Copyright © 2024, Massachusetts Medical Society. All rights reserved.

This is an Author Accepted Manuscript, which is the version after external peer review and before publication in the Journal. The publisher's version of record, which includes all NEJM AI editing and enhancements, will be available at https://ai.nejm.org/doi/full/10.1056/AIoa2401137.

Video-based biomechanical analysis captures disease-specific movement signatures of different neuromuscular diseases

Parker S. Ruth[1]*, Scott D. Uhlrich[2]*, Constance de Monts[3], Antoine Falisse[4], Julie Muccini[5], Sydney Covitz[4], Shelby Vogt-Domke[3], John Day[3], Tina Duong[3†], Scott L. Delp[4,6,7†]

*Contributed equally
[†]Contributed equally

[1]Department of Computer Science, Stanford University
[2]Department of Mechanical Engineering, University of Utah
[3]Department of Neurology, Stanford Medicine
[4]Department of Bioengineering, Stanford University
[5]Department of Radiology, Stanford Medicine
[6]Department of Mechanical Engineering, Stanford University
[7]Department of Orthopaedic Surgery, Stanford Medicine

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Corresponding Authors:
P.S.R. email: paru@stanford.edu

Abstract

Background

Assessing human movement is essential for diagnosing and monitoring movement-related conditions like neuromuscular disorders. Timed function tests (TFTs) are among the most widespread assessments due to their speed and simplicity, but they cannot capture disease-specific movement patterns. Conversely, biomechanical analysis can produce sensitive disease-specific biomarkers but is traditionally confined to laboratory settings. Recent advances in smartphone video-based biomechanical analysis enable quantification of 3D movement with the ease and speed required for clinical settings. However, the potential of this technology to offer more sensitive assessments of human function than TFTs remains untested.

Methods

To compare video-based analysis against TFTs, we collected an observational dataset from 129 individuals: 28 with facioscapulohumeral muscular dystrophy, 58 with myotonic dystrophy, and 43 controls with no diagnosed neuromuscular condition. We used OpenCap, a free open-source software tool, to capture smartphone video-based biomechanics of nine different movements in a median time of 16 minutes per participant. From these recordings we extracted 34 interpretable movement features. Using these features, we evaluated the ability of video-based biomechanics to reproduce four TFTs (10-meter walk, 10-meter run, timed up-and-go, and 5-time sit-to-stand) while capturing additional disease-specific signatures of movement.

Results

Video-based biomechanical analysis reproduced all four TFTs ($r > 0.98$) with similar test-retest reliability. In addition, video metrics outperformed TFTs at disease classification ($p = 0.021$). Unlike TFTs, video-based biomechanical analysis identified disease-specific signatures of movement such as differences in gait kinematics that are not evident in TFTs.

Conclusion

Video-based biomechanical analysis can complement existing functional movement assessments by capturing more sensitive, disease-specific outcomes from human movement. This technology enables digital health solutions for assessing and monitoring motor function, complementing traditional clinical outcome measures to enhance care, management, and clinical trial design for movement-related conditions.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Description

This study demonstrates that smartphone video-based biomechanical analysis can accurately replicate traditional timed function tests (TFTs) — commonly used to diagnose and assess movement-related conditions like neuromuscular disorders — while also capturing disease-specific movement patterns that TFTs fail to detect. By enabling sensitive and interpretable assessments in clinical settings, this approach offers a scalable and objective tool to enhance diagnosis, monitoring, and clinical trial design for neuromuscular disorders.

Introduction

Quantitative measurements of human movement can improve diagnosis, monitoring, and understanding of movement-related disorders. The state of the art for measuring human movement involves marker-based motion capture and force plates, which cost hundreds of thousands of dollars and require hours of expert labor.[1] While laboratory-based methods are the gold standard,[2] real-world clinical applications require greater speed and convenience.

Timed function tests (TFTs), which assess gross motor function by measuring the time taken to complete an activity, have wide clinical adoption due to their speed and simplicity. However, TFTs are less informative than laboratory-based biomechanical analysis. Reducing human movement to a single time score overlooks movement quality factors like compensatory gait changes or limited joint range-of-motion, which may better reflect disease progression. Thus, there is a gap between the rich measurement tools used in biomechanics research and those deployed in clinical settings.

With the ongoing development of disease-modifying therapies for neuromuscular diseases, improved measures of mobility and function are crucial for evaluating treatment efficacy, clinical decision-making, and disease management.[3,4] The sensitivity and accuracy of clinical endpoints greatly impact trial design and therapy advancement. For slower-progressing diseases, TFTs lack sensitivity to subtle changes in muscle function. Coarse metrics can fail to detect natural progression and the effects of therapeutic interventions. For example, individuals with early-stage disease may retain normal walking speeds (and hence normal TFT times) but display compensations or decreased movement quality indicating impaired muscle function. There is a strong clinical need for functional assessments sensitive to the unique phenotypical signatures of muscle dysfunction across neuromuscular diseases.[5] Prior work has demonstrated the promise of digital movement biomarkers to better quantify disease progression in several neuromuscular diseases, including Duchenne muscular dystrophy,[6] Friedreich's ataxia,[7] and Parkinson's disease.[8] Using full-body inertial measurement unit sensors and laboratory-based motion capture, these studies show how movement biomarkers can better predict disease trajectory and are more sensitive to short-term progression than standard clinical tests.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Here, we focus on two slow-progressing neuromuscular diseases: facioscapulohumeral muscular dystrophy (FSHD) and myotonic dystrophy (DM). There are no approved digital outcome measures for these conditions, and their slow progression necessitates sensitive outcome measures to accelerate the development of treatments. Both conditions impact gross motor function (e.g., reduce walking speed), but they affect different muscle groups, leading to distinct movement signatures. FSHD causes weakness in facial, scapulothoracic, shoulder, trunk, and ankle dorsiflexor muscles, resulting in impaired locomotion and overhead reaching.[9,10] DM primarily causes weakness in the lower trunk and distal muscles of the ankle, hand, trunk, knee, and hip, resulting in difficulty rising from a chair, squatting, and activities of daily living.[11]

Quantifying human movement with video can provide richer metrics with minimal added burden.[12] For example, machine learning models can predict clinical scores of Parkinson's disease severity from videos of gait[13] and sit-to-stand transitions.[14] The speed and ease of video capture facilitates scalable data collection in clinical and remote settings. For example, at-home videos of sit-to-stand transitions predict osteoarthritis and health outcomes.[15] Video capture is rich and flexible enough to tailor metrics for individual diseases. For example, "reachable workspace" is a video-derived metric quantifying upper-body functional volume developed for FSHD.[16] Reachable workspace is validated against other upper-limb metrics,[17] reflects longitudinal progression,[18] correlates with patient reported-outcomes, and is used as an outcome measure in clinical trials.[19]

We previously developed OpenCap, an open-source tool that computes three-dimensional (3D) human kinematics (i.e., joint positions and angles) from smartphone videos.[20] By eliminating time-consuming analysis and specialized equipment, this greatly expands the scale and accessibility of human movement analysis. For example, OpenCap can detect subtle movement patterns — such as steppage gait or circumduction gait — that go beyond simple metrics like walking speed.[21]

This study aimed to compare video-based 3D biomechanical analysis using OpenCap (Figure 1) to TFTs for quantifying movement of individuals with FSHD and DM. Specifically, we investigated two research questions: (1) Can OpenCap video-based metrics accurately reproduce clinician-graded TFTs in individuals with and without neuromuscular diseases? (2) Can OpenCap video-based biomechanical assessment capture disease-specific movement patterns in FSHD and DM that are more informative than TFT times alone? To answer these questions, we collected a dataset of OpenCap recordings and TFTs, extracted video-based movement features, and explored the potential for video-based digital metrics of human movement to both reproduce TFTs and provide additional relevant movement information about two neuromuscular diseases.

Methods

Participant screening and recruitment

One hundred twenty-nine individuals completed the study (Table 1), including 28 with FSHD, 58 with DM (49 myotonic dystrophy type 1 [DM1], 9 myotonic dystrophy type 2 [DM2]), and 43 controls with no diagnosed neuromuscular condition. (For detailed participant demographics see Table S1.) Prior to participation, participants provided written informed consent to a protocol approved by the Stanford University Institutional Review Board (IRB IRB00005136). The OpenCap system is HIPAA compliant and has been approved by the Stanford University Offices of Privacy and Data Security to process and store identifiable personal health information. Participants provided consent to the collection, processing, and optional sharing of their identifiable video data.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Inclusion criteria were the ability to rise from a chair and ambulate independently. Participants in the control group had no neuromuscular condition, movement disorder, or injury affecting movement in the last six months. Participants with movement disorders had confirmed diagnosis of DM (DM1 or DM2) or FSHD. Participants were recruited in three settings: 94 individuals participated in a high-throughput data collection, primarily in a university gymnasium using multiple parallel OpenCap motion capture stations; 30 participants were recruited on-site at support group meetings and national conferences; 5 participants were recruited during clinic visits.

Selection of physical activities

In consultation with the clinical domain experts on the study team, we selected a set of nine upper- and lower-body movements (Table S2) representing isolated muscle groups and whole-body function. We chose run, walk, timed up-and-go, and sit-to-stand movements because they are common TFTs. We added a 30-second calf raise to test balance and plantarflexor strength and a jump task to test maximal force of lower-extremity extensors. Finally, we chose three upper-body movements: shoulder abduction, elbow flexion, and arm range-of-motion. The arm abduction task is used in the existing Brooke Upper Extremity Scale, which ordinally scores functional ability of overhead arm abduction, hand-to-mouth movement, and distal hand function.[22,23] Together, these activities represent a wide spectrum of functional movement amenable to video-based biomechanical analysis.

Domain expert-informed feature engineering

We leveraged the expertise of clinical research team members and previously reported movement patterns in FSHD and DM to select a set of 34 interpretable, activity-specific features from the kinematic data (Table S3). To design features, we discussed common movement patterns exhibited by individuals with DM and FSHD and examined representative videos and skeletal kinematics from the dataset. The clinicians verbally identified common movement impairments and compensations, and engineers designing algorithms to quantify clinical observations. Some participants were unable to complete ankle-based activities (calf raise and jump); these missing values were imputed with zeros.

Scalable video-based human movement assessments

We used OpenCap, a free and open-source software platform enabling smartphone video-based biomechanical analysis.[20] OpenCap computes three-dimensional anatomical marker positions and skeletal joint angles from inverse kinematics. We selected the OpenCap settings that use HRNet pose estimation[24] and a deep learning model to augment sparse 3D keypoints into a richer anatomical marker set.[25] Within OpenCap, joint angles are computed by the OpenSim[2] Inverse Kinematics tool and a biomechanically constrained skeletal model.[26] All recordings were captured using either two or three cameras. A typical three-camera setup was developed (Figure S4); however, some trials were recorded with only two cameras due to space and logistical constraints (e.g., recording in a clinic hallway). In a prior validation experiment, OpenCap's accuracy did not differ between two- and three-camera setups.[20] All videos were recorded with at least 60 frames per second. The capability to record at 120 frames per second was developed mid-way into data collection, so some gait videos were collected at 120 frames per second to reduce the effects of motion blur during the run activity. Simultaneous with OpenCap recordings, TFTs were measured using stopwatches by evaluators trained by two physical therapists and neuromuscular disease outcomes experts in the Stanford Neuromuscular Clinic. TFTs are known to have very high inter-rater reliability with intraclass correlation above 0.99.[27,28]

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Analysis of video-based reproduction of TFTs

We computed Pearson correlations between our video-derived time scores analogous to TFTs and their corresponding human-measured TFTs. In the case of gait trials, we estimate the TFT time by dividing 10 meters by the gait speed measured in the video field of view. In the case of timed up-and-go and sit-to-stand trials, our algorithm automatically computes the time taken to complete the task. These video-based TFT surrogates do not exactly match what the human is measuring with TFT times. For walking and running, this is because the full ten meters are not visible in the camera field of view. For timed up-and-go and sit-to-stand, discrepancies arise because clinicians time when participants contact the chair seat, while our algorithm is based on center-of-mass elevation. Furthermore, all TFTs are subject to human error in measuring timing of heel strikes or chair contact. Thus, we sought to test whether video analysis can predict human-measured TFT times. To do this, we trained a linear regressor to predict human-measured TFTs from corresponding video-based TFT surrogates. We assessed prediction performance on unseen examples using leave-one-subject-out cross-validation, measuring accuracy with Pearson correlation and mean absolute error (MAE).

To measure test-retest reliability, 13 participants with DM repeated the data collection on the following day. Repeat trials were only used for test-retest analysis. Reliability was measured using intraclass correlation (ICC) with a two-way mixed effects model for absolute agreement of single measurements.[29]

Analysis of disease classification

To study whether video-based biomechanical metrics can detect disease-specific movement patterns that TFTs do not, we compared the performance of binary classification models trained on all participant data using leave-one-subject-out cross-validation on either video-based metrics or TFT times. For

participants in the test-retest experiment, only the first session was included. We used a linear support vector classifier, which is data-efficient and robust to input perturbations.[30] Input features were standardized (zero mean, unit standard deviation) prior to model fitting. The classifier's regularization parameter was chosen in each training fold using inner 3-fold cross validation. Model performance was quantified by balanced accuracy of predictions, which represents the percent of correct predictions, accounting for imbalance in class sizes. We compared different models' performance using the McNemar test of proportions which considers, in the cases where the predictions of two models disagree, whether one is more often correct.[31]

Analysis of disease-specific movement signatures

To explore which video-derived features were different between FSHD and DM, we performed two-sided Kolmogorov-Smirnov tests on all 34 engineered features. This was an exploratory analysis, and we applied Bonferroni corrections to control for type I errors due to multiple hypothesis testing.

Analysis of feature importance

To analyze the relative importance of the video metrics for each disease, we used all video-based features to train support vector classifiers distinguishing FSHD from controls and distinguishing DM from controls. We computed Shapley additive explanations (SHAP) scores,[32] which represent the amount each feature changes the classifier's probability of predicting an individual as having a neuromuscular disease. We measured feature importance as the mean absolute SHAP score across all participants.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Results

Video-based reproduction of TFTs

Video-based TFT surrogates are strongly correlated with human-measured TFT times with Pearson $r > 0.98$ (Figure S7). Linear models using video-based TFT surrogates accurately predict human-measured TFT times with Pearson $r > 0.98$ (Figure 2a-d). To contextualize these results, we compare the MAE of video-predicted TFTs against reported distribution-based minimum clinically important difference (MCID). MCIDs for DM are available for 10-meter walk and timed up-and-go. In both cases, our MAE is below the MCID (10-meter walk: MCID = 0.69 s, MAE = 0.31 s; timed up-and-go: MCID = 0.92 s, MAE = 0.22 s).[33] For the 10-meter run, our MAE (0.24 s) falls under the MCID for Duchenne muscular dystrophy (2.3 s).[34] The MCID for 5-time sit-to-stand in neuromuscular disease is not available in existing literature. Video-based TFT predictions showed similar test-retest reliability as human-measured TFT times and Brooke scores. Among other video features, 70% had moderate reliability or better (Figure S1).

Disease classification

A classifier trained on video metrics derived from all nine activities performed better (balanced accuracy = 82%) than one trained on TFT times (balanced accuracy = 50%, $p = 0.001$) and one trained on TFT times and Brooke scores (balanced accuracy = 72%, $p = 0.021$) (Figure 2e). Video models using features from individual activities performed as well or better than TFT models across all activities. TFT times from individual tasks classify diseases no better than random (balanced accuracy = 50%); however, video-based models classify diseases better than random using running features (balanced accuracy = 68%) or walking features (balanced accuracy = 66%). This is because video-based biomechanical analysis captures some disease-specific gait parameters that are not reflected in TFT times. For example,

video-based biomechanical analysis detected differences in running stride length between FSHD and DM participants ($p = 0.044$), while running speed (analogous to the TFT time) was not significantly different (Figure S2).

Disease-specific feature importance

The most important features for detecting FSHD are upper-limb activities representing shoulder and elbow range of motion (Figure 4a). By contrast, the most important features for detecting DM are from lower-limb activities such as running, walking, and jumping (Figure 4b). These observations are consistent with the known early clinical presentations of upper-limb and shoulder girdle weakness in FSHD and distal lower-extremity weakness in DM.

Disease-specific movement signatures

Compared to the DM group, the FSHD group had shorter stride lengths ($p = 0.037$) and higher ankle heights ($p = 0.008$) during the swing phase of walking (Figure 3a). This is consistent with the presentation of foot drop in FSHD due to ankle dorsiflexor weakness.[9] There was no significant difference in walking speed between FSHD and DM, consistent with the inability of the walking TFT model to classify diseases (Figure 2e).

Compared with DM, FSHD more strongly affects the proximal arm muscles, reducing shoulder and scapulothoracic function. We found significant differences in video-based biomechanical metrics of upper-limb tasks between diseases (Figure 3b). Ninety participants received a Brooke score of 6, indicating an ability to abduct hands overhead without elbow flexion; 23 received a score of 5 (able to raise hands overhead with some elbow flexion), and 21 received a score of 4 (unable to raise hands overhead). Our video metric comprising a composite of shoulder and elbow angles (Figure 3b) correlates with human-graded Brooke scores (Spearman's rank correlation $r = 0.80$). Despite this correlation, there is high variance in the video metric among individuals with the same Brooke score, highlighting the insensitivity of ordinal scores like Brooke. Additionally, we observed significant differences between diseases in other video-based features from upper-body tasks (e.g., reachable area and isolated elbow flexion angle), showing how these metrics can capture disease-specific patterns in upper-body movement (Figure S2). Together, these results highlight the strength of a continuous video-based metric to characterize upper-limb function compared to a human-graded ordinal score.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Discussion

Video-based 3D biomechanical analysis using OpenCap reproduces clinician-graded assessments of mobility and function while also encoding additional disease-specific insights. OpenCap video-based assessments accurately predicted TFT scores from videos (Figure 2a-d) with similar test-retest reliability (Figure S1). Unlike TFTs, video-based analysis differentiated the movement patterns of FSHD versus those of DM (Figure 2e) and identified disease-specific biomechanical impairments (Figure 3). These findings demonstrate the potential of using OpenCap-based biomechanical analysis to create functional outcome measures for neuromuscular diseases that are more sensitive than commonly used TFTs.

Video-based analysis is sensitive to disease-specific motor impairments caused by dysfunction in distinct muscle groups. For example, video-based analysis of running gait distinguishes between diseases while TFT times alone do not (Figure 2e). Video-based analysis also reveals specific gait parameter differences — such as step length and ankle height — between FSHD and DM (Figure 3a), indicating that video metrics capture salient aspects of disease signatures that TFT times do not. Furthermore, video-based metrics distinguish differences in upper-limb range of motion between FSHD

and DM and capture functional variability among individuals assigned the same human-graded Brooke score (Figure 3b). These findings suggest that video-based analysis enables more sensitive upper-limb functional assessments.

One reason video features outperform TFT times for classifying diseases is that they include measures of upper-limb mobility such as shoulder and elbow range of motion. Analysis of feature importance reveals that, unlike DM, the FSHD group could be distinguished from the control group mainly in upper limb features (Figure 4). This is consistent with the known clinical presentation: the scapular stabilizers, which help raise the arms above the head, are among the most affected muscles in FSHD[10] whereas they are less affected in DM.[11] However, even when Brooke scores are combined with TFTs, video-based analysis of all tasks still achieves significantly better classification (Figure 2e). Thus, the advantage of video-based analysis goes beyond simply including upper-limb activities.

Our set of video-based features are not comprehensive. Different features may demonstrate different disease classification performance. For example, our features extracted from the timed up-and-go features exclude gait parameters to avoid redundancy with the walk and run tasks. An alternative to our manual feature engineering process based on pathophysiological, clinical, and biomechanical domain-knowledge is to derive data-driven features using supervised or unsupervised dimensionality reduction.[35] It is possible that these methods could improve classification performance, but the features would be less interpretable than hand-engineered features.

The ultimate goal is to create digital outcome measures that are more sensitive and reliable than traditional TFTs. Video-based biomechanical analysis shows promise for achieving greater sensitivity than existing clinical outcome metrics.[6,7] Some video-based biomechanical metrics have low reliability scores among test-retest participants. With the present dataset, we cannot determine what proportion of variance is attributable to day-to-day variability in task performance versus measurement error. Unsurprisingly, for both the TFT and video measures, the maximal effort tasks (e.g., run) were more repeatable than sub-maximal tasks (e.g., walk). Greater variance in an outcome measure decreases sensitivity. Future studies could choose to focus on the metrics that we found to have high reliability.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

The advantages of novel digital health metrics must be balanced with the burden of data collection and analysis. In this study, the median time to record nine activities was 16 minutes, with kinematics calculations automated on OpenCap's cloud servers. For comparison, collecting the same data using typical laboratory-based motion capture would take three or more hours including calibration, equipment setup, marker placement, activity recording, and data post-processing. OpenCap achieves near lab quality kinematics with a much lower time burden.[20] This is especially significant for studying rare diseases. For example, our cohort of 58 individuals with DM is nearly four times larger than the largest existing study of 3D kinematics in DM.[36] By collecting data in five locations, including national conferences, we enrolled participants from 14 different states within the United States. Compared with traditional motion capture, OpenCap is inexpensive to scale. Our high-throughput data collection included six simultaneous recording stations. Taken together, the speed, portability, and scalability of OpenCap enabled a cohort seven times larger (n=129) than the average laboratory-based biomechanics study (n=18).[37,38]

It is important to acknowledge the limitations of this study. First, further research is required to determine if the advantages of video-based analysis generalize beyond FSHD and DM to other neuromuscular diseases. However, our findings imply that in these conditions, OpenCap video-based biomechanical analysis provided more information than a commonly used functional outcome metric and has the potential to improve the creation of movement biomarkers and functional outcome metrics

for other conditions. It is also important to consider how feature engineering can impact the generalizability of classification results. During feature engineering the research team was not blinded to all participant diagnoses since the research team also performed data collection. However, we believe this poses minimal risk of overfitting because clinicians viewed less than 30 of the 1,290 OpenCap recordings in our dataset during feature engineering. Second, our convenience sample includes a low proportion of female participants with FSHD because FSHD presents earlier and more severely in men (Table S1). The low number of females with FSHD impacts disease classification on this group (Table S4). However, a video-based classifier trained and tested on only male participants does not have worse disease classification (balanced accuracy = 83%), suggesting that our classifier does not depend on sex-linked movement patterns to differentiate diseases. Third, while video-based analysis was faster, cheaper, and easier than traditional methods, our multi-camera setup complicates at-home and continuous measurements. Wearable sensors could better capture long-term patterns in movement and behavior,[6,39–42] but video is better suited for periodic, high-resolution assessments of whole-body biomechanics in controlled settings. Single-camera biomechanical analysis will likely become possible in the near future,[43–45] which may enable more frequent at-home measurements or even opportunistic sensing of uninstructed activities of daily living. Finally, the present analysis is a cross-sectional assessment of construct and discriminant validity. Further work should determine repeatability, correlation with functional and patient-reported outcomes, minimum clinically important differences, and sensitivity to natural progression and therapeutic interventions.

This work shows that OpenCap video-based biomechanical analysis can both reproduce TFTs and measure patterns of movement in FSHD and DM that TFTs cannot. The future clinical implications of video-based functional assessments lie in their ability to enable more sensitive evaluation and monitoring of motor function, with applications in improving diagnosis, accelerating treatment development, and enabling personalized patient management.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

Declaration of interests

A.F. and S.D.U. are cofounders of Model Health, Inc., which supports the non-academic, commercial use of the open-source software OpenCap used for data collection in this study.

Data sharing

The datasets generated and analyzed for this study are freely available in the "Kinematics and timed function tests of facioscapulohumeral muscular dystrophy and myotonic dystrophy" data repository, available at https://doi.org/10.5281/zenodo.13788592. The code for feature extraction, figure generation, and statistics is available in the following GitHub repository: https://github.com/stanfordnmbl/opencap-fshd-dm-analysis.

Acknowledgements

We thank the 42 volunteers who helped facilitate recruitment, data collection, and data curation, especially Lin Karman, Sarah Ismail, Paxton Ataide, and Carmichael Ong. We are grateful to Trevor Hastie for consultation on statistical methods. Thank you to the Myotonic Dystrophy Foundation, and the FSHD Society for their support of our study. We especially wish to thank the 129 study participants without whom this research would not have been possible. This study was funded by the Wu Tsai

Human Performance Alliance, Mobilize Center under NIH Grant P41EB027060, and the Myotonic Dystrophy Foundation Early Career Research Grant. The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript.

Author contributions

P.S.R., S.D.U., C. M., J.D., T.D., and S.L.D. conceptualized and designed the study.
S.D.U., C.M., A.F., J.M., and T.D. developed the methods used to perform the experiments.
P.S.R., S.D.U., C.M., A.F., J.M., S.C., S.V., and T.D. conducted the experiments.
P.S.R. synthesized results and figures.
P.S.R. and S.D.U. drafted the manuscript.
C.M., A.F., J.M., S.C., S.V. J.D., T.D., and S.L.D. revised the manuscript.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

References

1 Uhlrich SD, Uchida TK, Lee MR, Delp SL. Ten steps to becoming a musculoskeletal simulation expert: A half-century of progress and outlook for the future. J Biomech 2023;154:111623. https://doi.org/10.1016/j.jbiomech.2023.111623.

2 Seth A, Hicks JL, Uchida TK, et al. OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. PLOS Comput Biol 2018;14:e1006223. https://doi.org/10.1371/journal.pcbi.1006223.

3 Pascual-Gilabert M, López-Castel A, Artero R. Myotonic dystrophy type 1 drug development: A pipeline toward the market. Drug Discov Today 2021;26:1765–72. https://doi.org/10.1016/j.drudis.2021.03.024.

4 Tihaya MS, Mul K, Balog J, et al. Facioscapulohumeral muscular dystrophy: the road to targeted therapies. Nat Rev Neurol 2023;19:91–108. https://doi.org/10.1038/s41582-022-00762-2.

5 Ghasemi M, Emerson CP, Hayward LJ. Outcome Measures in Facioscapulohumeral Muscular Dystrophy Clinical Trials. Cells 2022;11:687. https://doi.org/10.3390/cells11040687.

6 Ricotti V, Kadirvelu B, Selby V, et al. Wearable full-body motion tracking of activities of daily living predicts disease trajectory in Duchenne muscular dystrophy. Nat Med 2023;29:95–103. https://doi.org/10.1038/s41591-022-02045-1.

7 Kadirvelu B, Gavriel C, Nageshwaran S, et al. A wearable motion capture suit and machine learning predict disease progression in Friedreich's ataxia. Nat Med 2023;29:86–94. https://doi.org/10.1038/s41591-022-02159-6.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

8 Troisi Lopez E, Sorrentino P, Liparoti M, et al. The kinectome: A comprehensive kinematic map of human motion in health and disease. Ann N Y Acad Sci 2022;1516:247–61. https://doi.org/10.1111/nyas.14860.

9 Gambelli CN, Bredin J, Doix A-CM, et al. The effect of tibialis anterior weakness on foot drop and toe clearance in patients with facioscapulohumeral dystrophy. Clin Biomech 2023;102:105899. https://doi.org/10.1016/j.clinbiomech.2023.105899.

10 Dijkstra JN, Goselink RJM, Van Alfen N, et al. Natural History of Facioscapulohumeral Dystrophy in Children: A 2-Year Follow-up. Neurology 2021;97:. https://doi.org/10.1212/WNL.0000000000012882.

11 Bouchard J-P, Cossette L, Bassez G, Puymirat J. Natural history of skeletal muscle involvement in myotonic dystrophy type 1: a retrospective study in 204 cases. J Neurol 2015;262:285–93. https://doi.org/10.1007/s00415-014-7570-x.

12 Stenum J, Cherry-Allen KM, Pyles CO, Reetzke RD, Vignos MF, Roemmich RT. Applications of Pose Estimation in Human Health and Performance across the Lifespan. Sensors 2021;21:7315. https://doi.org/10.3390/s21217315.

13 Lu M, Zhao Q, Poston KL, et al. Quantifying Parkinson's disease motor severity under uncertainty using MDS-UPDRS videos. Med Image Anal 2021;73:102179. https://doi.org/10.1016/j.media.2021.102179.

14 Mehta D, Asif U, Hao T, et al. Towards Automated and Marker-Less Parkinson Disease Assessment: Predicting UPDRS Scores Using Sit-Stand Videos. Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Nashville, TN, USA.

15 Boswell MA, Kidziński Ł, Hicks JL, Uhlrich SD, Falisse A, Delp SL. Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study. Npj Digit Med 2023;6:1–7. https://doi.org/10.1038/s41746-023-00775-1.

16 Kurillo G, Chen A, Bajcsy R, Han JJ. Evaluation of upper extremity reachable workspace using Kinect camera. Technol Health Care 2013;21:641–56. https://doi.org/10.3233/THC-130764.

17 Han JJ, de Bie E, Nicorici A, et al. Reachable workspace and performance of upper limb (PUL) in duchenne muscular dystrophy. Muscle Nerve 2016;53:545–54. https://doi.org/10.1002/mus.24894.

18 Hatch MN, Kim K, Kurillo G, Nicorici A, McDonald CM, Han JJ. Longitudinal study of upper extremity reachable workspace in fascioscapulohumeral muscular dystrophy. Neuromuscul Disord 2019;29:503–13. https://doi.org/10.1016/j.nmd.2019.05.006.

19 Hatch MN, Kurillo G, Chan V, Han JJ. Motion sensor-acquired reachable workspace correlates with patient-reported upper extremity activities of daily living (ADL) function in facioscapulohumeral dystrophy. Muscle Nerve 2021;63:250–7. https://doi.org/10.1002/mus.27123.

20 Uhlrich SD, Falisse A, Kidziński Ł, et al. OpenCap: Human movement dynamics from smartphone videos. PLOS Comput Biol 2023;19:e1011462. https://doi.org/10.1371/journal.pcbi.1011462.

21 Horsak B, Eichmann A, Lauer K, et al. Concurrent validity of smartphone-based markerless motion capturing to quantify lower-limb joint kinematics in healthy and pathological gait. J Biomech 2023;159:111801. https://doi.org/10.1016/j.jbiomech.2023.111801.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

22 Brooke MH, Griggs RC, Mendell JR, Fenichel GM, Shumate JB, Pellegrino RJ. Clinical trial in duchenne dystrophy. I. The design of the protocol. Muscle Nerve 1981;4:186–97. https://doi.org/10.1002/mus.880040304.

23 Lue Y-J, Lin R-F, Chen S-S, Lu Y-M. Measurement of the Functional Status of Patients with Different Types of Muscular Dystrophy. Kaohsiung J Med Sci 2009;25:325–33. https://doi.org/10.1016/S1607-551X(09)70523-6.

24 Sun K, Xiao B, Liu D, Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation. Presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

25 Falisse A, Uhlrich SD, Chaudhari AS, Hicks JL, Delp SL. Marker Data Enhancement For Markerless Motion Capture 2024. https://doi.org/10.1101/2024.07.13.603382.

26 Rajagopal A, Dembia CL, DeMers MS, Delp DD, Hicks JL, Delp SL. Full-Body Musculoskeletal Model for Muscle-Driven Simulation of Human Gait. IEEE Trans Biomed Eng 2016;63:2068–79. https://doi.org/10.1109/TBME.2016.2586891.

27 Hutin E, Ghédira M, Loche C-M, et al. Intra- and inter-rater reliability of the 10-meter ambulation test in hemiparesis is better barefoot at maximal speed. Top Stroke Rehabil 2018;25:345–50. https://doi.org/10.1080/10749357.2018.1460932.

28 Ozsoy I, Kodak MI, Kararti C, Ozsoy G, Erturk A, Kahraman T. Intra- and Inter-Rater Reproducibility of the Face-to-Face and Tele-Assessment of Timed-up and Go and 5-Times Sit-to-Stand Tests in Patients with Chronic Obstructive Pulmonary Disease. COPD J Chronic Obstr Pulm Dis 2022.

29 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.

30 James G, Witten D, Hastie T, Tibshirani R, Taylor J. Support Vector Machines. In: James G, Witten D, Hastie T, Tibshirani R, Taylor J, editors. Introd. Stat. Learn. Appl. Python. Cham: Springer International Publishing; 2023. p. 367–98.

31 Kim S, Lee W. Does McNemar's test compare the sensitivities and specificities of two diagnostic tests? Stat Methods Med Res 2017;26:142–54. https://doi.org/10.1177/0962280214541852.

32 Lundberg SM, Lee S-I. A Unified Approach to Interpreting Model Predictions.

33 Knak KL, Sheikh AM, Andersen H, Witting N, Vissing J. Intrarater reliability and validity of outcome measures in myotonic dystrophy type 1. Neurology 2020;94:e2508–20. https://doi.org/10.1212/WNL.0000000000009625.

34 McDonald CM, Henricson EK, Abresch RT, et al. The 6-minute walk test and other clinical endpoints in duchenne muscular dystrophy: Reliability, concurrent validity, and minimal clinically important differences from a multicenter study. Muscle Nerve 2013;48:357–68. https://doi.org/10.1002/mus.23905.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

35 Winner TS, Rosenberg MC, Jain K, Kesar TM, Ting LH, Berman GJ. Discovering individual-specific gait signatures from data-driven models of neuromechanical dynamics. PLOS Comput Biol 2023;19:e1011556. https://doi.org/10.1371/journal.pcbi.1011556.

36 Tiffreau V, Detrembleur C, Van Den Bergh P, et al. Gait among Patients with Myotonic Dystrophy Type 1: A Three-Dimensional Motion Analysis Study. J Int Soc Phys Rehabil Med 2018;1:65. https://doi.org/10.4103/ijprm.ijprm_8_18.

37 Knudson D. Twenty-Year Trends of Authorship and Sampling in Applied Biomechanics Research. Percept Mot Skills 2012;114:16–20. https://doi.org/10.2466/11.PMS.114.1.16-20.

38 Knudson D. Confidence crisis of results in biomechanics research. Sports Biomech 2017;16:425–33. https://doi.org/10.1080/14763141.2016.1246603.

39 Knak KL, Sheikh AM, Witting N, Vissing J. Physical activity in myotonic dystrophy type 1. J Neurol 2020;267:1679–86. https://doi.org/10.1007/s00415-020-09758-8.

40 Zhuparris A, Maleki G, Koopmans I, et al. Smartphone and Wearable Sensors for the Estimation of Facioscapulohumeral Muscular Dystrophy Disease Severity: Cross-sectional Study. JMIR Form Res 2023;7:e41178. https://doi.org/10.2196/41178.

41 Servais L, Eggenspieler D, Poleur M, et al. First regulatory qualification of a digital primary endpoint to measure treatment efficacy in DMD. Nat Med 2023;29:2391–2. https://doi.org/10.1038/s41591-023-02459-5.

42 Mishra RK, Nunes AS, Enriquez A, et al. At-home wearable-based monitoring predicts clinical measures and biological biomarkers of disease severity in Friedreich's Ataxia. Commun Med 2024;4:1–9. https://doi.org/10.1038/s43856-024-00653-1.

43 Dong Y, Zou JJ, Liu J, et al. MD-Vibe: Physics-Informed Analysis of Patient-Induced Structural Vibration Data for Monitoring Gait Health in Individuals with Muscular Dystrophy. Presented at the New York, NY, USA.

44 Wei W-L, Lin J-C, Liu T-L, Liao H-YM. Capturing Humans in Motion: Temporal-Attentive 3D Human Pose and Shape Estimation from Monocular Video 2022. https://doi.org/10.48550/arXiv.2203.08534.

45 Shin S, Kim J, Halilaj E, Black MJ. WHAM: Reconstructing World-Grounded Humans with Accurate 3D Motion. Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

NEJM AI AUTHOR ACCEPTED MANUSCRIPT

	Total	FSHD	DM	CTL
Participants	129	28	58	43
Location				
High-throughput study	94	26	27	43
Community events	30	1	27	2
Neuromuscular clinic	5	1	4	0
Sex				
Female (%)	57 (44%)	4 (14%)	31 (53%)	22 (51%)
Age in years				
Mean (SD)	43 (16)	48 (16)	41 (14)	44 (17)
BMI				
Mean (SD)	25.2 (5.5)	25.8 (4.6)	25.4 (6.6)	24.7 (4.4)

Table 1: Study participant characteristics. (FSHD = facioscapulohumeral muscular dystrophy, DM = myotonic dystrophy, CTL = control)

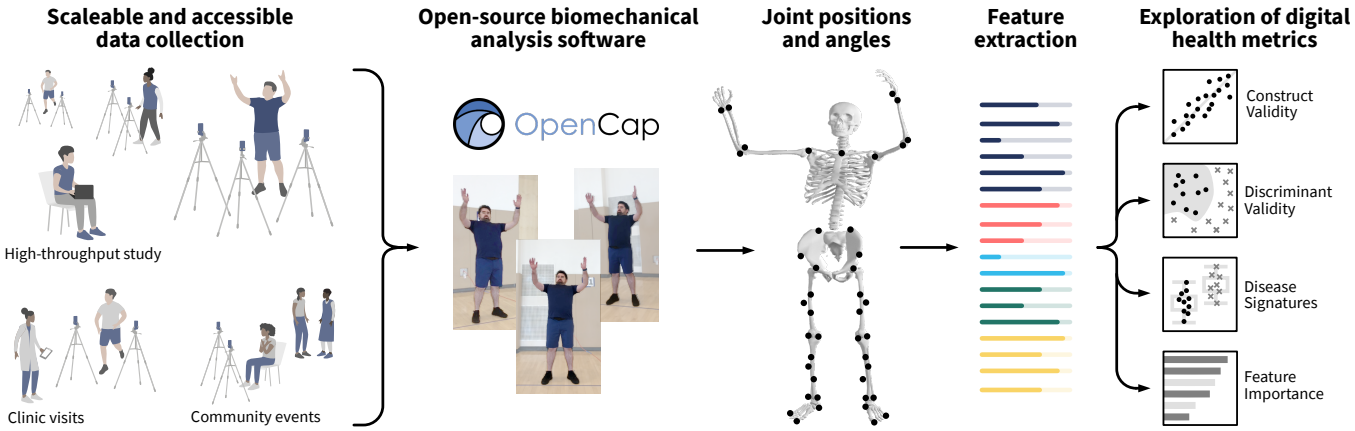

Figure 1: Video-based biomechanical analysis. Video-based biomechanical analysis enables scalable and portable data collection for studying human movement. We used the open-source OpenCap application[20] to convert smartphone videos to three-dimensional anatomical marker positions and joint angles. Leveraging clinical domain knowledge, we engineered a set of video-based metrics describing aspects of human motion known to have salience for the neuromuscular diseases we study. These features are then used to explore the design of novel digital health metrics. (Identifiable photos shared with informed consent.)

Source: created by authors

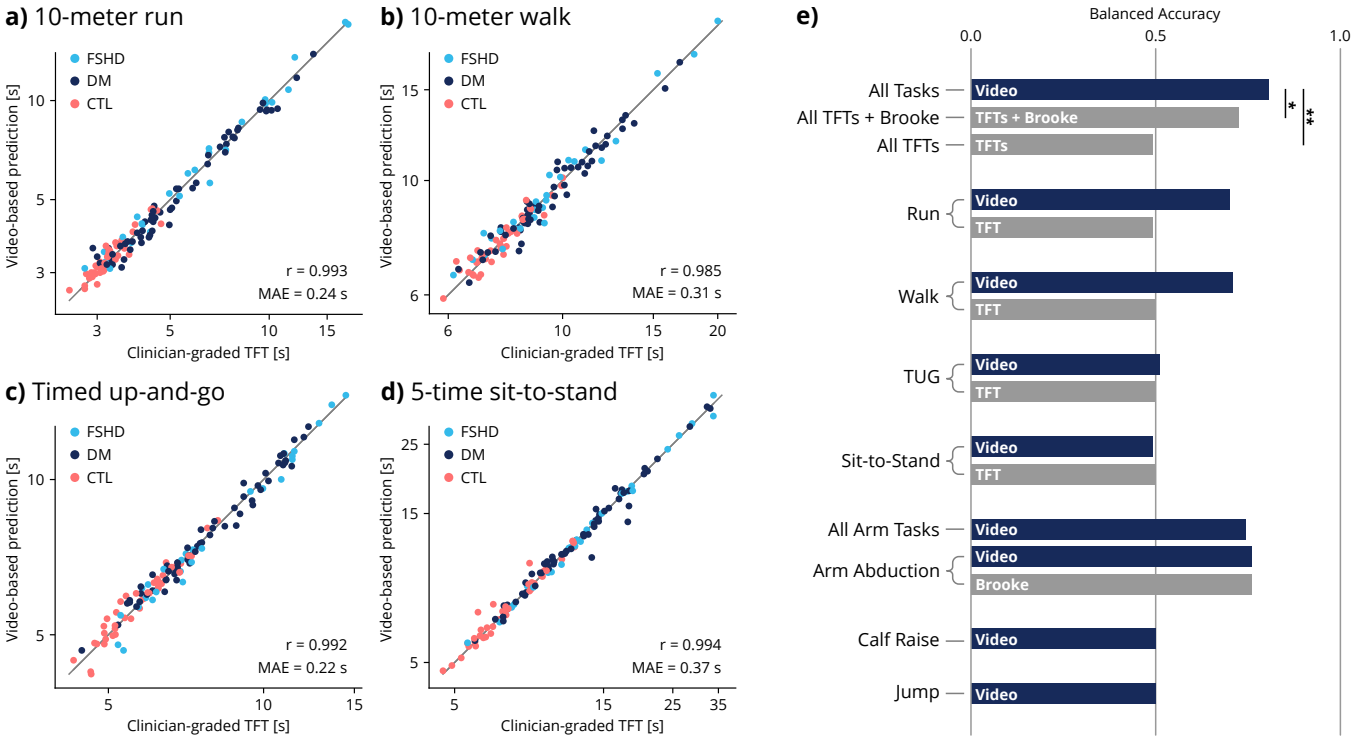

Figure 2: Construct and discriminant validity. To estimate human-graded timed function tests (TFT) from video-based biomechanical data, we computed related features (e.g., walk speed) and trained a linear regressor to predict the TFT time (e.g., walk time) based on the video-based features to account for systematic differences between the scores. Here, we show the video-predicted TFT times using a linear regressor fit with leave-one-subject-out cross validation for **a)** 10-meter run, **b)** 10-meter walk, **c)** timed up-and-go (TUG), and **d)** 5-time sit-to-stand (5xSTS) activities. **e)** Binary classification models were trained to distinguish between participants with facioscapulohumeral muscular dystrophy (FSHD) and myotonic dystrophy (DM). A model using video-based biomechanical metrics from nine activities is more accurate than models using timed function test tests (TFTs) and the Brooke Upper Extremity Scale. Video models using running and arm range of motion tasks also outperformed clinical scores. Classification model performance is measured by balanced accuracy, which corrects for unequal class sizes; 0.5 corresponds to always predicting the more common disease. Statistical differences in performance are tested using the McNemar test (*$p < 0.05$, **$p < 0.01$). (r = Pearson correlation, MAE = mean absolute error, CTL = control)

Source: created by authors

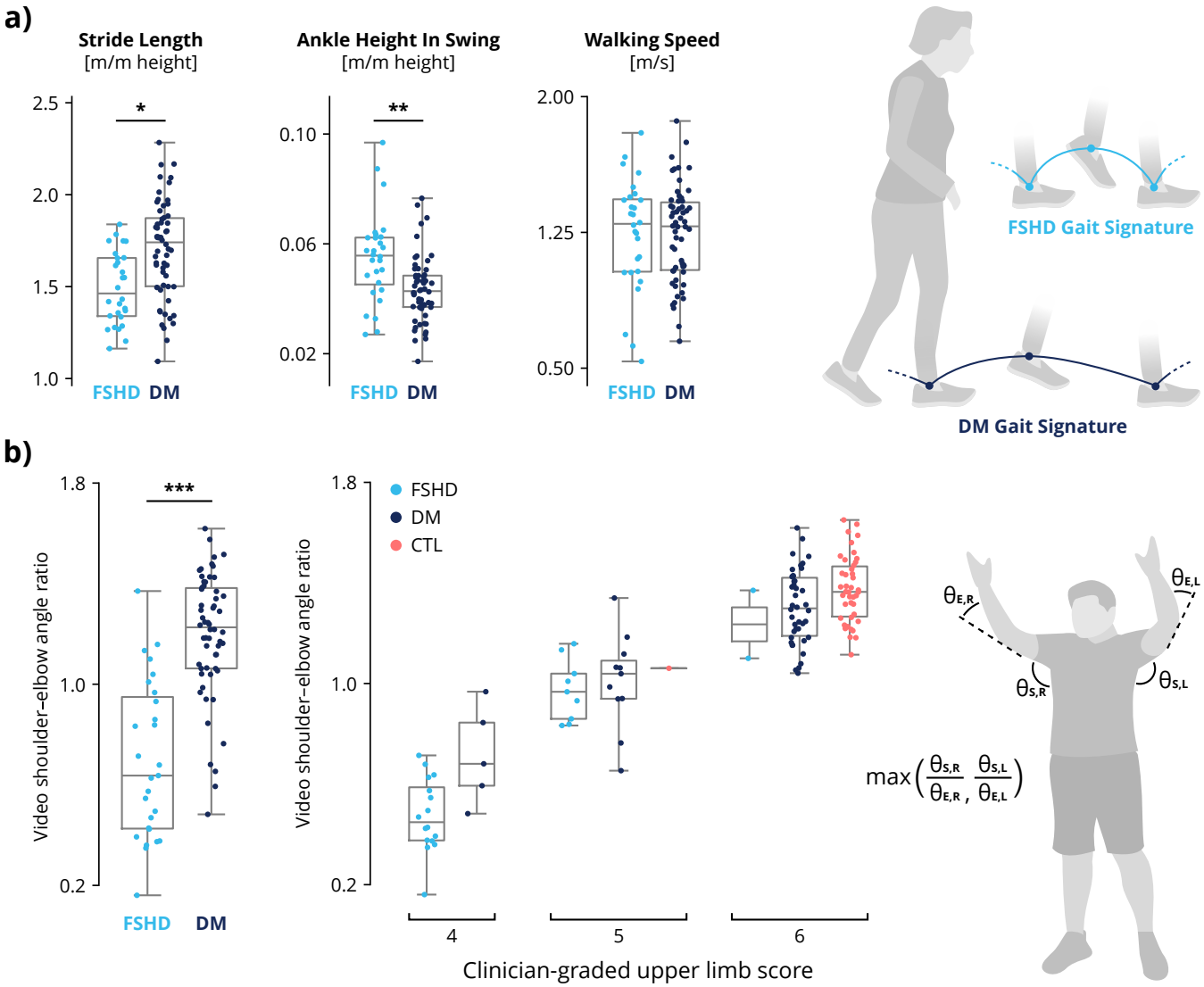

Figure 3: Disease-specific movement signatures. Video-based analysis reveals disease-specific movement signatures in gait and upper-limb range-of-motion. **a)** Patterns in walking gait differ between FSHD and DM. Despite similar walking speeds, the FSHD group walked with shorter stride lengths (p = 0.037) and higher ankle height during swing (p = 0.008) than did the DM group. ("m/m height" indicates units of meters normalized by participant height in meters, i.e. dimensionless.) **b)** Video-based biomechanical analysis of upper-body function shows differences between diseases (p = < 0.001). A video-based metric agrees with (Spearman's rank correlation r = 0.80). the human-graded Brooke Upper Extremity Scale, which ordinally scores the ability to raise arms overhead with straight elbows. To compute our analogous video-based metric, we calculate the ratio of shoulder abduction angle ($\boldsymbol{\theta_S}$) to elbow flexion angle ($\boldsymbol{\theta_E}$) and take the maximum ratio between right and left sides. Statistical significance is measured with a two-sided Kolmogorov–Smirnov test with Bonferroni corrections to control for Type I error across all 34 features. (FSHD = facioscapulohumeral muscular dystrophy, DM = myotonic dystrophy, CTL = control)

Source: created by authors

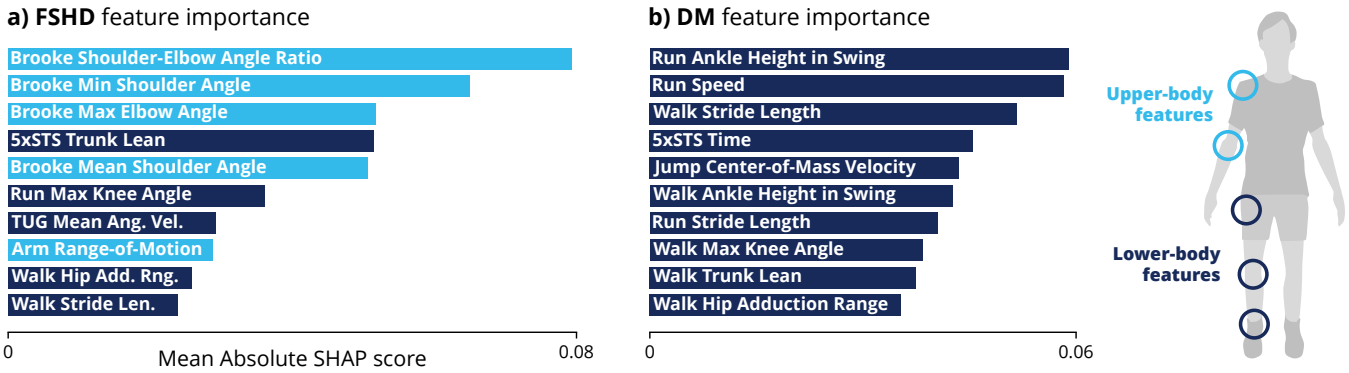

Figure 4: Disease-specific feature importance. Binary classification models were trained to distinguish control participants versus those with **a)** facioscapulohumeral dystrophy (FSHD) or **b)** myotonic dystrophy (DM). Feature importance is measured by the mean absolute Shapley additive explanations (SHAP) values, which indicate the average amount each feature impacts the classifier's probability of predicting disease. The top ten features are shown here; all feature importance scores are in Figure S3. The classification models are linear support vector machines. The FSHD model relies mainly on upper-limb features, while the DM model relies mainly on lower-limb features. Descriptions of video-based features are in Table S3. (5xSTS = 5-time sit-to-stand, TUG = timed up-and-go)

Source: created by authors