Zurich University
of Applied Sciences

School of
Health Sciences

Toward Ground Truth: Evaluating Therapists' Reliability in Video-Based Assessments of Upper Limb Compensatory Movements Post-Stroke

Lena Sauerzopf[1,2], Celina G. Chavez Panduro[3], Andreas R. Luft[4,5], Benjamin Kühnis[6], Elena Gavagnin[6,7], Tim Unger[8,9], Christopher Easthope Awai[8], Josef G. Schönhammer[4], Jürgen Degenfellner[10], Martina R. Spiess[1]

[1]ZHAW School of Health Sciences, Institute of Occupational Therapy, Winterthur, Switzerland, [2]Faculty of Medicine, University of Zurich, Zurich, Switzerland, [3]University for Continuing Education Krems, Krems, Austria, [4]Neuroscience of Motivation and Cognition in Rehabilitation (NeuroCoRe) Lab University of Zurich and University Hospital Zurich, Zurich, Switzerland, [5]Cereneo Center for Neurology and Rehabilitation, Vitznau, Switzerland, [6]ZHAW School of Management and Law, Institute of Business Information Technology CH, [7]ZHAW School of Engineering, Centre for Artificial Intelligence CH, [8]Data Analytics and Rehabilitation Technology (DART) Lab, Lake Lucerne Institute, Vitznau, Switzerland, [9]Rehabilitation Engineering Laboratory, ETH Zurich, Zurich, Switzerland, [10]ZHAW School of Health Sciences, Institute of Physiotherapy, Winterthur, Switzerland

Introduction

Upper limb compensatory movement patterns, possibly leading to negative long-term consequences, often occur during motor recovery post stroke [1]. Today, the evaluation of compensatory movements is usually based on therapists' visual observation from a 3D perspective [2]. Machine learning methods have shown promise in assessing movement quality [3,4]. However, the reliability of assessing compensatory movements through video data (2D-view) remains unclear.

Aim [5]

- To evaluate the reliability of therapists' visual assessments in rating compensatory movement patterns using video analysis.
- The motion analysis focused on the rating of upper limb and trunk movements during a standardized drinking task [6] in post-stroke individuals.

Methods [5]

- 22 experienced therapists rated 7 anonymized videos.
- The videos showed individuals with hemiparesis performing a standardized drinking task [6] using the impaired upper limb.
- Compensatory movements were rated on a scale from 0 (no compensation) to 100 (largest imaginable compensation).
- 2 labeling rounds with a 14-day interval.
- Determination of the degree of agreement within (first versus second labeling round) and between the different Raters [7].



Figure 1. Excerpt from one anonymized video presented to the therapists.

Results [5]

Categorizing the ICC according to [8] shows:

- The *drinking phase* demonstrated excellent intra-rater reliability and good inter-rater reliability
- Moderate to good reliability was observed in the *reaching*, *forward transport, back transport* and *returning phases.*
- The global domains of *smoothness and interjoint coordination*, evaluated across all five phases, indicated poor reliability.
- Wide CIs appeared in all domains

Phases of Movement [6]	ICC (95% CI)
Reaching	0.76 (0.44-1) 0.73 (0.22-1)
Forward Transport	0.68 (0.38-1) 0.84 (0.20-1)
Drinking	0.88 (0.60-1) 1 (0.24-1)
Back Transport	0.65 (0.31-1) 0.67 (0.18-1)
Returning	0.75 (0.43-1) 0.76 (0.21-1)
Global	**Overall ICC (95% CI)**
Smoothness	0.49 (0.18-1) 0.13 (0.02-0.86)
Interjoint coordination	0.40 (0.13-1) 0.19 (0.03-0.1)

Table 1. Overview of inter-rater and intra-rater reliability results.
ICC = Intraclass Correlation;
CI = Credible Interval

Conclusion [5]

- Based on our results, we cannot recommend establishing a ground truth for the automated analysis of compensatory movement patterns based solely on therapists' video ratings.
- Further research is needed to evaluate the key factors that enable therapists to reliably assess movement patterns in a two-dimensional perspective.

References

1. Levin, M. F., Kleim, J. A., & Wolf, S. L. (2009). What Do Motor "Recovery" and "Compensation" Mean in Patients Following Stroke? Neurorehabilitation and Neural Repair, 23(4), 313–319.
2. Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., & Bermúdez I Badia, S. B. (2021). A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–14.
3. Mennella, C., Maniscalco, U., De Pietro, G., & Esposito, M. (2023). The role of artificial intelligence in future rehabilitation services: a systematic literature review. IEEE Access, 11, 11024-11043.
4. Rahman, S., Sarker, S., Haque, A. N., Uttsha, M. M., Islam, M. F., & Deb, S. (2022). AI-driven stroke rehabilitation systems and assessment: A systematic review. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 192-207.
5. Sauerzopf, L., Panduro, C. G. C., Luft, A. R., Kühnis, B., Gavagnin, E., Unger, T., Schönhammer, J., & Spiess, M. R. (2024). Evaluating inter-and intra-rater reliability in assessing upper limb compensatory movements post-stroke: creating a ground truth through video analysis?. *Journal of NeuroEngineering and Rehabilitation*, *21*(1), 217.
6. Alt Murphy, M., Murphy, S., Persson, H. C., Bergström, U.-B., & Sunnerhagen, K. S. (2018). Kinematic Analysis Using 3D Motion Capture of Drinking Task in People With and Without Upper-extremity Impairments. Journal of Visualized Experiments : JoVE, 133, 57228.
7. Mehrholz, J., Thomas, S., & Elsner, B. (2016). Beurteilung von Assessments oder Testgütekriterien. neuroreha, 8(02), 62-67.
8. Portney, L. G., & Watkins, M. P. (2009). Foundations of clinical research: applications to practice (Vol. 892, pp. 11-15). Upper Saddle River, NJ: Pearson/Prentice Hall.

Contact

ZHAW Zurich University of Applied Sciences - School of Health Sciences - Institute of Occupational Therapy
Martina Spiess, martina.spiess@zhaw.ch | Celina G. Schweizer-Chavez, chavez.p.celina@gmail.com