{COM} vs. time for controls and "normal" NPH patients. **B** Mean perturbation pull intensity for controls vs. "normal" NPH patients. **C** Mean V_{COM} vs. time for controls and "normal" NPH patients for a "standardized" perturbation. **D** Initial step length and **E** reaction time responses to a "standardized" perturbation for controls and "normal" NPH patients

scale their step length to a given pull intensity (lower slope; $0.049 + 0.01$ vs. $0.083 + 0.009$, $p = 0.009$).

Discussion

We have described a kinematic method of measuring the $UPDRS_{PT}$ in an effort to establish a quantitative, clinically meaningful outcome measure that can be reliably used in the clinical setting and further developed as a biomarker of postural instability. The simplest description of the pull test's outcome, the patient's V_{COM} throughout the duration of the test, can reliably differentiate between $UPDRS_{PT}$ scores within NPH patients (Fig. 1). This indicated to us that V_{COM} discriminated between clinically meaningful criteria.

When limited to a narrow range of perturbations to approximate a standardized pull, differences between groups are reduced (Fig. 3), indicating that some of the differences in V_{COM} profiles may be attributed to how hard patients are pulled. Indeed, this is typically seen as one of the disadvantages of using the $UPDRS_{PT}$ in the clinical setting as it is difficult to standardize both between patients and providers leading to poor inter-rater reliability (Munhoz et al. 2004; Munhoz and Teive 2014). As we purposefully pulled at a variety of intensities, patients and controls with better balance would likely be pulled harder on average than those with postural instability. In fact, patients scored as "0" or "1" were—on average—pulled harder than those scored as "2" or "3" (Fig. 2), although there was significant overlap between the groups. This is likely reflective of better postural control in patients with lower scores as they were able to tolerate higher intensity pulls. Some patients actually demonstrated increases in mean pull intensity as their balance improved post-operatively (Supplemental Fig. 5). Thus, tracking this metric over time may be investigated in the future as an indicator of clinical improvement or worsening. Our use of "standardized" pulls also demonstrates that this process could be used to examine different patients' responses to a similar perturbation.

In our experience, most providers tend to pull quite lightly, perhaps due to concern over causing the patient to fall backwards. This will lead to overestimation of the patient's performance as even patients with poor balance can frequently recover from a light pull in a few steps. Performing the test with the examiner's back against a wall can mitigate this concern and allow the examiner to put the

patient through the entire range of pull intensities although this does require an examiner comfortable and skilled in the technique. Since changes in pull intensity are associated with differences in both patients and controls' postural response, pull intensity needs to be quantified in some manner. Quantifying differences in pull intensity between providers could also serve as a possible mechanism to standardize evaluations in the future if desired.

While trying to standardize a perturbation has obvious advantages in comparing reactive postural stability between patients, our data demonstrate the value in purposefully inducing variability into the pull test. Patients' responses vary according to the intensity of the perturbation, allowing the clinician to characterize the entirety of the patient's response (Figs. 6, 8). While clinicians often try to "standardize" their perturbation within and between patients, from a practical perspective, this task is impossible (especially when patients are seen by multiple providers) (Munhoz et al. 2004). Even if the pull test (or any test of postural instability) could be standardized, our data demonstrate that a patient's response at one standardized perturbation cannot necessarily be used as an approximation for the response at other perturbations without knowing how their response varies. In addition, depending where along the continuum of pull intensity the test was standardized, the same patient's response could be considered adequate (not falling for a low intensity perturbation) or inadequate (falling after a high intensity perturbation). This would not change their underlying postural response, only the arbitrary manner in which we have decided to categorize them. For example, while all patient groups demonstrate similar responses in the lower range of pull intensities, they vary considerably for high intensity pulls. In addition, "normal" NPH patients and control participants looked identical in the middle range of standardized perturbations but their responses are statistically significantly different in the lower or higher ranges of pull intensity. Without testing the full range of pull intensities, it would be impossible to distinguish between "normal" NPH patients and controls. As a result, quantifying the inherent variability present in the test is much more likely to be fruitful in terms of both understanding and classifying postural instability in movement disorder patients.

While the gait of NPH patients has been investigated in detail (Stolze 2001), postural instability has typically not been very well studied and its underlying mechanisms are poorly understood. The neural control of postural stability depends on integration of somatosensory, vestibular, and visual inputs and maintaining equilibrium requires coordinating these systems to produce motor strategies that maintain one's COM in a stable range to prevent falls (Horak 2006). Deficits in any of these systems impair coordination and the ability to maintain an upright posture. NPH patients have a larger static sway area and higher backwards
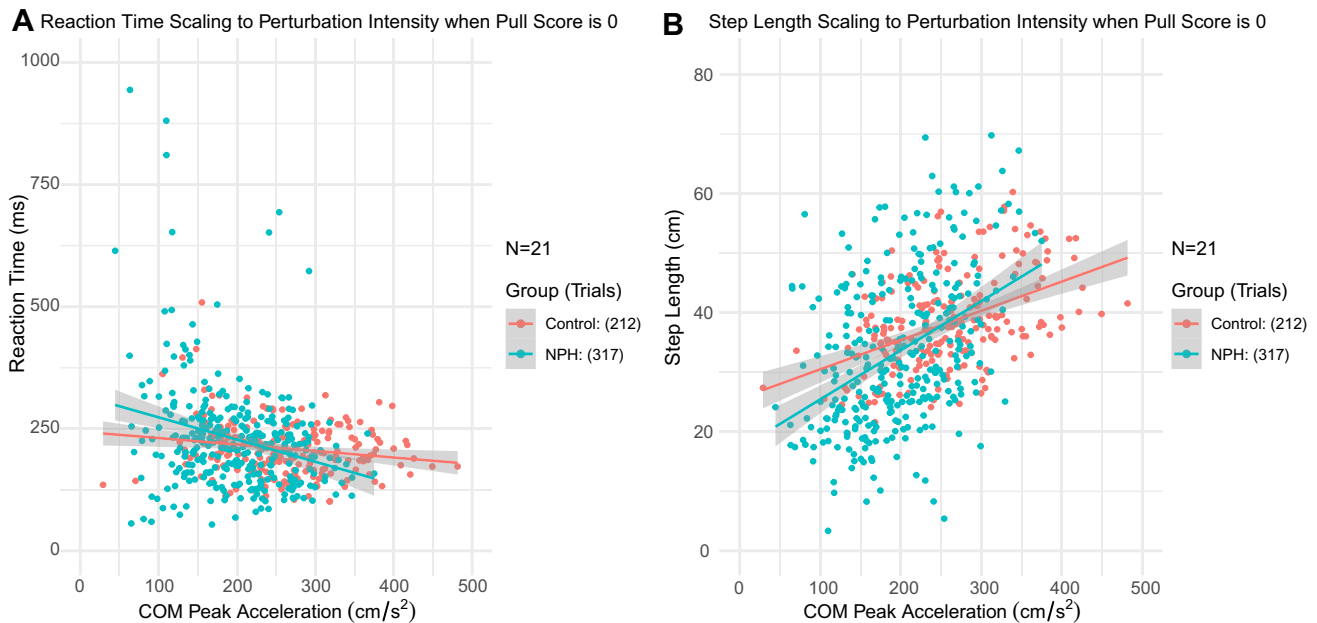
Fig. 8 Reactive postural response scaling of control and "normal" NPH patients. **A** Reaction time as a function of pull intensity. **B** Initial step length as a function of pull intensity

directed COM velocity during upright stance (Blomsterwall et al. 2000). Some studies have suggested that NPH patients rely less on visual input than healthy controls (Bäcklund et al. 2017), but others have shown typical improvements in balance with eyes open vs. closed (Lundin et al. 2013). NPH patients whose motor performance improves after VPS demonstrate bilateral increases in supplementary motor area (SMA) activation (Lenfeldt et al. 2008), an area known to be important for postural control planning and execution (Slobounov et al. 2005; Richard et al. 2017; Fujimoto et al. 2017) though this study used a simple measure of finger tapping to measure motor performance. Thus, investigating the role of SMA in the pathogenesis of postural instability in NPH represents a potential area for future mechanistic studies.

No dynamic studies have examined NPH patients' postural step response in detail prior to this study. When examining their underlying postural responses by UPDRS$_{PT}$ scores, NPH patient groups differed in both their ability to scale step length to increasing pull intensity and their overall step length for a given intensity (Fig. 6B). These differences were much more robust than differences in reaction time, signifying that planning and execution of the initial step backwards is likely the most important kinematic parameter for determining postural control and for clinically differentiating between NPH patients. Studies examining the postural step response using standardized perturbations in PD patients have generally shown that PD patients take smaller steps than healthy subjects, (King et al. 2010; Smulders et al. 2014; Di Giulio et al. 2016; Schlenstedt et al. 2017) with slower reaction times that may be due to an increase

in the number (King et al. 2010) and size (Schlenstedt et al. 2017) of anticipatory postural adjustments prior to stepping. This suggests that the kinematic deficits captured within these parameters are not necessarily disease specific. With their step length kinematic deficits now well-defined, there is much mechanistic work in future studies to determine whether these deficits are due to NPH affecting visual, vestibular, somatosensory systems or some combination of all.

Age-matched controls and NPH patients who scored "0" on the UPDRS$_{PT}$ had identical kinematic responses to a "standardized" perturbation (Fig. 7), but had differing step length and reaction time scaling responses when examining the entirety of the pull intensity spectrum (Fig. 8). Interestingly, NPH patients with normal UPDRS$_{PT}$ scores demonstrated improved ability to scale their initial step length and reaction time to increasing pull intensity as compared to controls (Fig. 8). This kinematic response is typically seen in younger control subjects with increasing age reducing the ability to scale step size or reaction time (McGovern et al. 2020). As all these patients had surgical intervention, this suggests that surgery may fundamentally impact the reactive postural response in NPH patients. Further studies that specifically compare age- and UPDRS$_{PT}$-matched controls to movement disorder patients with varying disease processes (e.g., PD and NPH) and interventions such as medication or surgery will need to be performed to identify age-related, disease-specific and interventional phenomenon.

This study has limitations related to its sample size, kinematic approach, technology and patient selection. Given the small number of patients and controls in this study, it

is possible that the study is underpowered and, therefore, may be unable to detect more subtle differences between groups, particularly given the heterogeneity of patients with NPH. Kinematics is focused on the end result of the body's motion in space, and therefore, it gives limited insight into the underlying control mechanisms producing that result. To gain insight into the involved motor control pathways, kinematics need to be combined with other techniques such as intracranial local field potential recordings, EMG, EEG, and/or fMRI. While easily incorporated into our clinical visits, our motion capture system uses 15 IMUs and is expensive. Future work needs to be focused on reducing the relevant set of sensors to 5 or fewer that can be cheaply purchased. With regards to patient selection, this study may be limited in its generalizability as we only included patients with NPH. It is possible that patients with Parkinson's disease will demonstrate different kinematic patterns based on their UPDRS$_{PT}$ score. In addition, as our NPH patients consisted mostly of veterans, almost all our patients were men, meaning that our results need to be confirmed in more women.

The strengths of this approach lie in its practical advantage of incorporating the most widely used clinical test of postural instability and its analysis of the patient's response throughout a range of perturbation intensities. Therefore, as a bridge between the clinic and laboratory, an easily adopted quantitative UPDRS$_{PT}$ should have significantly more impact on diagnosis, monitoring, and clinical trial assessment than a test that can only be applied in certain laboratories or less widely applicable settings.

Author contributions SD: software, formal analysis, writing—original draft, review, and editing. JH: investigation, project administration, data curation, and writing—original draft. VM: investigation. AG: software and formal analysis. JJ: software, data curation, and formal analysis. AJ: investigation, project administration, and data curation. SL: conceptualization, methodology, and writing—review and editing. JA: conceptualization, methodology, writing—review and editing. JML: conceptualization, methodology, investigation, writing—review and editing. RM: supervision, conceptualization, methodology, investigation, software, formal analysis, writing—original draft, review, and editing. All the authors read and approved the final manuscript.

Availability of data and material Available upon request.

Code availability Available upon request.

Declarations

Conflict of interest None.

Ethics approval This study was approved by the MVAHCS and University of Minnesota Institutional Review Boards and was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki.

Consent to participate All the subjects provided informed consent to participate in this study.

Consent to publish All subjects provided informed consent to allow their data to be published.

References

Bäcklund T, Frankel J, Israelsson H et al (2017) Trunk sway in idiopathic normal pressure hydrocephalus—quantitative assessment in clinical practice. Gait Posture 54:62–70. https://doi.org/10.1016/j.gaitpost.2017.02.017

Berg K, Wood-Dauphine S, Williams JI, Gayton D (1989) Measuring balance in the elderly: preliminary development of an instrument. Physiother Can 41:304–311. https://doi.org/10.3138/ptc.41.6.304

Blomsterwall E, Svantesson U, Carlsson U et al (2000) Postural disturbance in patients with normal pressure hydrocephalus. Acta Neurol Scand 102:284–291

Burns ER, Stevens JA, Lee R (2016) The direct costs of fatal and nonfatal falls among older adults—United States. J Saf Res 58:99–103. https://doi.org/10.1016/j.jsr.2016.05.001

Di Giulio I, St George RJ, Kalliolia E et al (2016) Maintaining balance against force perturbations: impaired mechanisms unresponsive to levodopa in Parkinson's disease. J Neurophysiol 116:493–502. https://doi.org/10.1152/jn.00996.2015

Dorsey ER, Omberg L, Waddell E et al (2020) Deep phenotyping of Parkinson's disease. JPD. https://doi.org/10.3233/JPD-202006

Erb MK, Karlin DR, Ho BK et al (2020) mHealth and wearable technology should replace motor diaries to track motor fluctuations in Parkinson's disease. NPJ Digit Med. https://doi.org/10.1038/s41746-019-0214-x

Fujimoto H, Mihara M, Hattori N et al (2017) Neurofeedback-induced facilitation of the supplementary motor area affects postural stability. Nph 4:045003. https://doi.org/10.1117/1.NPh.4.4.045003

Goetz CG, Tilley BC, Shaftman SR et al (2008) Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 23:2129–2170. https://doi.org/10.1002/mds.22340

Horak FB (2006) Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? Age Ageing 35(Suppl 2):ii7–ii11. https://doi.org/10.1093/ageing/afl077

Horak FB, Mancini M (2013) Objective biomarkers of balance and gait for Parkinson's disease using body-worn sensors. Mov Disord 28:1544–1551. https://doi.org/10.1002/mds.25684

King LA, George RJS, Carlson-Kuhta P et al (2010) Preparation for compensatory forward stepping in Parkinson's disease. Arch Phys Med Rehabil 91:1332–1338. https://doi.org/10.1016/j.apmr.2010.05.013

Lenfeldt N, Larsson A, Nyberg L et al (2008) Idiopathic normal pressure hydrocephalus: increased supplementary motor activity accounts for improvement after CSF drainage. Brain. https://doi.org/10.1093/brain/awn232

Lundin F, Ledin T, Wikkelsø C, Leijon G (2013) Postural function in idiopathic normal pressure hydrocephalus before and after shunt surgery: a controlled study using computerized dynamic posturography (EquiTest). Clin Neurol Neurosurg 115:1626–1631. https://doi.org/10.1016/j.clineuro.2013.02.015

Mancini M, Zampieri C, Carlson-Kuhta P et al (2009) Anticipatory postural adjustments prior to step initiation are hypometric in untreated Parkinson's disease: an accelerometer-based approach. Eur J Neurol 16:1028–1034. https://doi.org/10.1111/j.1468-1331.2009.02641.x

Mancini M, Salarian A, Carlson-Kuhta P et al (2012) ISway: a sensitive, valid and reliable measure of postural control. J NeuroEng Rehabil 9:59. https://doi.org/10.1186/1743-0003-9-59

McGovern RA, Cortes JC, Wilson AP et al (2020) Impaired scaling of step length in Parkinsonian postural instability. medRxiv. https://doi.org/10.1101/2020.08.25.20179945

Munhoz RP, Teive HA (2014) Pull test performance and correlation with falls risk in Parkinson?s disease. Arq Neuro-Psiquiatr 72:587–591. https://doi.org/10.1590/0004-282X20140082

Munhoz RP, Li J-Y, Kurtinecz M et al (2004) Evaluation of the pull test technique in assessing postural instability in Parkinson's disease. Neurology 62:125–127

Nonnekes J, Goselink R, Weerdesteyn V, Bloem BR (2015) The retropulsion test: a good evaluation of postural instability in Parkinson's disease? J Parkinson's Dis 5:43–47. https://doi.org/10.3233/JPD-140514

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Richard A, Van Hamme A, Drevelle X et al (2017) Contribution of the supplementary motor area and the cerebellum to the anticipatory postural adjustments and execution phases of human gait initiation. Neuroscience 358:181–189. https://doi.org/10.1016/j.neuroscience.2017.06.047

Rucco R, Sorriso A, Liparoti M et al (2018) Type and location of wearable sensors for monitoring falls during static and dynamic tasks in healthy elderly: a review. Sensors 18:1613. https://doi.org/10.3390/s18051613

Schepers M, Giuberti M, Bellusci G (2018) Xsens MVN: consistent tracking of human motion using inertial sensing. Xsens Technologies, pp 1–8. https://www.researchgate.net/publication/324007368_Xsens_MVN_Consistent_Tracking_of_Human_Motion_Using_Inertial_Sensing/link/5ab8be2f0f7e9b68ef51f7ba/download

Schlenstedt C, Mancini M, Horak F, Peterson D (2017) Anticipatory postural adjustment during self-initiated, cued, and compensatory stepping in healthy older adults and patients with Parkinson disease. Arch Phys Med Rehabil 98:1316-1324.e1. https://doi.org/10.1016/j.apmr.2017.01.023

Slobounov S, Hallett M, Stanhope S, Shibasaki H (2005) Role of cerebral cortex in human postural control: an EEG study. Clin Neurophysiol 116:315–323. https://doi.org/10.1016/j.clinph.2004.09.007

Smith BA, Carlson-Kuhta P, Horak FB (2016) Consistency in administration and response for the backward push and release test: a clinical assessment of postural responses. Physiother Res Int 21:36–46. https://doi.org/10.1002/pri.1615

Smulders K, Esselink RA, De Swart BJ et al (2014) Postural inflexibility in PD: does it affect compensatory stepping? Gait Posture 39:700–706. https://doi.org/10.1016/j.gaitpost.2013.10.003

Stolze H (2001) Comparative analysis of the gait disorder of normal pressure hydrocephalus and Parkinson's disease. J Neurol Neurosurg Psychiatry 70:289–297. https://doi.org/10.1136/jnnp.70.3.289

Valkovič P, Brožová H, Bötzel K et al (2008) Push and release test predicts better Parkinson fallers and nonfallers than the pull test: comparison in OFF and ON medication states. Mov Disord 23:1453–1457. https://doi.org/10.1002/mds.22131

Weiss A, Herman T, Plotnik M et al (2010) Can an accelerometer enhance the utility of the Timed Up & Go Test when evaluating patients with Parkinson's disease? Med Eng Phys 32:119–125. https://doi.org/10.1016/j.medengphy.2009.10.015

Weiss A, Brozgol M, Dorfman M et al (2013) Does the evaluation of gait quality during daily life provide insight into fall risk? A novel approach using 3-day accelerometer recordings. Neurorehabil Neural Repair 27:742–752. https://doi.org/10.1177/1545968313491004

Yingyongyudha A, Saengsirisuwan V, Panichaporn W, Boonsinsukh R (2016) The Mini-Balance Evaluation Systems Test (Mini-BESTest) demonstrates higher accuracy in identifying older adult participants with history of falls than do the BESTest, Berg Balance Scale, or Timed Up and Go Test. J Geriatr Phys Therapy 39:64–70. https://doi.org/10.1519/JPT.0000000000000050

Zampieri C, Salarian A, Carlson-Kuhta P et al (2010) The instrumented Timed Up and Go test: potential outcome measure for disease modifying therapies in Parkinson's disease. J Neurol Neurosurg Psychiatry 81:171. https://doi.org/10.1136/jnnp.2009.173740