At-home Use of a Computer-based Pointing Task Accurately and Reliably Estimates Motor Impairments

VINEET PANDEY, John A Paulson School of Engineering and Applied Sciences, Harvard University, USA
NERGIS C. KHAN, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, USA
ANOOPUM S. GUPTA, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, USA
KRZYSZTOF Z. GAJOS, John A Paulson School of Engineering and Applied Sciences, Harvard University, USA

Obtaining valid, reliable, and low-burden quantitative assessment of motor impairments is a key challenge in the longitudinal care of people with motor impairments. Assessments with specialized instruments in controlled settings produce high-quality measures of motor performance; it remains unknown if similar measures can be obtained using common technologies in the home environment. We contribute to this body of research by evaluating the validity, reliability, and acceptability of at-home use of a recent computer-based system, called Hevelius, for quantifying motor impairments in the dominant arm. Hevelius presents pointing tasks, computes 32 measures from users' mouse movement trajectories, and reports age-specific z-scores; the z-scores separate the effects of a motor impairment from the effects of development/aging. In our study with participants with a pediatric movement disorder, Hevelius measures from at-home use demonstrate strong test-retest reliability (ICC = 0.9) and produce estimated severity scores correlated with clinician-assigned severity scores (r=0.67). Additionally, the participants and their caregivers found the tool simple to use. Our results highlight the promise of obtaining reliable quantitative assessments of motor impairments in unsupervised settings.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: remote assessment, neurological disorder

ACM Reference Format:

Vineet Pandey, Nergis C. Khan, Anoopum S. Gupta, and Krzysztof Z. Gajos. 2022. At-home Use of a Computer-based Pointing Task Accurately and Reliably Estimates Motor Impairments. In *Proceedings of Woodstock '22*. ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Obtaining valid, reliable, and low-burden quantitative assessment of motor impairments is a key challenge in longitudinal clinical care, in medical research, and in design of accessible technologies [16, 36]. Technology-supported approaches for producing high-quality quantitative assessments demonstrate trade-offs in quality of measurements and costs/efforts; they are promising when performed under researcher supervision in controlled settings but produce data is noisy and difficult to interpret in natural environments. At home assessments can be cheaper and more accessible; furthermore, such assessments might reflect behavior in more realistic settings than the lab [13]. However, the measurements might be of low quality due to imperfect compliance with instructions, possibility of interruptions, and diversity of hardware and contextual settings [8, 13]. This paper tackles the problem of collecting valid and reliable measures of motor impairments without researcher supervision in real-world environments.

Manuscript submitted to ACM

To address challenges with compliance and motivation at home, passive tracking efforts capture and quantify natural use by instrumenting computer use and collecting data from smartphone sensors and specialized hardware [16]. While these approaches reduce participant burden, they might still require expensive devices–such as specialized sensors–or produce data that are difficult to interpret [26]. Thus, despite considerable progress, there is still no consensus on how to perform high quality assessments of motor impairments in natural environments like home. We contribute to this body of research by evaluating the validity, reliability, and acceptability of at-home use of a recent computer-based system, called Hevelius, for quantifying motor impairments in the dominant arm.

Hevelius presents people with pointing tasks and analyzes the trajectories of the mouse movements they perform. Hevelius computes 32 measures from the movement trajectories, many of which have been introduced or used by HCI researchers [13, 14]. Unlike previous approaches, Hevelius reports measures as age-specific z-scores computed in comparison to a normative data set obtained from more than 200,000 healthy controls. Because motor abilities change substantially throughout a person's lifetime, the age-specific z-scores reported by Hevelius make it possible to separate the effects of a disease from the effects of development or aging. To relate the measures produced by Hevelius to clinically-meaningful ground truth, a regression model has been developed and validated for estimating the severity scores of people with ataxia (using the Brief Ataxia Rating Scale, or BARS) from the measures reported by Hevelius.

We conducted a study involving 18 children with Ataxia-telangiectasia (A-T) and 14 healthy children. The children with A-T were assessed by a clinician. They also used Hevelius once on researcher-provided equipment and under researcher supervision. Subsequently, the children used Hevelius at home–on their own computers and without researcher supervision–once a week for up to 12 weeks. The data obtained at home demonstrate strong test-retest reliability (ICC = 0.9). Severity scores estimated by Hevelius from at-home sessions correlated with severity scores assigned by the clinician as well as the severity scores estimated from the sessions when the children used Hevelius under researcher supervision. These findings demonstrate that unsupervised use of Hevelius produces data that are as valid as the data obtained in supervised settings. Lastly, people with motor impairments and their caregivers found the tool acceptable. Taken together, our results contribute evidence that it is possible to obtain valid and reliable quantitative assessments of motor impairments in unsupervised settings.

This submission is an original contribution that builds on prior work of using the Hevelius system under researcher supervision [9].

2 RELATED WORK

In this section, we summarize research from technology-supported approaches to quantify motor impairments across different settings.

2.1 Quantitative assessments of motor impairments across lab and home

Accurate and reliable assessments of motor impairments can benefit clinical work, medical research, and the design of accessible technologies [11, 16, 36]. Technology-supported assessments in lab/clinical settings have produced high quality measurements. Wearable sensors track movement data using accelerometer (for movements), gyroscope (swinging, turning), and electrocardiography (heart rate/rhythm to classify intense vs light exercises) [23]. Finger tapping tasks on touchscreens can distinguish people with Parkinson's Disease from healthy controls [3]. While such approaches produce quantitative measures of motor impairment, they rely on expensive hardware (like sensors) and researcher supervision. At home assessments can be cheaper and more accessible; furthermore, such assessments might reflect "natural" behavior [13]. However, in the absence of researcher supervision and controlled settings, the measurements might be of low quality. People might not comply with the instructions, come across interruptions, or use devices in settings that don't yield useful data [14].

To reduce participant burden, common consumer devices have been used to create quantitative representations of motor impairments. Passive tracking captures and quantifies natural performance by instrumenting computer use, tracking smartphone sensors, and using specialized hardware [16]. Such efforts–an instance of digital phenotyping [16, 27]–have expanded our understanding of movement disorders in the real world. Keystrokes derived from typing on a laptop identified response to dopamine therapy [25]. More recently, a single wrist sensor has been shown to provide accurate, reliable, and interpretable information about the severity of motor impairments in a rare pediatric disorder [18]. Such approaches reduce logistical and cognitive burdens: participants don't need to visit a lab/clinic or alter their behavior. Since the data is tracked during free-living activities, however, it might be noisy and difficult to interpret [26]. Additionally, such approaches might still require devices or sensors that are not easily accessible to many. To the best of our knowledge, there is no consnsus on how to perform high-quality assessment of motor impairments beyond supervised settings with readily accessible technologies. We contribute to this body of research by evaluating the validity, reliability, and acceptability of at-home use of a recent computer-based system, called Hevelius, for quantifying motor impairments in the dominant arm.

2.2 Pointing tasks assess motor impairments in controlled settings

Cost-effective, frequent, and unobtrusive assessments of motor performance during everyday computer use is an important goal in accessibility research [13]. One common technique for quantifying people's motor performance while using a device is analyzing a user's performance on *pointing tasks* where people move the cursor to indicate a particular target [13]. For personal computers, trajectories and events from such mouse-based pointing task are converted to features [6, 8]; such features are then compared across people with motor impairments and age-matched healthy controls. Prior research has developed multiple measures that discriminate between healthy controls and people with motor impairments. For instance, individuals with Mild cognitive impairment (MCI) demonstrated fewer total mouse movements and greater variability in duration of pauses between mouse movements [32]. Velocity profile (speed, acceleration, and jerk) can help distinguish deliberate, targeted pointer movements from "noisy" ones that are more common in those with motor impairments [8]. Additionally, participants with motor impairments make more curved and looped movements in their mouse trajectories. Many such features are naturally informative for clinical assessments of motor impairments. For instance, participants with a motor phenotype called ataxia make oscillatory movements during the finger to nose test; such movement results from over- and undershoot of the target and is called dysmetria.

While quantifying and using features from pointing tasks have been useful in the lab setting, two bottlenecks limit such techniques' utility in assessing motor impairments more pervasively. First, such assessments have been performed under researcher supervision; the quality of unsupervised pointing tasks is less understood. People's performance on pointing interactions in natural settings differs substantially from those in lab settings [8]. Participant's motor abilities might change over time due to "medication, fatigue, and changes in the underlying medical condition" [14]. Performance might even change across usage device: user interface evaluations with paid remote participants yield the same conclusions as in-lab studies for desktop interfaces [19], but not necessarily for mobile interfaces [7].

Second, such studies require age-matched healthy controls. Because motor abilities change substantially throughout a person's lifetime, data from age-matched health controls enables separating the effects of a disorder from the effects of development or aging. More recently, supervised assessments with a computer-based system, called Hevelius, have quantified motor impairments in the dominant arm [9]. Hevelius presents participants with pointing tasks and analyzes the trajectories of the mouse movements they perform. Hevelius computes 32 measures from the movement trajectories, many of which–e.g. changes in movement direction, noise-to-force ratio–have been introduced or used by HCI researchers [24, 34]. Unlike previous approaches, Hevelius reports measures as age-specific z-scores computed in comparison to a normative data set obtained from more than 200,000 healthy controls. The age-specific z-scores reported by Hevelius make it possible to separate the effects of a disease from the effects of development or aging. In terms of study pragmatics, age-specific z-scores reduce the burden on unsupervised assessments by removing the need to find age-matched healthy controls.

2.3 At-home use of Hevelius by participants with a rare neurological disorder

Ataxia-telangiectasia (A-T) is a progressive, life-limiting neurological disorder. Typically apparent during childhood, this disorder is characterized by impaired coordination of movement, impaired immunity, increased cancer risk, and telangiectasias (small widened blood vessels). Since it affects multiple body systems, A-T requires a complex care team comprising multiple specialists making it extra daunting for caregivers to both understand and manage the condition.

Valid, reliable, and low-effort at-home assessments of motor impairment can potentially benefit A-T families. A-T is a *rare disorder*; rare disorders are disorders that affect fewer than 200,000 people. This quantitative distinction in the number of people leads to differences in the availability of experts (both in numbers and location), quality of care, general awareness about the condition, and the current state of research [15]. For instance, a clinical trial for a rare disorder enrolled 39 people in 10 years [21]. At-home assessments can potentially improve people's access to better medical research.

However, there are three challenges in conducting at-home assessments with pointing tasks for participants with A-T. First, since A-T is a pediatric disorder, any assessment needs to be appropriate for cognitive and motor skills of children. Since the assessments happen without expert supervision, any issues due to a mismatch between participants' abilities and tool's expectations cannot be debugged or fixed in real time. Second, participants in A-T show a broad range of symptoms and performance. Third, younger participants might lack the necessary knowledge or articulation capabilities to share contextual factors (like mood or interruptions) that might be needed to accurately assess their performance. Therefore, a tool for assessing motor impairments in unsupervised settings (like home) would need to be robust and easy-to-use; support ability-appropriate customization by the user; and support reports from adult caregivers.

While Hevelius' utility under researcher supervision is promising, the value and acceptability of Hevelius for unsupervised sessions is unknown. Specifically, we ask four questions:

- Are Hevelius' measures reliable over unsupervised sessions?
- How well do measures from unsupervised sessions estimate existing ground truth?
- How well do estimates over the measures compare across supervised and unsupervised sessions?
- Do participants and caregivers find the tool acceptable to use without researcher supervision?

We contribute a rigorous evaluation of the reliability, validity, and acceptability of unsupervised assessments for pointing problems with a rare disorder community.

3 HEVELIUS FOR AT-HOME ASSESSMENTS

Hevelius comprises a computer mouse-based tool that provides objective, granular, interpretable, quantification of motor impairment in the dominant arm with a few minutes of use [9]. Hevelius presents participants with pointing tasks, collects movement data, and computes 32 measures (Table 3). The measurements are reported as age-specific

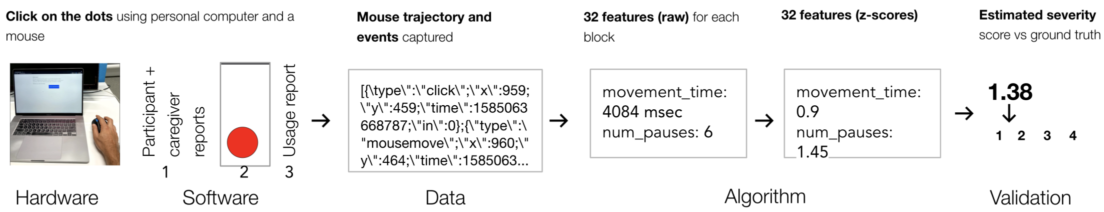

Fig. 1. Families with A-T access Hevelius at home using a mouse and a browser on a personal computer. Activities comprise pointing tasks and reports from participant & caregivers. Supplementary material provides the complete list of 32 Hevelius measures. Figure inspired by Figure 3 in [35].

z-scores by comparing them to baseline data collected from health volunteers of the same age. Hevelius is accessed using an online browser on a personal computer and the pointing tasks are performed using a mouse. Supplementary methods in a previous publication elaborate on Hevelius' design and techniques for data collection and processing [9]. In this section, we summarize Hevelius' pointing task and movement data processing.

3.1 Pointing task

Participants use the mouse to click on a target (presented as a red circle) as accurately and quickly as they can after it appears. When clicked correctly, the clicked target is replaced with another target. Clicking on one target comprises a complete trial; a sequence of nine targets constitutes a block. Participants can take a break between blocks; they're instructed to completed each block without interruption. Targets in a block are of the same size ; targets across blocks are selected over multiple target sizes. The distance between successive targets in a block is the same. Four combinations (repeated once each) of distances (A) and target sizes (W) are included in the pointing task in the same order across sessions; (A,W) = (30, 90), (16, 240), (90, 270), (26, 390), (16, 240), (30, 90), (26, 390), (90, 270).

3.2 Movement Data Processing

Participants' mouse movement is represented as basic movement statistics (location of endpoints, timing) as well as detailed movement trajectories. Each movement is decomposed into several components including initiation (from the target onset to the first mouse move event), execution (from first to last mouse move event), verification (time spent inside the target between last mouse move event and the start of the click), and click (mouse down to mouse up event). 32 measures computed from the movement components are converted to age-specific z-scores by comparing the movement features to those from healthy online volunteers of the same age.

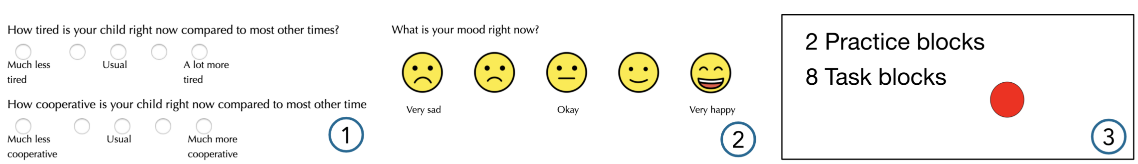

Fig. 2. Summary of At-home version of Hevelius. 1) Caregivers provide reports on participants' well-being as a 5-point Likert scale and numeric responses. Prompts included tiredness and degree of co-operation; and a numeric response scale for significant events (stumbling or tripping) since last use of Hevelius. 2) Participants provide self-reports for mood, alertness, and sleep; and 3) Participants perform pointing tasks. Pointing tasks comprised two practice blocks and eight task blocks.

3.3 Instrumenting Hevelius for at-home assessments

Four components were added to the in-clinic version of Hevelius. Two components–caregiver reports and participant self-reports–are presented to the participant before the pointing task (Figure 2). Caregiver reports sought responses about participants' well-being; prompts included 5-point scale about tiredness and degree of co-operation; and a numeric response scale for significant events (stumbling or tripping) since last use of Hevelius. Participant self-reports sought responses about mood, alertness, and sleep on a 5-point Likert scale; the scale was shown with large face icons to make the options more accessible to children. Like the previous version, pointing tasks comprised two practice blocks and eight task blocks. Unlike the previous version of Hevelius–where the target sizes varied from 20 pixels to 60 pixels–Hevelius for at-home use had target sizes vary from 16 pixels to 90 pixels. The third component is an option for participants to select a target size as the minimum target size across all sessions. The fourth and final component is caregivers' usage reports after the pointing task; these included categorical responses (for length of the task, interruptions, and other issues) and comments (on length and ease of the task, interruptions, and technical difficulties). Hevelius is implemented in PHP, HTML, and Javascript and hosted online at [anonymized].

4 STUDY

An longitudinal study of at-home use of Hevelius evaluated whether Hevelius produces valid, reliable, low-burden measures from longitudinal use at home. Participants comprised children with A-T and healthy controls.

4.1 Research Questions

Our study addressed three research questions to understand the validity, reliability, and acceptability of Hevelius at home.

(1) Do Hevelius' measures provide reliable assessments from at-home use?
  (a) Which Hevelius' measures are reliable? How reliable are they?
  (b) Do Hevelius' measures from at-home use reliably estimate clinically-meaningful ground truth?
(2) Do Hevelius' measures provide valid assessments from at-home use?
  (a) Do Hevelius' measures from at-home use accurately estimate clinically-meaningful ground truth?
  (b) How well do Hevelius' measures from at-home use compare to supervised use for estimating clinically-meaningful ground truth?
(3) Do participants and caregivers find the tool acceptable to use without researcher supervision?
  (a) What challenges do families face in using Hevelius at home?

4.2 Methods

Our study has two components for the same study population: 1) Supervised use: using Hevelius once (on researcher-provided equipment and under researcher supervision), and 2) At-home use (on personal computers without researcher supervision) once a week for 12 weeks.

4.2.1 Supervised use. All participants with A-T completed clinical assessment and performed pointing tasks on Hevelius. Clinical assessment comprised recording video data for a neurological exam; this video was later used by a clinician to assign *severity scores* to participants' motor impairment in the dominant arm based on the Brief Ataxia Rating scale (BARS) [30]. All healthy controls were assigned BARS scores of 0 without clinical examination. While using Hevelius, participants had the choice to increase the target size in the second practice task if they felt the smallest target size

Table 1. Participant demographics. A total of 32 A-T and healthy controls were enrolled in the study. The severity scores represents the severity assigned by a clinician for the dominant arm on the Brief Ataxia Rating Scale (0–4).

Diagnosis	N	Age		Sex		Handedness		Severity Score
		Median	Range	Female	Male	Left	Right	Scale: 0-4
A-T	18	10	[4,15]	8	10	2	16	0.5-3 (M: 2.03, SD: 0.74)
Control	14	11	[4,16]	5	9	1	13	0

(16 pixels) was too small; the selected target size was used as the minimum target size across supervised and at-home assessments for the participant. Caregivers provided reports using Hevelius' prompts. Two members of the research team were present during participants' supervised use to answer any questions. At the completion of supervised use, the research team suggested families use at-home Hevelius once a week for up to 12 weeks and encouraged them to note a day and time of the week for using the tool. Researchers provided families with a USB 3 Optical Mouse [1] for at-home use; in some cases, families mentioned they were comfortable using their mouse. Caregivers were told that they could communicate with two members of the research team via emails if they faced any issues.

4.2.2 At-home use. Participants and caregivers used Hevelius without supervision on their personal computers using a mouse. A partner organization sent two emails to all participating families: 1) a reminder mail 2 weeks after their supervised use; 2) a summary of researchers' response to questions from the families. Three research team met weekly among themselves to share weekly usage data, identify outliers, and discuss usability changes to the tool. If a family did not use the tool for two weeks, the research team updated the designated contact person at rare disease foundation whose team reached out to the caregivers (over email/phone) to understand concerns (if any).

4.3 Participants

4.3.1 Approval. Written informed consent and assent were obtained from all participants prior to participation. This study was approved by the Internal Review Board at [anonymized]. Participants received a $50 American Express gift card.

4.3.2 Recruitment. Thirty-two children—eighteen with A-T, fourteen without–were enrolled in the study (Table 1). Identified healthy controls were siblings of the A-T participants. All participants were identified in partnership with the Ataxia-telangiectasia Children's Project (A-TCP) which is a 501(c)(3) nonprofit organization that supports biomedical research projects for Ataxia-telangiectasia (A-T)[2]. All children with A-T were genetically confirmed to have the disorder. Children were excluded from the study if they were younger than 4 years old, unable to perform the computer mouse task, and demonstrated another movement disorder or other condition that affects arm function or mobility. The median age of A-T and healthy controls was 10 and 11 years respectively. Three participants, one control and two with A-T, indicated that left hand was their dominant hand.

4.3.3 Data collection and analysis. Data was collected between January and September 2020. Three participants with A-T (between the ages of 4 and 10) did not receive clinical assessment; four participants with A-T did not use Hevelius in supervised setting; one participant did not use the tool at home. Of these eight participants, five were not cooperative,

[1] https://www.amazon.com/gp/product/B0029L0IM8/
[2] We refer to A-TCP as rare disease foundation in the text

tired, or resting during different activities of the supervised use; one participant was enrolled too late in the day to perform the clinical assessment; one participant did not receive clinical assessment due to escalation of the COVID-19 pandemic; and another participant–who did not use the tool at home–did not report to follow-ups. Two healthy controls did not perform at-home use. Data from an at-home session was excluded if the session produced measures that were null/outlier. An outlier is a value that does not fall in [Q1 - 1.5 IQR, Q3 + 1.5 IQR]; outlier analysis were done on estimated BARS scores. Data from the first eight at-home sessions were used in analyses; participants' attrition is shown in Figure 4A. Analyses was performed using R, Microsoft Excel, and JMP.

4.4 Measures

Hevelius features were computed from each session. Supervised use of Hevelius produced a z-score for each Hevelius feature. Median over at-home sessions' data was taken to produce a z-score for each Hevelius feature for at-home use. We developed a LASSO regression model for estimating the severity scores of participants with A-T (using the Brief Ataxia Rating Scale, or BARS, for dominant arm) from the measures reported by Hevelius.

4.4.1 Reliability of Hevelius' measures. Measures comprised

(1) the test-retest reliability of the 32 measures over at-home use for participants with A-T and healthy controls;
(2) the test-retest reliability of severity scores estimated from Hevelius' measures over at-home use for participants with A-T and healthy controls;

Reliability of estimated scores and 32 Hevelius measures was computed as Intraclass Correlation Coefficient and their 95% confidence intervals; these calculations were performed using R with irr package based on a single rating, absolute-agreement, 2-way mixed-effects model. The measures were compared four sessions at a time. Estimated scores were compared one session, two sessions, and four sessions at a time. As per common heuristics [20], we interpreted the Intraclass Correlation Coefficient (*ICC*) using the following thresholds: 1) below 0.50: poor; 2) between 0.50 and 0.75: moderate; 3) between 0.75 and 0.90: good; 4) above 0.90: excellent.

4.4.2 Validity of Hevelius' measures. Measures comprised

(1) Pearson correlation (r) between severity scores estimated from supervised use of Hevelius and severity scores assigned by a clinician (ground truth) **for participants with A-T**
(2) Pearson correlation (r) between severity scores estimated from at-home use of Hevelius and severity scores assigned by a clinician (ground truth) **for participants with A-T**

As per common heuristics [31], we interpreted Pearson's correlation (r) using the following thresholds: 0.00–0.10: Negligible; 0.10–0.39: Weak; 0.40–0.69: Moderate; 0.70–0.89: Strong; 0.90–1.00: Very strong.

4.4.3 Acceptability of at-home assessments for participants with A-T. Measures comprise

(1) Time taken on the pointing task;
(2) Participant self-reports about well-being (responses about mood, alertness, and sleep on a 5-point Likert scale)
(3) Caregiver reports (responses to 5-point scale about tiredness and degree of co-operation);
(4) Categorical and open-ended reports from caregivers about the length of the task, interruptions and other concerns

Table 2. Summary of at-home use of Hevelius by participants with A-T. *Severity* is the severity score assigned by a clinician. *# At-home sessions* counts the number of attempted sessions (*Attempted*) and those sessions (*Accepted*) that yielded measures that were not null/outliers; *Time taken* (measured in minutes) represents the median time taken by the participant during the pointing task over at-home sessions.

P	Age	Severity	**At-home sessions**		
id	(yrs)	(0-4)	# Attempted	# Accepted	Median Time taken (mins)
P01	06	2.0	04	03	16
P02	13	2.0	12	10	6.6
P03	07	0.5	12	12	7
P04	09	2.5	03	03	15.3
P05	09	2.0	12	12	10.4
P06	06	1.0	12	11	11.5
P07	10	1.0	09	08	7.4
P08	07	1.5	06	03	10.7
P09	15	2.0	12	11	6.3
P10	15	3.0	08	08	4.6
P11	12	2.5	12	12	10.6
P12	10	2.5	05	05	7.7
P13	11	2.5	06	04	11.6
P14	10	2.5	13	10	8.8

5 RESULTS

14 participants with A-T received a severity score from a clinician and used Hevelius at home for at least one session (median: 9 sessions; range: 3 to 12 sessions) (Table 2). 11 A-T participants received a severity score from a clinician and used Hevelius under researcher supervision. 9 healthy controls used Hevelius at home and in the supervised setting.

For participants with A-T, most at-home sessions yield measures and take a few minutes on the pointing task. Participants used the tool at home for 126 sessions of which 111 sessions (88%) yielded measures and usage reports from caregivers. Of the sessions that didn't yield measures, 06 sessions provided data that our analysis scripts couldn't parse; 04 sessions provided data that were outliers for the corresponding participants; 03 sessions comprised incomplete use; and 02 sessions had a technical glitch while participants used the tool. Caregivers reported 11 of 111 sessions as too long. The per-participant median session time for pointing tasks ranged from 4.6 minutes to 16 minutes.

5.1 Reliability of estimated scores among participants with A-T and healthy controls

5.1.1 Many features from at-home use demonstrated good test-retest reliability among both A-T and controls participants. Basic aspects of performance demonstrated good reliability among both participants with A-T and healthy controls. Such aspects include overall efficiency (measured by movement time; ICC(A,1)=0.985 (participants with A-T), 0.931 (healthy controls) and the ability to control movement speed (measured by normalized jerk; ICC(A,1)= 0.799 (participants with A-T), 0.99 (healthy controls). Features related to time showed up to be the most reliable; these included duration of the longest pause, number of pauses, movement time, verification time, click duration, and execution time (with and without pauses). Table 3 lists Hevelius measures that demonstrated excellent and good reliability for participants with A-T. Supplementary material lists the reliability of all 32 Hevelius measures for participants with A-T and healthy controls.

Table 3. Hevelius features (with moderate and high reliability) sorted in descending order of reliability for A-T participants. Intraclass Correlation is also shown for controls participants for comparison. A list of all 32 features with their reliability is provided in the Supplementary material.

#	Feature name	Feature description	ICC (A-T)	ICC (Controls)
1	Duration of the longest pause	Duration of the longest pause of 100ms or longer. If not such pause occurred, 0ms is recorded for this measure	0.988	0.899
2	Number of pauses	Number of pauses of 100ms or longer	0.987	0.906
3	Movement time	Complete movement time from target onset to the end of the successful click on the target	0.985	0.931
4	Verification time	The time interval between the end of a movement inside a target and the beginning of the click (i.e., the time when the mouse button was pressed)	0.909	0.9
5	Click duration	The time between mouse button press and release during the correct click on the target	0.907	0.94
6	Execution time	Time from the first to the last mouse movement (excluding any movement that occurred while the mouse button was pressed)	0.887	0.951
7	Click slip	Distance between the point where the mouse button was pressed down and where it was released during click on the target	0.825	0.758
8	Execution time variability	Coefficient of variation of execution times in a block of trials	0.808	0.899
9	Verification time	Standard deviation of verification times in a block of trials	0.799	0.915
10	Normalized jerk [4, 10]	$\text{normalized jerk} = \frac{(ET)^3}{v_{max}^2} \int_t \left(\frac{da}{dt}\right)^2 \mathrm{d}t$ where $\frac{da}{dt}$ is the jerk, ET is the execution time without pauses and v_{max} is the peak speed during the movement.	0.799	0.95
11	Execution time without pauses	Like execution time, but excludes pauses of 100ms or longer	0.75	0.943

5.1.2 Estimated scores from at-home assessments demonstrated moderate test-retest reliability among participants with A-T. Fig. 3A shows the per-participant distribution of the estimated scores. Moderate reliability was found between the eight sessions' estimated scores from the A-T participants when taken one at a time. The reliability improved to excellent when at-home sessions were taken four at a time (Fig. 3B);
(one session at a time) ICC(A,1) = 0.744, 95% CI [.554, .905]; (two sessions at a time) ICC(A,1) = 0.87, 95% CI [.72, .958]; (four sessions at a time) ICC(A,1) = 0.9; 95% CI [.679, .972].

5.1.3 Estimated scores from at-home use demonstrated moderate to good test-retest reliability among healthy controls. Moderate reliability was found between the eight sessions' estimated scores from healthy controls. The reliability improved to excellent when at-home sessions were taken four at a time (Fig. 3B);
(one session at a time) ICC(A,1) = 0.596, 95% CI [.346, .857]; (two sessions at a time) ICC(A,1) = 0.814, 95% CI [.584, .948]; (four sessions at a time) ICC(A,1) = 0.911, 95% CI [.652, .979].

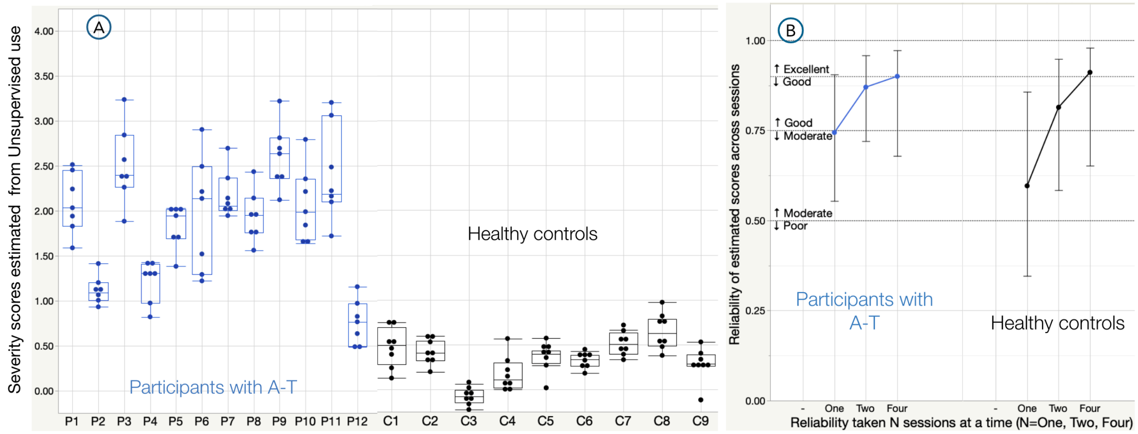

Fig. 3. (A) Participant-wise distribution of estimated scores from at-home assessments for participants with A-T and for controls (B) Test-retest reliability was moderate for participants with A-T (ICC=0.839) and for controls (ICC=0.707) participants when sessions were compared individually. Reliability improved to good for both A-T (ICC=0.855) and Controls (ICC=0.874) participants when comparisons were made four sessions at a time.

5.2 Hevelius' measures provide valid assessments from at-home use

5.2.1 Estimated scores from at-home assessments are moderately correlated with clinician-assigned scores. Estimated severity scores for at-home sessions correlated moderately (r=0.67; F(1,12) = 9.91, p<0.01) with the clinician-assigned

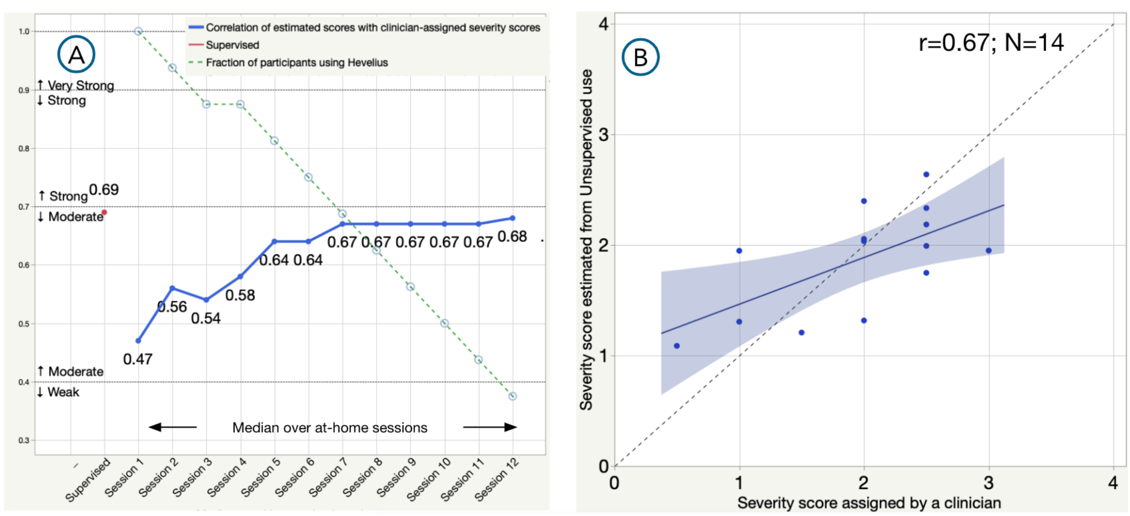

Fig. 4. Severity scores estimated from at-home assessments were compared to severity scores assigned by a clinician. (B) Estimated scores correlated moderately with clinician-assigned scores (r=0.67; F(1,12) = 9.91, p=0.0084). Scatter plot shows the data. Linear regression line is shown in blue. The dashed y=x line is shown for comparison. The shaded bands represent the 95% Confidence Interval. Pearson correlation coefficient (r) and number of participants (N) are shown. (A) Pearson's correlation between scores estimated from the first S at-home sessions (where $1 \leq S \leq 12$) and clinician-assigned severity scores. Scores estimated from more at-home assessments were better correlated with clinician-assigned severity scores. The dashed line shows the fraction of participants using the tool over at-home sessions.

severity scores for the dominant arm on the Brief Ataxia Rating Scale (Fig. 4B). The Bland-Altman plot for estimated scores and clinician-assigned scores showed mean difference of 0.13 between estimated scores and clinician-assigned scores with 95% Confidence Interval in [0.9,-1.18] (Fig. 5B - blue). Including more at-home sessions improved the correlation between estimated scores with clinician-assigned severity scores (Fig. 4A).

5.2.2 Estimated scores from at-home assessments and supervised assessments are similarly correlated with clinician-assigned scores. We found two results Clinician-assigned severity scores correlated similarly with estimated scores from at-home use (r=0.67; F(1,12) = 9.91, p=0.0084) as with supervised use (r=0.69; F(1,9) = 8.09, p=0.02) (Fig. 5A). The Bland-Altman plot for supervised sessions showed mean difference of 0.09 between estimated scores and clinician-assigned scores with 95% Confidence Interval in [1.06,-1.26] (Fig. 5B - red). Compared to at-home sessions, the mean difference is lower by 0.04 points and the 95% Confidence Interval limits is higher by 0.24 points.

5.3 Acceptability of at-home assessments for participants with A-T

We share participants' and caregivers' acceptability of using Hevelius at home. We draw on quantitative data from usage logs and qualitative data from participants' self-reports (responses about mood, alertness, and sleep on a 5-point Likert scale) and caregiver reports (responses to 5-point scale about tiredness and degree of co-operation; usage reports about issues).

5.3.1 Most at-home sessions yield measures and take a few minutes on the pointing task. The median time on pointing tasks was less than 10 minutes for 11/14 participants with A-T. 11 of 111 sessions were reported to have technical issues. Such issues included clicks not registered on the targets and participants making mis-clicks. Caregivers reported

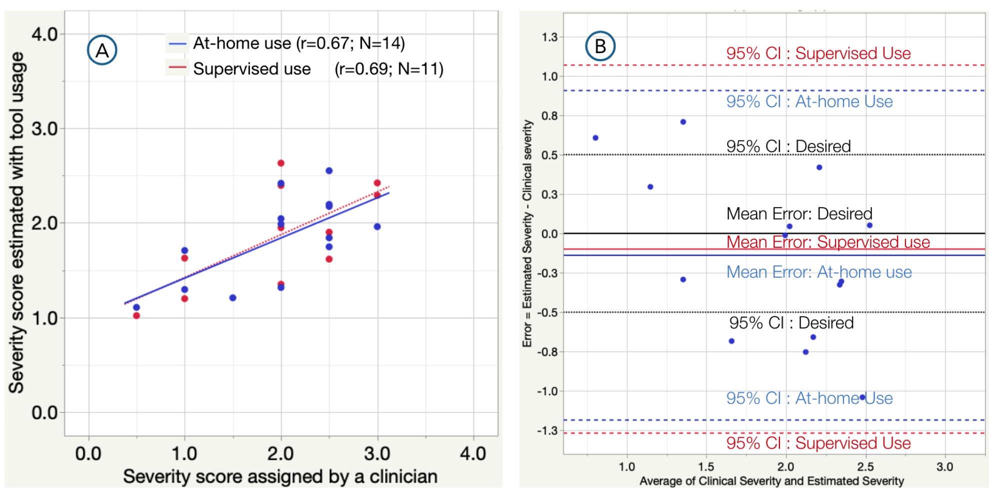

Fig. 5. (A) Severity scores estimated from supervised and at-home use were plotted against the severity score assigned by a clinician. Linear regression line is shown (blue for at-home use and red for supervised use). Pearson correlation coefficient (*r*) and number of participants (N) are shown. (B) Estimated scores and clinician-assigned severity scores were plotted using the Bland-Altman plot to visualize their difference. Scatter plot shows the data points (x,y) for at-home use; the data points were created by calculating the average and difference of the two scores. Solid lines represent the mean and dashed lines represent the 95% Confidence Interval. Blue represents at-home use; Red represents supervised use; Black represents desired mean error and 95% Confidence Interval. CI = Confidence Interval.

restarting the task to perform the task. One concern was when the targets appeared cut off on the screen; in response, the research team altered the pointing task systematically for all participants so the targets appeared in full on smaller desktop screens as well. Caregiver for P05 reported slowing the cursor speed and making the cursor bigger.

5.3.2 Participants developed strategies to perform the task. In 22 of 111 sessions, participants were reported to be altering their sitting posture while performing the pointing task. Some common heuristics emerged: participants used their non-dominant hand (hand not used for the pointing task) to stabilize themselves. Caregivers reported–sometimes across multiple sessions–that participants used their non-dominant hand to brace themselves on the on chair/bench they were sitting on (P3); to steady the wrist of the main hand (P6); to hold the table on which the laptop with the task was used (P5). Another strategy was to lean in closer to the laptop; 3 Caregivers reported that participants leaned forward to the screen (presumably) to see more clearly, especially for smaller targets.

5.3.3 Challenges in using the tool. One caregiver added a note to "reduce practice time and number of rounds" (P11, Session 2); caregiver for two participants requested "less dots" as well. Caregivers for 3 participants also reported the participants' frustration while clicking the targets with the mouse.

> "[Participant] move (sic) the mouse back and forth across the screen in frustration, as I'm sure you'll see in the results" (P3, Session 10)
>
> "Bigger dots would be helpful and less of them. Also the mouse is difficult to use, as his fingers keep hitting the rolling piece in the middle which causes google chrome to ask if we want to close out the program. The combination becomes frustrating for him." (P13, Session 4)

03 families also mentioned health and lifestyle challenges in using the tool regularly; lifestyle concerns included travel and sports tournaments. Some caregivers provided suggestions to improve the assessment. Caregiver(s) for P13 suggested numerous ways to improve the task such as including audio feedback and gamifying the task with a pointer as a super mario kart driving to the target.

5.3.4 Participant self-reports and caregiver reports about health and well-being did not better explain the variation in at-home assessments. A goodness of fit analysis estimated the proportion of variance explained in estimated severity scores by models that included three parameters: a unique participant code (nominal), at-home session number (ordinal), and caregiver & participant responses (ordinal) in that order. R^2 values demonstrate participant code explained 73.7% of the variance in estimated severity scores; adding the session number accounted for 76.2% of the variance. Further including caregiver and participant responses accounted for 79.5% of the variance.

5.3.5 Caregivers shared details about participant. Some caregivers provided additional details about the participant' health in response to a prompt about sharing current events for the participants.

> "[Participant] was standing still and fell at drama class and bruised left leg; did not wear CPAP 2 nights this week, had a sleep study with titration 1 night additional; Blood draw for bi-annual immune panel, 10 vials blood drawn." (P2, Session 3)
>
> "He still has some nasal congestion and cough, is doing breathing treatments to keep his lungs clear." (P5, Session 8)
>
> "She seems to have a cold with nasal congestion. She also had a low-grade fever off and on all week. " (P6, Session 2)

Some caregivers shared sotheories about performance and compensatory strategies

> P3, S10: "She has been using the mouse more this week for online games like ABC Mouse. I'm not sure if she is just over-tired or if the methods she uses for clicking on those games has changed her"
> "Lots of exhaustion and wobbling this week. Not sure if we're seeing a regression in her condition or if she's just extra tired from all of the activities we have been up to." (P3, Session 11)

6 DISCUSSION

Cost-effective, frequent, and unobtrusive assessments of health state is an important goal in health research [16]. The success of "lab-quality" pointing data in automatically detecting movement disorders is known [12]. While promising, lab-based assessments pose challenges too: they are effort-intensive for both researchers and participants to perform frequently. Additionally, one session of pointing tasks (in the lab or outside) does not develop realistic measures for variance of performance over time. Our results demonstrate that Hevelius' measures from unsupervised use are as good as measures from a supervised use in estimating clinically-meaningful ground truth. Furthermore, some measures are highly reliable while others are not. 1) A related concern for doctors during telemedicine encounters is identifying subtle differences in motor performance over video-based interactions [2]. At home assessments–like the ones supported by Hevelius–can enrich telemedicine by improving the specificity of such assessments without requiring clinician time.

Our results provide a conservative estimate of the value of unsupervised assessments due to multiple reasons. While the ataxia phenotype is common across many neurological disorders, the participant set of A-T demonstrates ataxia that can be more severe. Overall, our results extend the current understanding of measures of motor performance from the lab to unsupervised settings. For our setup, target sizes varied between 16 to 90 pixels. Since pixel sizes vary according to the display hardware configuration, ideally we'd require all participants to use the same monitors at home. However, compared to distributing mouse, this is a more costly exercise. The quality of our results demonstrate that even with likely heterogeneity in people's monitor configuration, the measures were valid and reliable.

6.1 At-home assessments complement supervised assessments

We believe there are two values to supervised assessments that goes beyond collecting high-quality data: i) observing participant's interactions and hacks; and ii) building trust and answering questions. In-situ participant observation is a key technique for developing a deeper understanding of people's interactions with technology [5]; the above-mentioned example of changing posture was identified with multiple participants during supervised visits.However, performing such observations in people's homes is infeasible beyond a small number of participants. Supervised use of a tool shows researchers whether users follow the prompts appropriately. During the supervised visit, the research team also fielded clarifying questions and shared details. This more *humane* aspect of supervision is important too; HCI work has shown that trust breaks down in electronic contexts [29].

Many in-clinic visits for movement disorders happen once every six months. Quantitative, unsupervised assessments can reduce the reliance on patients' and caregivers' self-reports. Additionally, at home assessments enable frequent assessments which can tackle challenges such as recency bias in patients' self-reports. Caregivers reports of patient health and updates over multiple sessions provided insights into participant mood and rest. A summary of such reports can be beneficial when in-clinic interactions are infrequent.

Any future clinical use of unsupervised assessments must design around known challenges of increased clinician and caregiver burden. For instance, prior integrative review has found that patients expect providers to constantly monitor data to dispel their doubts [28]. Another challenge is potentially increasing caregiver burden. Caregivers spend a lot of effort taking care of people with disorders [1]; depending on them for unsupervised assessments can be more challenging in more longitudinal settings. While answering these questions was out of scope for our research, we suggest developing heuristics to better integrate unsupervised assessments in clinical processes.

6.2 Technical challenges with unsupervised assessments

We reflect on the challenges of conducting at-home assessments and how they compare to supervised assessments. Our tool design was successful in being straightforward to use for participants and caregivers. We see some possibilities of improvement based on participants' interaction with the tool: 1) *first session issues*; 2) outlier data; and 3) choice of technical infrastructure. We suspect that the peculiar dip in the quality of data for unsupervised session#1 (r=0.47; Figure 4A) might have been due to people's challenges using the tool for the first time by themselves at home. Further work can investigate whether such "first session issues" are common in remote deployments; if so, how might health researchers reduce such challenges? One approach could include providing remote supervision for the first session and not using data from the first session. The presence of outlier at-home sessions for many participants suggest that the interaction between the ataxia phenotype, the severity and the tool requires more work. Performing longitudinal assessments is more important in this case. Since some families reported concerns with participants' lack of familiarity with mouse, future research can consider exploring technical infrastructures and input devices (like touchscreens) that might be more suitable to pediatric population.

6.2.1 Challenges of lived experience. Successful data tracking in real-world settings demonstrates tension between scientific validity and lived experience [17]. Our data was collected from eight weeks of tool use. Caregivers shared that travel and sports tournaments led to skipping some sessions. For A-T participants in our study, eight weeks isn't long enough to see progression. However, there are other effects that at-home assessments should tackle; these include learning effects (where participants learn the task with frequent use leading to performance that is less representative of the underlying conditions) and compensatory effects, where participants create motor and cognitive strategies that compensate for the challenges in motor performance [22]. For instance, the primary author observed that participants altered their sitting posture while using Hevelius. After adding this as a prompt for caregiver reports during at-home use, the trend became clearer: in 20% of the sessions, participants demonstrated such compensatory strategies (e.g. leaning in to the monitor or bracing their non-dominant hand on furniture).

6.3 Support from partner organizations like rare disease foundations

Sociotechnical systems can amplify the efforts of committed and knowledgeable individuals [33]. However, finding such individuals, getting them started, and keeping them invested requires complementary efforts. Additionally, prior research has noted the difficulty of finding participants for clinical trials for rare disroders [21]. Overall, Hevelius' success in collecting useful measures from unsupervised settings relied on both the design of the tool and our partners, specifically the rare disease foundation A-T Children's Project. Having worked with the patient community and multiple researchers, the rare disease foundation provided multiple contributions that are difficult to achieve for a small research group. First, the rare disease foundation assisted the research team in publicizing the study and finding participants. Second, the supervised use of our tool happened at a gathering of families organized by the rare disease foundation. Meeting multiple knowledgeable and interested families in one place would have otherwise been challenging. Additionally, such face-to-face interactions can also improve trust in electronic contexts [29]. Third, the foundation reminded participants when they did not use Hevelius (after receiving usage updates from the research team). For cases where such patient-researcher understanding/trust might be more nascent, not visible, or clearly lacking, researchers can better direct efforts towards better understanding the community. As our experience suggests, such trust-building and knowledge sharing can be mediated and accelerated by trusted organizations.

7 CONCLUSION

This paper reports an at-home longitudinal study of a recent computer-based system, called Hevelius, that quantifies motor impairments in the dominant arm [9]. Hevelius computes 32 measures from the movement trajectories and does not require expert judgement or age-matched healthy controls to compute severity scores. Our findings demonstrate that unsupervised use of Hevelius produces data that are as valid as the data obtained in supervised settings. Further, the data obtained at home demonstrate strong test-retest reliability. Lastly, participants and their caregivers contributed data without major concerns. Taken together, our results contribute evidence that it is possible to obtain valid and reliable quantitative assessments of motor impairments in unsupervised settings.

ACKNOWLEDGMENTS

We thank Jennifer Thornton and Sara Reiling from A-TCP for helping the research team recruit families for the study and communicating with them. We thank Aleksandra Koralczyk for conducting an interview with some families. Vineet thanks Herman Saksono for his inputs framing this work for an HCI audience. This research was funded in part by Sony, A-TCP, Biogen, and NIH.

REFERENCES

[1] Ronald D Adelman, Lyubov L Tmanova, Diana Delgado, Sarah Dion, and Mark S Lachs. 2014. Caregiver burden: a clinical review. *Jama* 311, 10 (2014), 1052–1060.

[2] [Anonymized]. [n.d.]. Understanding Clinician Perspectives to Identify Opportunities for Telemedicine Beyond COVID-19. In *Submission at CHI'22*.

[3] Teresa Arroyo-Gallego, María Jesus Ledesma-Carbayo, Alvaro Sánchez-Ferro, Ian Butterworth, Carlos S Mendoza, Michele Matarazzo, Paloma Montero, Roberto López-Blanco, Veronica Puertas-Martin, Rocio Trincado, et al. 2017. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. *IEEE Transactions on Biomedical Engineering* 64, 9 (2017), 1994–2002.

[4] Sivakumar Balasubramanian, Alejandro Melendez-Calderon, and Etienne Burdet. 2011. A robust and sensitive metric for quantifying movement smoothness. *IEEE transactions on biomedical engineering* 59, 8 (2011), 2126–2136.

[5] Joel Brandt, Philip J Guo, Joel Lewenstein, Mira Dontcheva, and Scott R Klemmer. 2009. Two studies of opportunistic programming: interleaving web foraging, learning, and writing code. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1589–1598.

[6] Abigail Evans and Jacob Wobbrock. 2012. Taming wild behavior: The input observer for obtaining text entry and mouse pointing measures from everyday computer use. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1947–1956.

[7] Leah Findlater, Joan Zhang, Jon E Froehlich, and Karyn Moffatt. 2017. Differences in crowdsourced vs. lab-based mobile and desktop input performance data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6813–6824.

[8] Krzysztof Gajos, Katharina Reinecke, and Charles Herrmann. 2012. Accurate measurements of pointing performance from in situ observations. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 3157–3166.

[9] Krzysztof Z Gajos, Katharina Reinecke, Mary Donovan, Christopher D Stephen, Albert Y Hung, Jeremy D Schmahmann, and Anoopum S Gupta. 2020. Computer mouse use captures ataxia and parkinsonism, enabling accurate measurement and detection. *Movement Disorders* 35, 2 (2020), 354–358.

[10] Neville Hogan and Dagmar Sternad. 2009. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *Journal of motor behavior* 41, 6 (2009), 529–534.

[11] Gordon Holmes. 1939. The cerebellum of man. *Brain* 62, 1 (1939), 1–30.

[12] Amy Hurst, Scott E Hudson, Jennifer Mankoff, and Shari Trewin. 2008. Automatically detecting pointing performance. In *Proceedings of the 13th international conference on Intelligent user interfaces*. 11–19.

[13] Amy Hurst, Scott E Hudson, Jennifer Mankoff, and Shari Trewin. 2013. Distinguishing users by pointing performance in laboratory and real-world tasks. *ACM Transactions on Accessible Computing (TACCESS)* 5, 2 (2013), 1–27.

[14] Amy Hurst, Jennifer Mankoff, and Scott E Hudson. 2008. Understanding pointing problems in real world computing environments. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. 43–50.

[15] Maia Jacobs, Galina Gheihman, Krzysztof Z Gajos, and Anoopum S Gupta. 2019. " I think we know more than our doctors" How Primary Caregivers Manage Care Teams with Limited Disease-related Expertise. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–22.

[16] Sachin H Jain, Brian W Powers, Jared B Hawkins, and John S Brownstein. 2015. The digital phenotype. *Nature biotechnology* 33, 5 (2015), 462–463.

[17] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine Zia. 2017. Tummytrials: a feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 6850–6863.

[18] Nergis C Khan, Vineet Pandey, Krzysztof Z Gajos, and Anoopum S Gupta. 2021. Free-Living Motor Activity Monitoring in Ataxia-Telangiectasia. *The Cerebellum* (2021), 1–12.

[19] Steven Komarov, Katharina Reinecke, and Krzysztof Z Gajos. 2013. Crowdsourcing performance evaluations of user interfaces. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 207–216.

[20] Terry K Koo and Mae Y Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine* 15, 2 (2016), 155–163.

[21] Stephen W Lagakos. 2003. Clinical trials and rare diseases.

[22] Timothy D Lee, Stephan P Swinnen, and Deborah J Serrien. 1994. Cognitive effort and motor learning. *Quest* 46, 3 (1994), 328–344.

[23] Rex Liu, Albara Ah Ramli, Huanle Zhang, Esha Datta, Erik Henricson, and Xin Liu. 2021. An Overview of Human Activity Recognition Using Wearable Sensors: Healthcare and Artificial Intelligence. *arXiv preprint arXiv:2103.15990* (2021).

[24] I Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. 2001. Accuracy measures for evaluating computer pointing devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 9–16.

[25] Michele Matarazzo, Teresa Arroyo-Gallego, Paloma Montero, Verónica Puertas-Martín, Ian Butterworth, Carlos S Mendoza, María J Ledesma-Carbayo, María José Catalán, José Antonio Molina, Félix Bermejo-Pareja, et al. 2019. Remote monitoring of treatment response in Parkinson's disease: the habit of typing on a computer. *Movement Disorders* 34, 10 (2019), 1488–1495.

[26] Helena M Mentis, Anita Komlodi, Katrina Schrader, Michael Phipps, Ann Gruber-Baldini, Karen Yarbrough, and Lisa Shulman. 2017. Crafting a view of self-tracking data in the clinical visit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5800–5812.

[27] Jukka-Pekka Onnela and Scott L Rauch. 2016. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 41, 7 (2016), 1691–1696.

[28] Meghan J Reading and Jacqueline A Merrill. 2018. Converging and diverging needs between patients and providers who are collecting and using patient-generated health data: an integrative review. *Journal of the American Medical Informatics Association* 25, 6 (2018), 759–771.

[29] Elena Rocco. 1998. Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 496–502.

[30] Jeremy D Schmahmann, Raquel Gardner, Jason MacMore, and Mark G Vangel. 2009. Development of a brief ataxia rating scale (BARS) based on a modified form of the ICARS. *Movement Disorders* 24, 12 (2009), 1820–1828.

[31] Patrick Schober, Christa Boer, and Lothar A Schwarte. 2018. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia* 126, 5 (2018), 1763–1768.

[32] Adriana Seelye, Stuart Hagler, Nora Mattek, Diane B Howieson, Katherine Wild, Hiroko H Dodge, and Jeffrey A Kaye. 2015. Computer mouse movement patterns: A potential marker of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 1, 4 (2015), 472–480.

[33] Kentaro Toyama. 2015. *Geek heresy: Rescuing social change from the cult of technology.* PublicAffairs.

[34] Neff Walker, David A Philbin, and Arthur D Fisk. 1997. Age-related differences in movement control: Adjusting submovement structure to optimize performance. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 52, 1 (1997), P40–P53.

[35] Edward Jay Wang, Junyi Zhu, Mohit Jain, Tien-Jui Lee, Elliot Saba, Lama Nachman, and Shwetak N Patel. 2018. Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera. In *Proceedings of the 2018 CHI conference on human factors in computing Systems*. 1–9.

[36] Jacob O Wobbrock, Shaun K Kane, Krzysztof Z Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-based design: Concept, principles and examples. *ACM Transactions on Accessible Computing (TACCESS)* 3, 3 (2011), 1–27.