Hevelius Report: Visualizing Web-Based Mobility Test Data For Clinical Decision and Learning Support

Hongjin Lin
School of Engineering and Applied Sciences
Harvard University
Allston, Massachusetts, United States
hongjin_lin@g.harvard.edu

Tessa Han
Bioinformatics and Integrative Genomics
Harvard University
Cambridge, Massachusetts, United States
than@g.harvard.edu

Krzysztof Z. Gajos
School of Engineering and Applied Sciences
Harvard University
Allston, Massachusetts, United States
kgajos@g.harvard.edu

Anoopum S. Gupta
Neurology
Massachusetts General Hospital
Boston, Massachusetts, United States
agupta@mgh.harvard.edu

Abstract

Hevelius, a web-based computer mouse test, measures arm movement and has been shown to accurately evaluate severity for patients with Parkinson's disease and ataxias. A *Hevelius* session produces 32 numeric features, which may be hard to interpret, especially in time-constrained clinical settings. This work aims to support clinicians (and other stakeholders) in interpreting and connecting *Hevelius* features to clinical concepts. Through an iterative design process, we developed a visualization tool (*Hevelius Report*) that (1) abstracts six clinically relevant concepts from 32 features, (2) visualizes patient test results, and compares them to results from healthy controls and other patients, and (3) is an interactive app to meet the specific needs in different usage scenarios. Then, we conducted a preliminary user study through an online interview with three clinicians who were *not* involved in the project. They expressed interest in using *Hevelius Report*, especially for identifying subtle changes in their patients' mobility that are hard to capture with existing clinical tests. Future work will integrate the visualization tool into the current clinical workflow of a neurology team and conduct systematic evaluations of the tool's usefulness, usability, and effectiveness. *Hevelius Report* represents a promising solution for analyzing fine-motor test results and monitoring patients' conditions and progressions.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools**; **Visualization systems and tools**.

Keywords

Digital phenotyping; Clinical decision-making; Mobility impairment; Parkinson's disease; Ataxia

1 Introduction

About one out of ten people in the U.S. experiences some form of mobility impairment [15], ranging from difficulties with gross motor skills (e.g., walking) to fine motor skills (e.g., picking up an object by hand). Clinical support, medical research, and treatment design for patients with mobility impairments rely on accurate assessments of motor ability. These assessments usually entail infrequent clinician-performed evaluations in a clinic setting. However, monitoring progression calls for more timely, frequent, and convenient assessments. *Hevelius* [10], a web-based computer mouse test, was developed to quantify dominant arm mobility. It has been shown to accurately measure disease severity for patients with Parkinson's disease and ataxias [10, 22]. A *Hevelius* session spans only a few minutes and can be conducted at home [7, 22]. Participants are asked to perform up to 8 sets of 9 pointing movements using a computer mouse. At the end of each session, *Hevelius* produces 32 features related to mouse movement profiles.

To integrate *Hevelius* into clinicians' existing workflow for decision-making and research, we aim to reduce the cognitive burden required to interpret the movement data as much as possible. Taken together, our work makes the following contributions.

First, working closely with a practicing clinician-engineer, we speculated three specific obstacles clinicians may face when interpreting *Hevelius* results and three potential usage scenarios. The three main obstacles are: 1) connecting the 32 fine-grained features to interpretable and clinically relevant concepts, 2) visualizing the features in digestible graphical forms, and 3) accessing and exploring the data in different user scenarios quickly and conveniently. The three main usage scenarios are: 1) reviewing a patient's data a few minutes *before* a clinical appointment, 2) discussing the visualizations with patients *during* the appointment (a typical clinic visit

(a) Summary plots   (b) Trajectory plots   (c) Speed plots   (d) Additional information
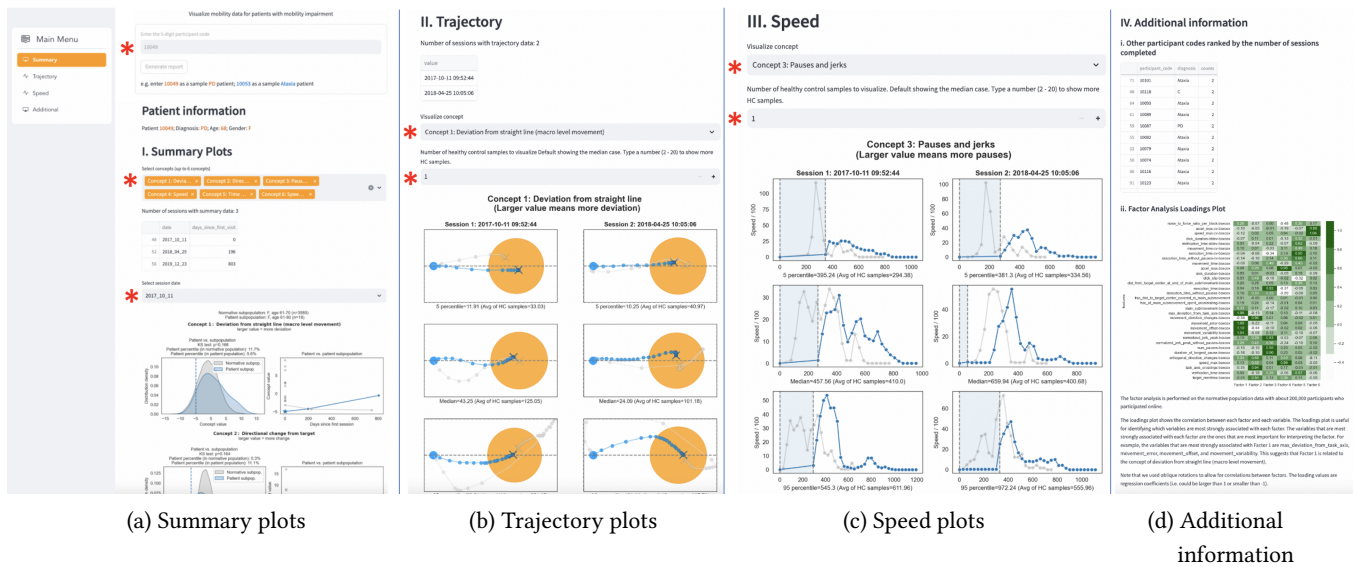
Figure 1: Interactive visualization tool (*Hevelius Report*) developed to help users analyze mobility test results. After the user enters a patient's unique code, the tool provides a report of the patient's test results with (a) summary plots, (b) trajectory plots, (c) speed plots, and (d) additional information on the analyses. The interface is a single continuous page with a sidebar that allows users to jump to a particular section. The red asterisks (annotations in the figure; not present in the tool) indicate opportunities for user interactions.

lasts 30–60 minutes), and 3) taking time to learn about the patient's data more deeply *after* the appointment.

Second, to address each of these obstacles and support the three scenarios, we developed a secure web-based visualization platform *Hevelius Report* (Figure 1) that (1) extracts six clinically relevant concepts from 32 features with factor analysis, (2) visualizes various aspects of test results (concepts, raw movement trajectories, and speed profile) using the "small multiple" concept from data visualization [31], enabling comparisons to results from healthy controls, other patients, and the patient themself over time, and (3) is interactive to meet specific needs in different user scenarios.

Third, we conducted one-hour semi-structured interviews with three additional clinicians not involved in the project to obtain preliminary user feedback on the platform. They affirmed the speculated obstacles and usage scenarios, expressing interest in learning about *Hevelius* and using the *Hevelius Report*. The combination of raw movement trajectories presented in the "small multiple" format and summary plots could effectively assist clinicians in identifying subtle changes not apparent in existing clinical tests.

Our iterative design process, visualization techniques, and findings on clinicians' needs and perspectives could be generalized to other fine-motor test results for clinical decision-support. In future work, we plan to integrate the visualization tool into the current clinical workflow of a neurology team and conduct systematic evaluations of the tool's usefulness, usability, and effectiveness.

2 Related Work

Mobility test visualizations vary based on the motor impairments and clinical environments. We focused our literature review on Parkinson's disease, which comes with some of the most common mobility problems like rigidity, slowness, and tremor [4].

Jusufi et al. [17] visualized spiral drawing data of Parkinson patients collected using a telemetry touch screen device. The authors showed the raw spiral trajectory data alongside two time-series graphs showing drawing speed and deviation from the desired trajectory, suggesting a (manual) effort to reduce the raw trajectory data to two concepts: speed and position. Hixson et al. [14] designed a data visualization platform to represent the Movement Disorder Society Unified Parkinson's Disease Rating Scale Results (MDS-UPDRS). The authors displayed the MDS-UPDRS scores on human anatomical locations, which made it straightforward for clinicians to identify areas of the patient's body that need more treatment attention.

Prior works have also visualized the 3D movement of Parkinson's patients. For example, Piro et al. [23] visualized 3D motion data using a computer vision approach to animate patients' movements while they completed a hand movement task in a clinical setting. Synnott et al. [28] visualized 3D trajectory data of patients completing daily activities like raising a cup. Jombík et al. [16] displayed tremor planes in 3D to find spatial features that distinguish tremors from Parkinson's patients from healthy volunteers.

Our work extends existing visualization efforts by providing data-driven (instead of manual) concept extraction of a web-based mobility test without requiring a specialized physical device. While our work is based on *Hevelius*, our findings and visualizations are generalizable to other tests that track fine motor movement trajectories.

3 HCI design frameworks and collaboration

Our team comprises Human-Computer Interaction (HCI) scholars and a practicing neurologist-engineer, who is motivated to integrate *Hevelius* into his clinical workflow. We held weekly meetings to discuss clinician needs, clinical workflows, and prototype feedback. We speculated the specific obstacles that clinicians may face when interpreting *Hevelius* results:

- **Obstacle 1. Clinicians may struggle to connect the 32 fine-grained features to clinical concepts.** Example features include movement offset, movement variability [19], and normalized jerk [5] (see [22] and Appendix C for the full list). These features are too detailed and are not directly related to clinical aspects of movement disorders. This problem is exacerbated by time constraints.
- **Obstacle 2. Clinicians may find interpreting *Hevelius* test results in their current numerical form difficult.** Clinicians may want to visualize these numerical values in graphical forms.
- **Obstacle 3. Clinicians may find it difficult to access and explore the data conveniently and quickly.** The raw *Hevelius* data is currently stored in CSV files. Clinicians may want to obtain the results more conveniently (i.e., without coding) and more quickly (i.e., without performing each analysis step themselves).

We also identified three main usage scenarios. First, clinicians may review patients' data *a few minutes before a clinical appointment*, allowing them to grasp high-level information about their patients' arm mobility. Second, clinicians may discuss the visualization report with patients *during the appointment* (a typical clinic visit lasts 30-60 minutes). Lastly, they may take more time to digest the report *after the appointment*. These different usage scenarios motivate flexibility and interactivity to address distinct needs from different scenarios.

Furthermore, we co-determined the design values and revisited them throughout the process, following the *Value Sensitive Design* framework [8, 11]. They were 1) *convey useful information* from *Hevelius* for diagnosis, severity measurement, and progression tracking, 2) *facilitate learning* for clinicians (and patients) about disease progression and treatment, and 3) *be mindful* about how certain visualizations might make patients feel discouraged and confused. While we focus on clinicians' needs in this work, an ongoing study from our team seeks patients' perspectives to update this set of values (and our work) to better serve the patients as key stakeholders [27].

Lastly, we presented multiple designs to solicit informal feedback in our weekly meetings with our clinician-engineer collaborator, following a *Parallel Design* principle [6, 30].

4 Data

We used two data sources in this work. One dataset was collected from 247,667 healthy control participants in a web-based study hosted on *LabintheWild* [24] in 2013. We used this dataset for factor analysis to extract clinically relevant concepts. Another dataset was collected from 286 participants from a clinical study (2016–2020) conducted by the clinician in our team, with 124 of the participants diagnosed with ataxias, 61 diagnosed with Parkinson's, 59 of them serving as healthy controls, and the rest with other diagnoses. About 70 (24%) participants completed more than one session. We used this dataset to develop visualizations for individual patients. Both datasets were anonymized before the authors could access them.

5 Factor Analysis and Data Visualizations

We describe three aspects of the visualization tool, each addressing an obstacle described in Section 3. We further describe other solutions explored and explain why the chosen solution is optimal in Appendix B.

First, the visualization tool extracts high-level, clinically relevant concepts. Addressing the obstacle of interpreting fine-grained features (Obstacle 1), we performed factor analysis [29] on 32 *Hevelius* features and obtained six higher-level concepts, using the *factor_analyzer* Python package [1]. We implemented the *promax* rotation (an oblique rotation method) to allow the factors to be correlated since many clinical concepts, such as speed and accuracy of movement trajectories, are correlated [33].

Figure 2 (Appendix A) represents the loading between each factor and *Hevelius* feature. Features with the highest loading values for a factor are the most salient for interpreting the factor. For example, the variables that are most strongly associated with Factor 1 are the maximum deviation from the task axis, movement errors, movement offset, and movement variability. This list of variables led us to infer that Factor 1 is related to deviation from a straight line. We mapped each factor to a concept relevant to aspects of movement that clinicians are likely to assess when examining patients' motor skills. The resulting six concepts are 1) deviation from the straight line, 2) directional change from the target, 3) pauses and jerks, 4) speed, 5) time inconsistency, and 6) speed inconsistency. Two categories emerged from the six concepts: the first two are related to the computer mouse trajectory, while the latter four are related to time and speed. We are conducting additional analysis to investigate the statistical relationship between these six concepts and clinical scores like the Brief Ataxia Rating Scale (BARS) [25] and Patient-Reported Outcome Measure of Ataxia (PROM-Ataxia) [26].

Second, we visualized various aspects of *Hevelius* test results (concepts, trajectory profile, and speed profile), addressing the obstacle of interpreting the data in its current numerical form (Obstacle 2). First, we created visualizations to summarize information about each of the six concepts (**"summary plots"**). An example summary plot for the "deviation from the straight line" concept is shown in Figure 3 (Appendix A). We overlaid the patient's value over the distributions of comparable healthy controls (i.e., same gender and age group) and comparable patients (i.e., same diagnoses, gender, and age group) (Figure 3, left). We also plotted the patient's values and those of comparable patients over time (Figure 3, right). These visualizations enable clinicians to compare the patient to healthy controls and other patients and track the patient's progression.

Next, we visualized how the patient moved the computer mouse to the target (**"trajectory plots"**), using the "small multiple" concept from data visualization [31]. An example trajectory plot is shown in Figure 4 (Appendix A). For a given concept, we identified the 5*th*, 50*th*, and 95*th* percentile of trials of the patient (in blue)

and the healthy control population (in grey). We plotted these trajectories in three rows (one row for each of the three percentiles). Each column represents one *Hevelius* session. This visualization enables one to compare the mouse trajectories of the patient and those of healthy controls, see the variation in trajectories across trials at a given time point (best, median, and worst), and see how trajectories change over time.

Furthermore, we visualized the mouse movement speed profile during each trajectory (**"speed plots"**). An example speed plot is shown in Figure 5 (Appendix A). The speed plot shows three phases of the task: initiation (between the start of the trial and the start of mouse movement), execution (between when the patient first moves the mouse and before the patient clicks the mouse), and click duration (the start of the click to the end of the click, i.e., from mouse down to mouse up). This visualization enables one to compare the computer mouse speed of the patient and healthy controls, see the variation in speed across trials at a given time point ($5th$, $50th$, and $95th$ percentile trials), and see how movement speed changes over time.

Through representing contrasts [20], the visualizations can provide insights into a patient's mobility condition and progression compared to other populations. For example, patients with Parkinson's might move the mouse more slowly but follow a straight line. In contrast, healthy control participants might move quickly but deviate from the straight line more drastically. Healthy controls might also start the task much sooner than patients in the initiation phase, which connects with the clinical concept of movement initiation.

Third, the visualization tool is interactive. Upon developing a design prototype (Appendix E, Figure 10), we built an interactive app—*Hevelius Report*—using *Streamlit.io* [3] to address the clinicians' need to conveniently explore *Hevelius* results (Obstacle 3). When using the visualization tool, users are first asked to enter the unique code of a patient. Then, users are presented with basic patient information (diagnosis, age, gender) (Figure 1a), summary plots (with the option to view different dates) (Figure 1a), trajectory plots (Figure 1b), speed plots (Figure 1c), and additional information about the analyses (specific factor analysis results) (Figure 1d). Users can select a subset of concepts and set the number of health control comparisons (in grey) as they scroll through the report.

6 Preliminary user study

Recruitment and Analysis. To elicit early feedback on the usefulness and usability of our visualization platform, we conducted a one-hour semi-structured group interview with three clinicians who treat patients with movement disorders, including Parkinson's disease. None of them were involved in the project or had prior knowledge of the system. We recruited the clinicians through personal connections. Since clinicians are highly specialized experts with many time constraints, we minimized the number of scheduling emails and agreed to meet at a time that worked for all three of them. We did not record the meeting nor collect any personal information. We only reported high-level findings without using any quotes in this paper.

Results. Overall, the clinicians affirmed the speculated obstacles and usage scenarios, expressing interest in using this visualization

tool. The combination of raw movement trajectories and summary plots allowed them to identify subtle changes not available in existing clinical tests. The clinicians confirmed that the high-level concepts extracted using factor analysis were clinically relevant. The clinicians indicated that those concepts are aspects of movement they examine when assessing a patient's motor skills. They also expressed interest in further analysis of the statistical associations between the factors and established clinical measures.

When it comes to the visualizations, clinicians held diverging preferences, highlighting the importance of interactivity. For example, regarding the ordering of data visualizations, two clinicians preferred seeing summary plots before diving into the details; one clinician preferred seeing trajectory plots first to do most of the synthesis himself before seeing the summary plots. This indicates different preferences for cognitive engagement [9]. Designing to meet human experts' diverse needs in clinical tasks is crucial for reducing over-reliance on automation and the resulting high-stake erroneous decisions [12].

7 Limitations and Future Work

One limitation of our current work is the relatively small number of end users involved. Future work will seek perspectives from more clinicians, as well as other key stakeholders like patients (e.g., to understand how they might want to see and use their own data [27]). Future work also includes integrating *Hevelius Report* into an examination program *Neurobooth* [2] that uses comprehensive digital tests, including *Hevelius*, to examine a patient's motor skills, eye movements, speech, and cognition. In the meantime, we will seek additional user perspectives on user scenarios and obstacles. Upon further improvement, we will formally evaluate the tool and its impact on clinical decision-making by conducting systemic evaluations with clinicians and other end users–with usability testing like the system usability scale (SUS) questionnaire and controlled studies.

Another limitation is the small number of longitudinal data to visualize patients' progression of mobility over time. We believe that ongoing data collection efforts will enable better usage of the progression plot. We also emphasize the importance of recruiting a demographically and socio-economically diverse group of patients and healthy control participants to mitigate downstream bias [18]. A diverse and accurate data presentation of patients will improve the population estimates of *Hevelius* outcomes and facilitate more equitable clinical research and treatment.

References

[1] [n. d.]. factor_analyzer 0.5.1 documentation. https://factor-analyzer.readthedocs.io/en/latest/ Accessed: 2024-08-13.

[2] [n. d.]. Neurobooth at Massachusetts General Hospital. https://neurobooth.mgh.harvard.edu Accessed: 2024-08-13.

[3] [n. d.]. Streamlit. https://streamlit.io/. Accessed: 2024-06-19.

[4] Melissa J Armstrong and Michael S Okun. 2020. Diagnosis and treatment of Parkinson disease: a review. *JAMA* 323, 6 (2020), 548–560.

[5] Sivakumar Balasubramanian, Alejandro Melendez-Calderon, and Etienne Burdet. 2012. A robust and sensitive metric for quantifying movement smoothness. *IEEE Transactions on Biomedical Engineering* 59, 8 (2012), 2126–2136. https://doi.org/10.1109/TBME.2011.2179545

[6] Steven Dow, Julie Fortuna, Dan Schwartz, Beth Altringer, Daniel Schwartz, and Scott Klemmer. 2011. Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2011-05-07) *(CHI '11)*. Association for Computing Machinery. https://doi.org/10.1145/1978942.1979359

[7] Nicole M Eklund, Jessey Ouillon, Vineet Pandey, Christopher D Stephen, Jeremy D Schmahmann, Jeremy Edgerton, Krzysztof Z Gajos, and Anoopum S Gupta. 2023. Real-life ankle submovements and computer mouse use reflect patient-reported function in adult ataxias. *Brain Communications* 5, 2 (March 2023), fcad064. https://doi.org/10.1093/braincomms/fcad064

[8] Batya Friedman. 1996. Value-sensitive design. 3, 6 (1996), 16–23. https://doi.org/10.1145/242485.242493

[9] Krzysztof Z Gajos and Lena Mamykina. 2022. Do people engage cognitively with AI? Impact of AI assistance on incidental learning. In *International Conference on Intelligent user Interfaces*. 794–806.

[10] 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.

[11] Sucheta Ghoshal and Sayamindu Dasgupta. 2023. Design Values in Action: Toward a Theory of Value Dilution. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2347–2361. https://doi.org/10.1145/3563657.3596122

[12] Kate Goddard, Abdul Roudsari, and Jeremy C. Wyatt. 2012. Automation bias: a systematic review of frequency, effect mediators, and mitigators. 19, 1 (2012), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

[13] John A Hartigan and Manchek A Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society* 28, 1 (1979), 100–108.

[14] Jaimie Hixson, Jorge E Quintero, Ashley Guiliani, Tritia R Yamasaki, Greg A Gerhardt, John T Slevin, and Craig van Horne. 2023. Visualization of the Movement Disorder Society Unified Parkinson's Disease Rating Scale Results. *Journal of Parkinson's Disease* (2023), 1–6.

[15] Lisa I Iezzoni, Ellen P McCarthy, Roger B Davis, and Hilary Siebens. 2001. Mobility difficulties are not only a problem of old age. *Journal of General Internal Medicine* 16 (2001), 235–243.

[16] Peter Jombík, Pavel Spodniak, Vladimír Bahỳl', and Ján Necpál. 2020. Visualization of Parkinsonian, Essential and Physiological Tremor Planes in 3D Space. *Physiological Research* 69, 2 (2020), 331.

[17] Ilir Jusufi, Dag Nyholm, and Mevludin Memedi. 2014. Visualization of spiral drawing data of patients with Parkinson's disease. *International Conference on Information Visualisation* (2014), 346–350.

[18] Rie Kamikubo, Lining Wang, Crystal Marte, Amnah Mahmood, and Hernisa Kacorri. 2022. Data Representativeness in Accessibility Datasets: A Meta-Analysis. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (Athens, Greece) *(ASSETS '22)*. Association for Computing Machinery, New York, NY, USA, Article 8, 15 pages. https://doi.org/10.1145/3517428.3544826

[19] Simeon Keates, Faustina Hwang, Patrick Langdon, P. John Clarkson, and Peter Robinson. 2002. Cursor measures for motion-impaired computer users. In *Assets '02: Proceedings of the fifth international ACM conference on Assistive technologies* (Edinburgh, Scotland). ACM, New York, NY, USA, 135–142. https://doi.org/10.1145/638249.638274

[20] Ference Marton. 2014. *Necessary conditions of learning*. Routledge.

[21] Fionn Murtagh and Pedro Contreras. 2012. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2, 1 (2012), 86–97.

[22] Vineet Pandey, Nergis C. Khan, Anoopum S. Gupta, and Krzysztof Z. Gajos. 2023. Accuracy and Reliability of At-Home Quantification of Motor Impairments Using a Computer-Based Pointing Task with Children with Ataxia-Telangiectasia. *ACM Transactions on Accessible Computing* 16, 1 (March 2023), 10:1–10:25. https://doi.org/10.1145/3581790

[23] Neltje E Piro, Lennart K Piro, Jan Kassubek, and Ronald A Blechschmidt-Trapp. 2016. Analysis and visualization of 3D motion data for UPDRS rating of patients with Parkinson's disease. *Sensors* 16, 6 (2016), 930.

[24] Katharina Reinecke and Krzysztof Z. Gajos. 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. ACM, New York, NY, USA, 1364–1378. https://doi.org/10.1145/2675133.2675246

[25] 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. 24, 12 (2009), 1820–1828. https://doi.org/10.1002/mds.22681

[26] Jeremy D. Schmahmann, Samantha Pierce, Jason MacMore, and Gilbert J. L'Italien. 2021. Development and Validation of a Patient-Reported Outcome Measure of Ataxia. *Movement Disorders* 36, 10 (Oct. 2021), 2367–2377. https://doi.org/10.1002/mds.28670

[27] Jianna So, Faye X Yang, Anoopum S Gupta, and Krzysztof Z Gajos. 2024. "It's Better to be Grounded in Reality": a Speculative Exploration of Patient-Centered Digital Phenotyping for Neurological Conditions. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, Newfoundland and Labrador, Canada) *(ASSETS '24)*. Assoiation for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3663548.3688486

[28] Jonathan Synnott, Liming Chen, CD Nugent, and George Moore. 2010. Assessment and visualization of Parkinson's disease tremor. In *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*. IEEE, 1–4.

[29] Mohsen Tavakol and Angela Wetzel. 2020. Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education* 11 (2020), 245.

[30] Maryam Tohidi, William Buxton, Ronald Baecker, and Abigail Sellen. 2006. Getting the right design and the design right. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montréal, Québec, Canada) *(CHI '06)*. Association for Computing Machinery, New York, NY, USA, 1243–1252. https://doi.org/10.1145/1124772.1124960

[31] Edward Tufte. 1990. *ENVISIONING INFORMATION*. Graphics Pr.

[32] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, 11 (2008).

[33] Marley W. Watkins. 2018. Exploratory Factor Analysis: A Guide to Best Practice. *Journal of Black Psychology* 44, 3 (April 2018), 219–246. https://doi.org/10.1177/0095798418771807
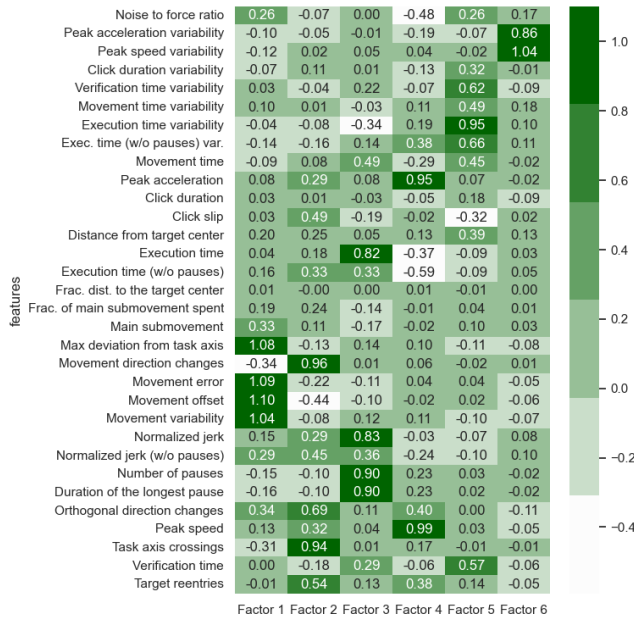
A   Figures of visualization

Figure 2: Factor analysis results. We performed factor analysis on the *Hevelius* test results from 247,667 healthy control participants and reduced the 32 *Hevelius* features to 6 factors. The coefficients representing the relationship between each factor and the features are displayed above. We mapped the 6 factors to 6 concepts: 1) deviation from straight line, 2) directional change from target, 3) pauses and jerks, 4) speed, 5) time inconsistency, and 6) speed inconsistency (Factors 1-6, respectively). Discussions with clinicians indicated that these concepts are clinically relevant.
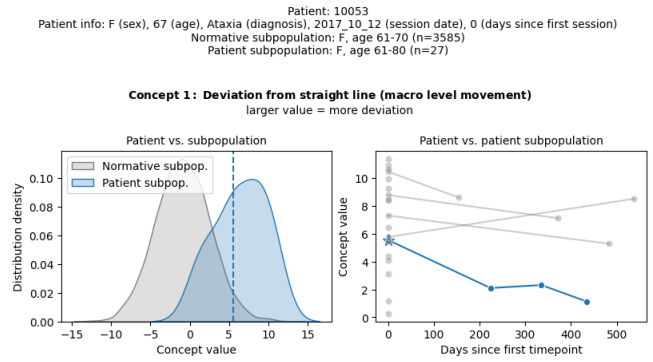

Figure 3: Example summary plot for a Parkinson's patient for a given timepoint and concept. The plot on the left compares the patient's result with that of the relevant patient and healthy control sub-populations. The plot on the right shows the progression of the patient's condition over time and compares the patient's progression with that of other patients.
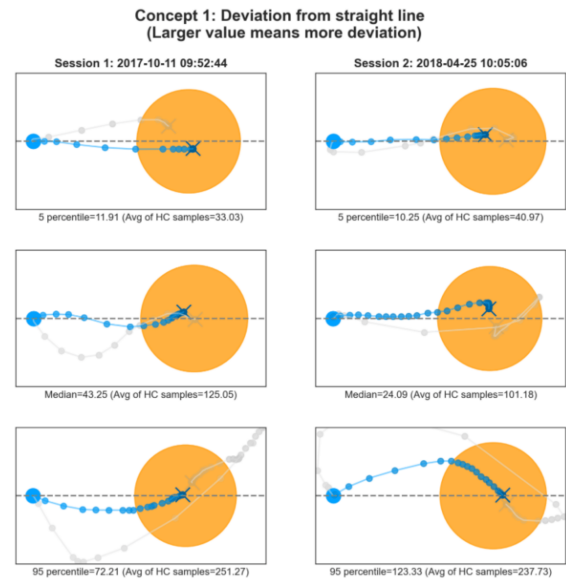

Figure 4: Example trajectory plots for a Parkinson's patient. The plot compares the mouse trajectory of the patient (blue) to example trajectories from healthy controls (gray) over two *Hevelius* sessions (two columns). Users can choose to see one average healthy control trajectory or a distribution of up to 20 trajectories. The 5th, 50th, and 95th percentile trajectories are selected based on a concept (in this example, the "deviation from straight line" concept).
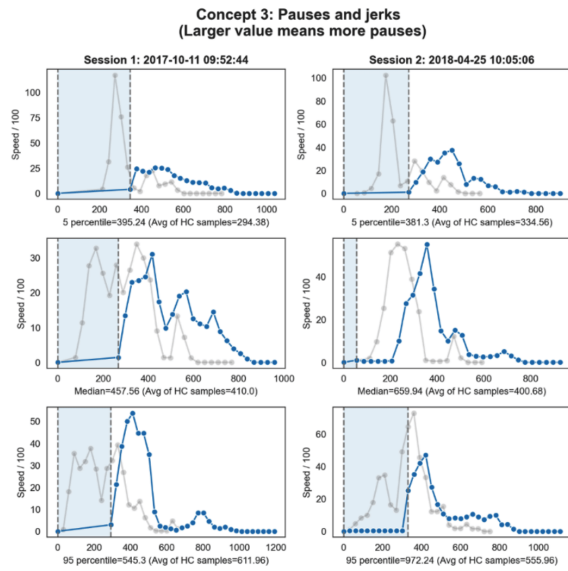
Figure 5: Example time-speed plots for a Parkinson's patient over two *Hevelius* sessions. The plot shows the speed of the mouse during the corresponding trajectories in the trajectory plot (Figure 4). Users can choose to see one average healthy control speed line (gray) or a distribution of up to 20 speed lines. The initiation phase (time between the start of the task and the start of mouse movement) is highlighted in the light blue shaded area.

B Other solutions explored

B.1 Concept Learning

In addition to factor analysis, we also explored other methods to perform dimensionality reduction, including k-means clustering [13], t-SNE [32], and hierarchical clustering [21]. However, these methods had the following shortcomings. Although k-means clustering and t-SNE can reduce a large number of features into a smaller number of clusters, the relationship among features in each cluster and, thus, the meaning of each cluster is difficult to interpret. For hierarchical clustering, although the resulting dendrogram depicts how features are related to each other in terms of distance in space, it is unclear how to summarize a given cluster of features mathematically. For details, please see Figure 6. Because factor analysis not only clearly indicates the relationship between each factor and the features but also provides a way to mathematically summarize the features (i.e., each factor is a linear combination of the features), we chose to use factor analysis for the study. When performing factor analysis, we explored the possibility of reducing the 32 features to 4, 5, 6, and 7 factors. We ultimately selected 6 factors because they provided a good balance between a small number of factors explaining a large proportion of variance in the data.
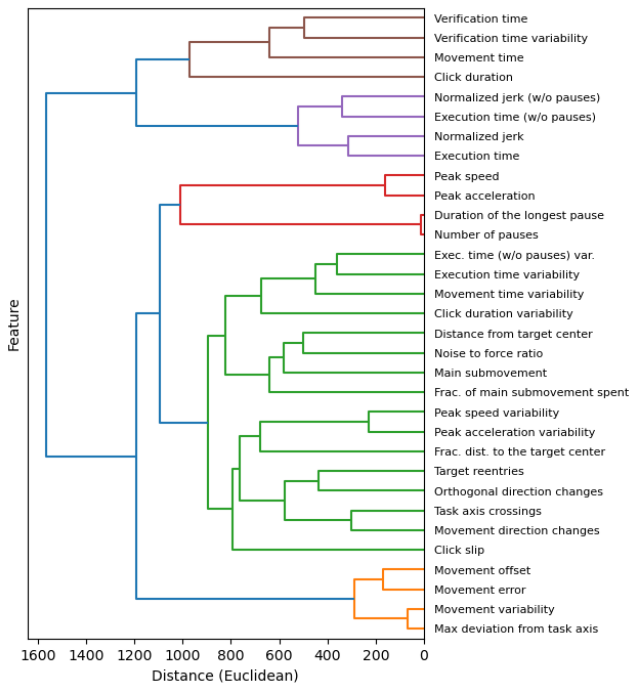

Figure 6: Hierarchical clustering. In addition to factor analysis, we also explored hierarchical clustering as another method to reduce the dimensionality of the *Hevelius* features. The dendrogram from the hierarchical clustering analysis is displayed above. Although hierarchical clustering shows how features are related to each other in terms of distance in space, it is unclear how to summarize a given cluster of features mathematically. Thus, we did not end up using hierarchical clustering to reduce the dimensionality of the features.

We also conducted factor analysis on patient subgroups instead of healthy control populations to understand how the concept groups differ. We found a high level of overlap among these results and decided to use the concept groups based on the healthy control populations due to their generalizability across different patient groups.

B.2 Visualization

During the design of *Hevelius* visualizations, one key question that came up repeatedly was how to show the distributions of the healthy control subpopulation and the patient subpopulation. For the summary plots, we obtained the distribution of values, performed kernel density estimation to obtain a smooth curve, and plotted this curve to represent the distribution. For the trajectory and speed plots, based on discussions with our clinician collaborator, we initially plotted three trajectories from the specified quantiles from healthy control participants to provide some information about the distribution of trajectories. However, discussions with other clinicians revealed that some prefer seeing a single average trajectory from the healthy control population instead. To account for these different preferences, we designed the app to allow users to specify the number of trajectories they want to see. For examples of different prototypes during the development of the visualizations, please see Appendix D (Figures 7, 8, and 9).

Another point of discussion was whether to normalize data points to facilitate the comparison of patients' progression in the summary plots (Figure 3, right). In the plot, normalizing each patient's data with respect to results from each patient's first *Hevelius* session would allow all the lines to start at the same starting point and would allow for a clearer comparison of how each patient has progressed since their first *Hevelius* session. However, the unnormalized data shows large variability in the patients' starting points. This information would be lost upon normalization and may misguide the interpretation of the data, misleading one to believe that patients are more similar than they actually are. Given these concerns with normalized data, we opted to visualize the unnormalized data.

B.3 Visualization Platform

In addition to *Streamlit.io*, we also considered building the user interface using *R Shiny* and via our own website. Building our own website would allow for great flexibility, but would also entail unnecessary additional work of building various website structures from scratch. In contrast, *Streamlit.io* and *R Shiny* are packages that offer support in building interactive web apps, so we opted for the latter two. We chose *Streamlit.io* over *R Shiny* to build the user interface because *Streamlit.io* integrates seamlessly with Python (the coding language used to create the data visualizations).

During the design process, we discussed the interface options, such as having different pages for each data visualization or a continuous scroll. Based on discussions with our collaborators, we opted for a continuous scroll so that users could see multiple visualizations simultaneously (if they choose to) and added a sidebar menu to jump to different sections quickly.

C   List of features in *Hevelius* test results

The *Hevelius* test results include 3 features related to patient and session identification and 32 features related to mouse movement. The 3 features related to patient and session identification are 1) patient ID, 2) diagnosis, and 3) patient-session ID. The 32 features related to mouse movement are:

(1) Standard deviation (computed over all trials in a block) of the distance from the target center at the end of the first submovement divided by the mean of peak accelerations

(2) Peak acceleration: maximum smoothed acceleration recorded during a movement

(3) Peak speed: the maximum smoothed speed recorded during a movement

(4) Click duration variability: standard deviation of click durations in a block of trials

(5) Verification time variability: standard deviation of verification times in a block of trials

(6) Movement time variability: coefficient of variation of movement times in a block of trials

(7) Execution time variability: coefficient of variation of execution times in a block of trials

(8) Execution time variability without pauses: coefficient of variation of execution times without pauses in a block of trials

(9) Movement time: complete movement time from target onset to the end of the successful click on the target

(10) Peak acceleration: maximum smoothed acceleration recorded during a movement

(11) Click duration: time between mouse button press and release during the correct click on the target

(12) Click slip: distance between the point where the mouse button was pressed down and where it was released during click on the target

(13) Distance from target center at the end of main submovement: 2D distance from the mouse pointer location at the end of the main submovement to the target center

(14) Execution time: time from the first to the last mouse movement, excluding any movement that occurred while the mouse button was pressed

(15) Execution time without pauses: like execution time, but excludes pauses of 100ms or longer

(16) Fraction of remaining distance to the target center covered in main submovement: fraction of the remaining distance along the task axis covered during the main submovement

(17) Fraction of the main submovement spent accelerating: fraction of the time from the start of the submovement to the time when acceleration reached its peak value divided by the overall duration of the submovement

(18) Main submovement: submovement with the highest peak speed

(19) Maximum deviation from task axis: maximum distance of the mouse pointer from the task axis during a movement

(20) Movement direction changes: number of times the movement component orthogonal (vertical) to the task axis changes sign

(21) Movement error: average absolute distance of the mouse pointer from the task axis

(22) Movement offset: average (non-absolute) distance of the mouse pointer from the task axis

(23) Movement variability: standard deviation of the distance of the actual path from the task axis

(24) Normalized jerk: dimensionless measure computed as normalized jerk based on Pandey et al. [22]

(25) Normalized jerk without pauses: like normalized jerk, but excludes parts of the movement when the mouse pointer was paused for 100ms or longer

(26) Number of pauses: number of pauses of 100ms or longer

(27) Duration of the longest pause: duration of the longest pause of 100ms or longer

(28) Orthogonal direction changes: number of times the movement component parallel to the task axis changes sign

(29) Peak speed: maximum smoothed speed recorded during a movement

(30) Task axis crossings: number of times the mouse pointer crossed the task axis during the movement

(31) Verification time: 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)

(32) Target re-entries: number of times the mouse pointer leaves the target and then re-enters it before the start of the click

D Development of data visualizations

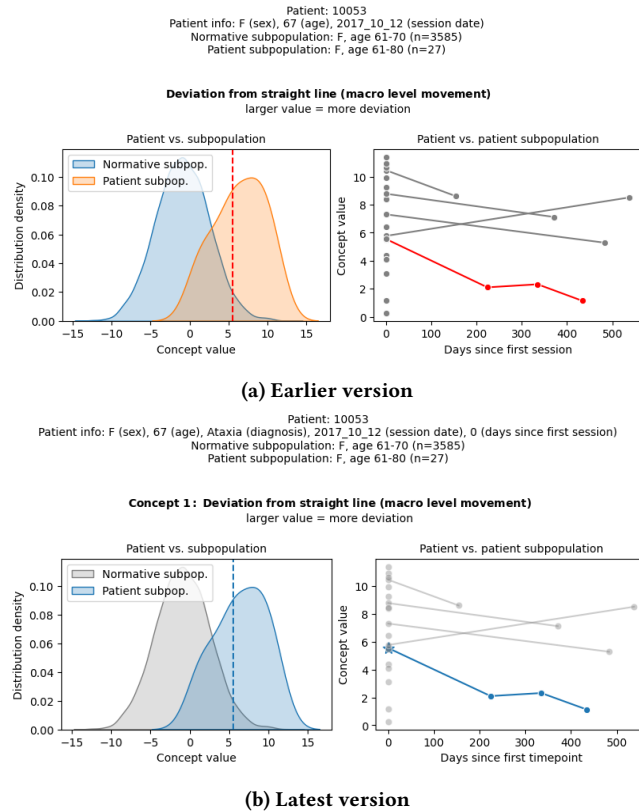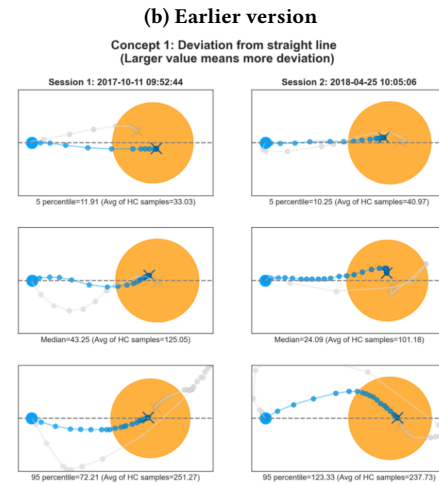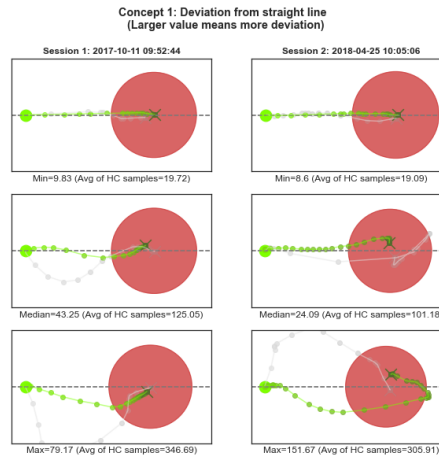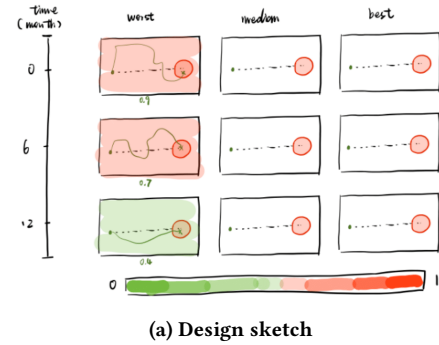
(a) Earlier version

(b) Latest version

Figure 7: Development of summary plots, from (a) an earlier version to (b) the latest version (same as Figure 3 in main paper). In the latest version, we adjusted the color scheme for consistency with trajectory and speed plots, added a star indicating the visualized timepoint in the progression plot, statistically tested the difference between the patient population and control population, and noted the patient's percentile with respect to the patient population and healthy control population.

(a) Design sketch

(b) Earlier version

(c) Latest version

Figure 8: Development of trajectory plots, from (a) a design sketch to (b) an earlier version to (c) the latest version (same as Figure 8 in main paper). After sketching potential designs, we chose the design shown in (a) and implemented it in (b). In contrast to the earlier version (b), which visualizes the best, median, and worst trajectories, shows the average healthy control trajectory in gray, and uses a red-green color scheme, the latest version (c) visualizes the $5th$, $50th$, and $95th$ percentile trajectories, can show a single average or up to 20 healthy control trajectories in gray (user decides), and uses an orange-blue color scheme consistent with the color scheme in the summary and speed plots.
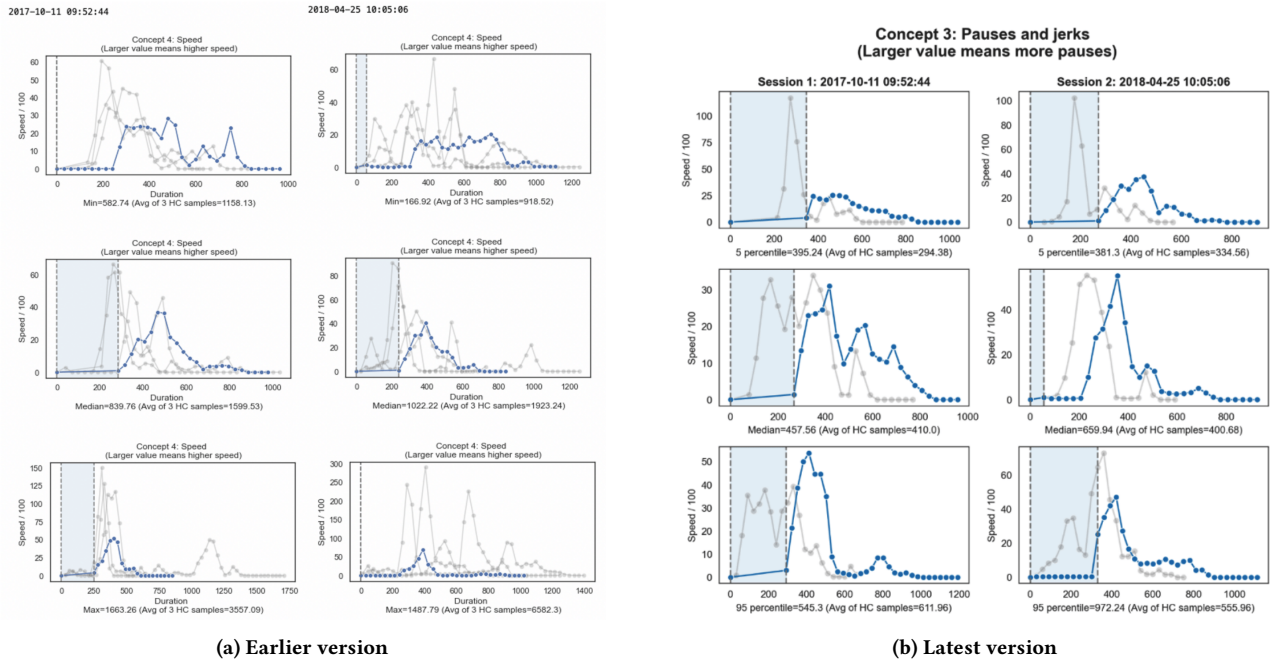
(a) Earlier version

(b) Latest version

Figure 9: Development of speed plots, from (a) an earlier version to (b) the latest version (same as Figure 9. In the earlier version, users could only visualize the average speed line for a healthy control in gray. In the latest version, users can visualize a single average or up to 20 healthy control speed lines in gray.
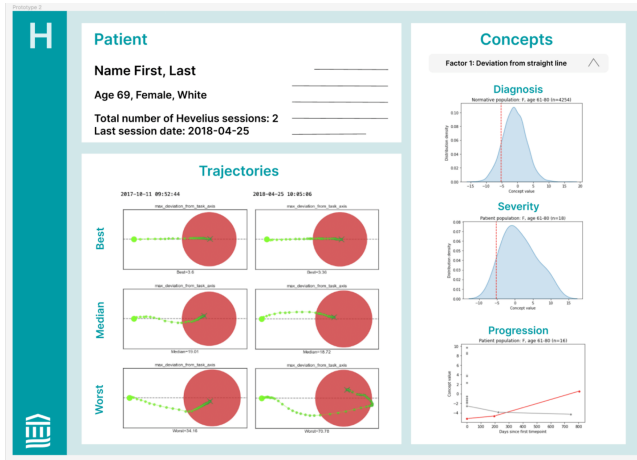
E Development of user interface

Figure 10: Design sketch for the user interface of the visualization tool. Later, during the implementation, because the amount of information and visualizations was too much to display on one screen, we opted for a single-scroll continuous page with a sidebar that allows users to jump to a section of interest.