Imperial College
London

12 June 2024

Modelling Heterogeneity in the Trajectory of Duchenne Muscular Dystrophy

Vignesh Balaji (CID: 01849526)

Supervisor: Dr C.Hallsworth

Imperial College London

Duchenne Muscular Dystrophy (DMD)

- A devastating muscle wasting condition that affects young boys, caused by a mutation in the dystrophin (DMD) gene [1–4].
- The expected life expectancy ranges from 20 to 40 years [3–5].
- The worldwide prevalence for this disease is 4.78 per 100,000 - very rare [1, 2].
- The DMD gene is the largest known gene in the human genome - a lot of variability in patient symptoms because a lot can go wrong [6, 7].

Imperial College London

North Star Ambulatory Assessment (NSAA)

- A functional (motor ability) measure of disease severity - consists of integers between 0 and 34 [1–3, 8].
- Composed of 17 elements e.g. Walk, Stand, Jump etc.
- Each item is scored on an integer scale 0-2, which is then summed to give the NSAA [1].

The NSAA Component Integer Scale	
0	Unable to perform task
1	Able to perform task with assistance
2	Able to perform task independently

Imperial College London

The problem

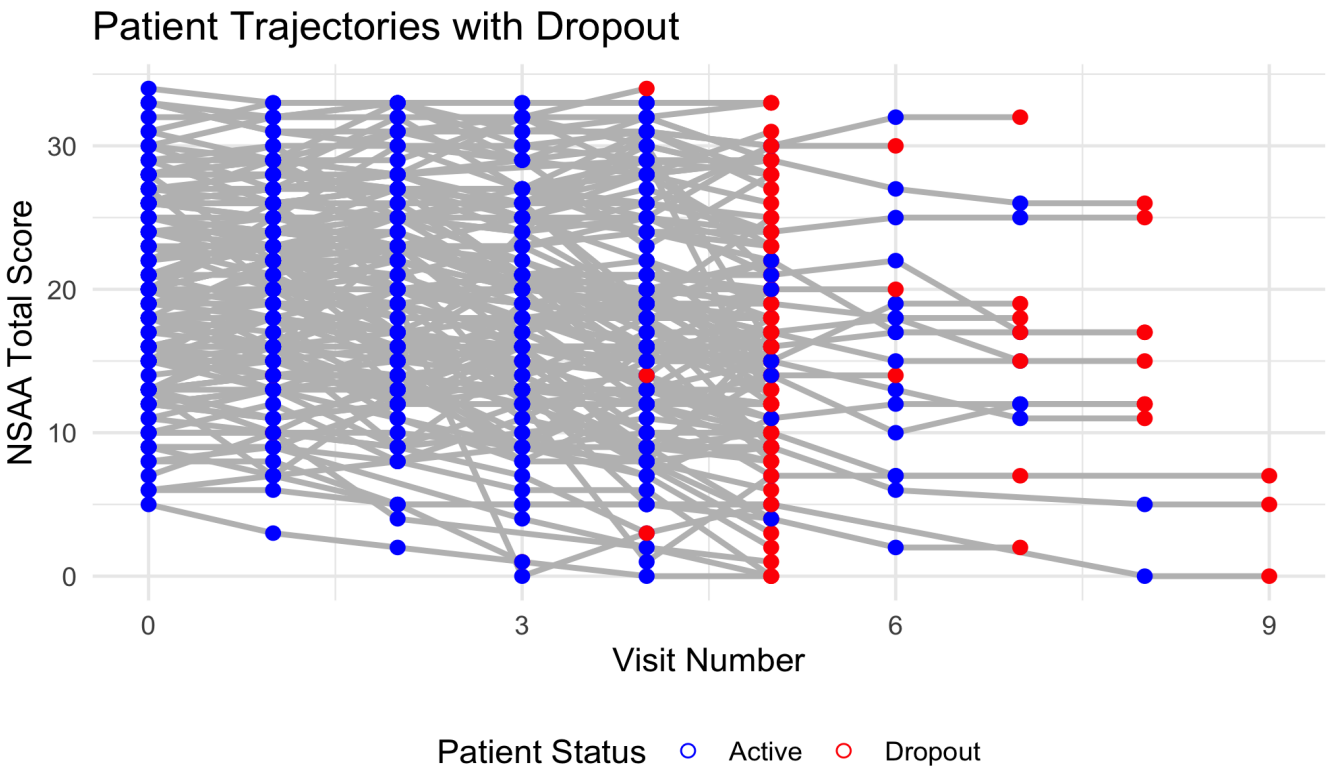

Figure 1: Individual placebo group patient trajectories, from a phase 3 trial [3]. Large cross-sectional and temporal variability in patient scores makes it hard to identify nuanced treatment effects.

Imperial College London

Project Aim

Modelling Clinical Heterogeneity

Advise clinicians on how to better capture the differences in symptoms between patients.

How did I achieve this?

- Analysed a dataset of placebo group patients [3], and fitted a generalised least squares (GLS) model.
- Performed statistical diagnostics on the GLS model and interpreted estimated parameters.
- Specified an ideal clinical trial design (sample size and patient characteristics) through simulation - power analysis.
- Proposed a more stable and meaningful measurement - principal component analysis.

Imperial College London

What is wrong with the NSAA?

- The NSAA assumes each component is equally important and useful at differentiating between patient severity.
- Easy & difficult elements of the NSAA are not particularly variable.
- “Jump”, “Hop”, “Climb Box Step” appear to be the most variable - most useful.

Imperial College London

What is wrong with the NSAA?

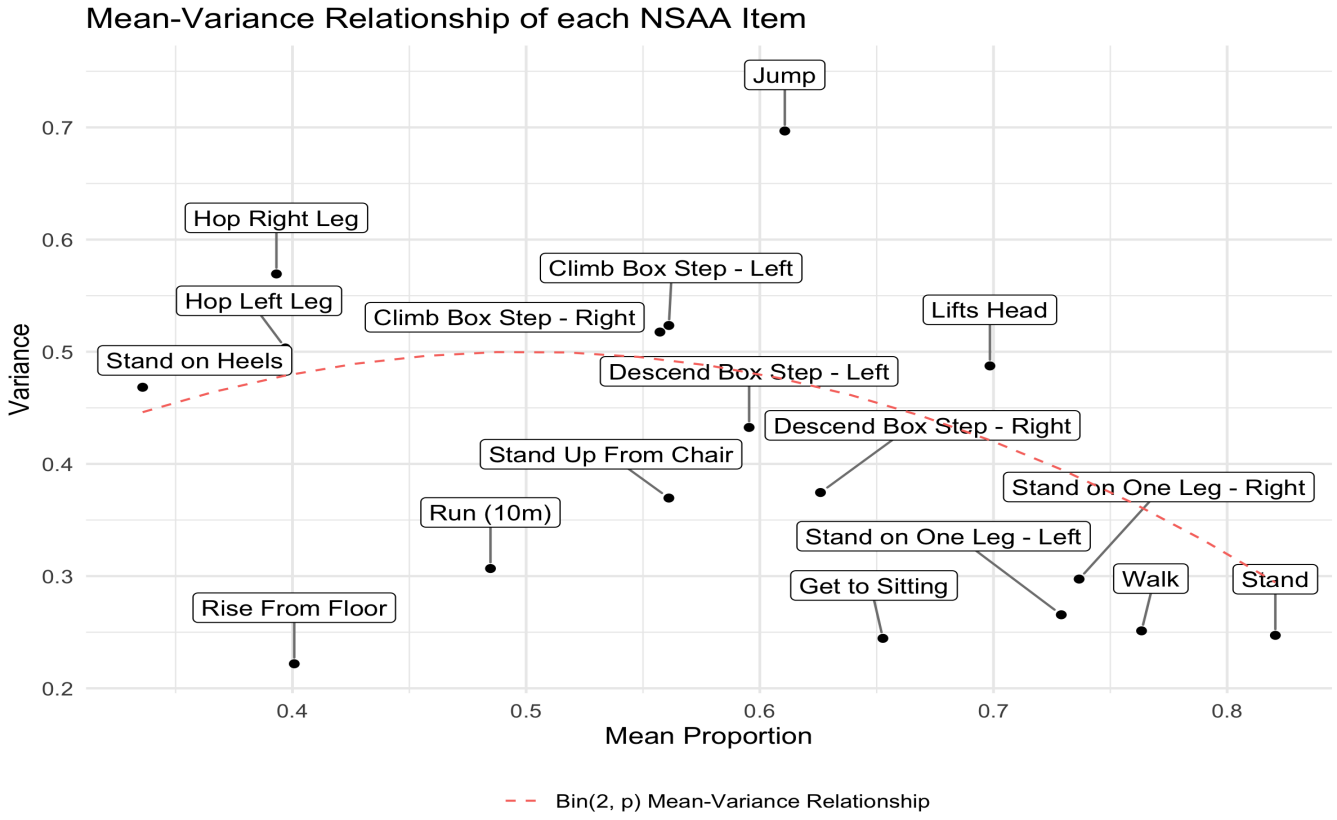

Figure 2: The mean-variance relationship of NSAA components (at baseline) is loosely aligned with the mean-variance relationship from a Bin$(2, p)$ - scaled Bernoulli.

Imperial College London

What is wrong with the NSAA?

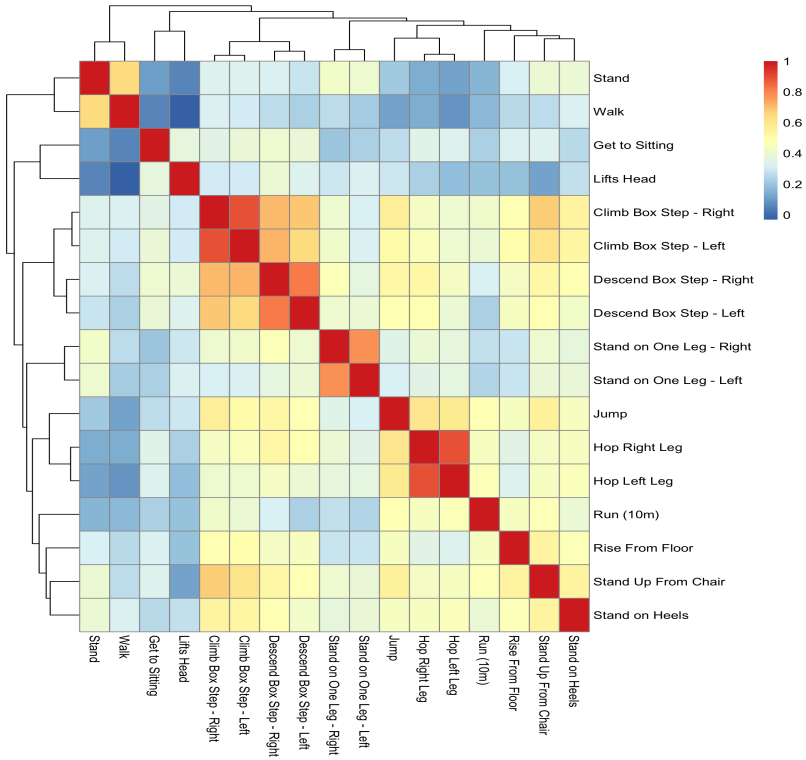

Figure 3: The heatmap and dendogram illustrate correlations and clustering between certain activites.

- Do we need all 17 items of the NSAA? Do they add value?
- Elements performed on left and right legs are highly correlated - symmetric vs asymmetric? [9, 10]
- Correlations and clustering between items that are similar e.g. “Jump” & “Hop”, “Stand” & “Walk”

Imperial College London

Temporal Mean-Variance Relationship

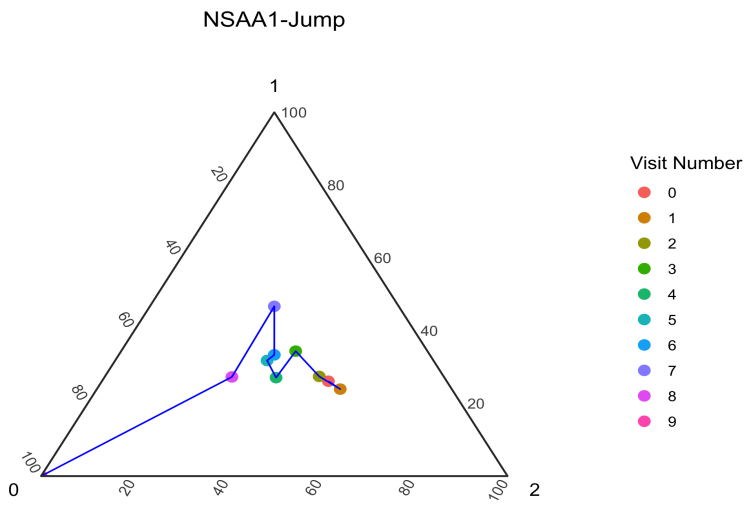

Figure 4: The proportion of 0s, 1s and 2s are near the centroid, thus are balanced through time.

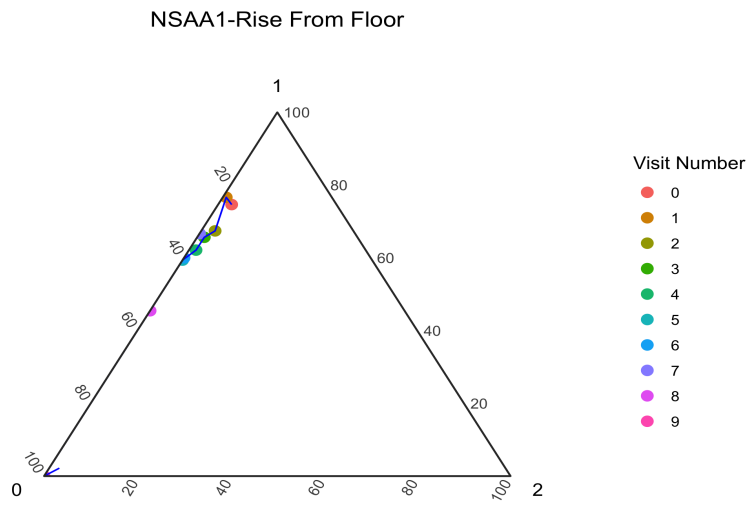

Figure 5: The proportion of 0s, 1s and 2s are biased towards the 0-1 edge, indicating a challenging task.

Principal Component Analysis (PCA)

- PCA is a statistical technique commonly used for condensing the most vital information in a dataset
- It linearly transforms the original variables into orthogonal directions of maximum variance - known as principal components or axes (PCs).
- PC1 the direction of largest possible variance in the dataset, PC2 is second largest possible and orthogonal to the first ...
- It can be shown that the principal components of any centred dataset matrix X comes from the eigendecomposition of its covariance matrix $\frac{1}{n}X^TX$ [11].

Imperial College London

A PCA Solution

Proposed Solution

Linearly transform the current NSAA framework, which aggregates the marginal scores, using PCA.

- Perform PCA on the unstandardised baseline marginal scores from placebo group [3] - justification later.
- Use the coefficients/weights/loadings of PC1 to specify a new linear combination.
- We allow the higher utility items to have more influence on a patient's overall score.

Imperial College London

The Weights

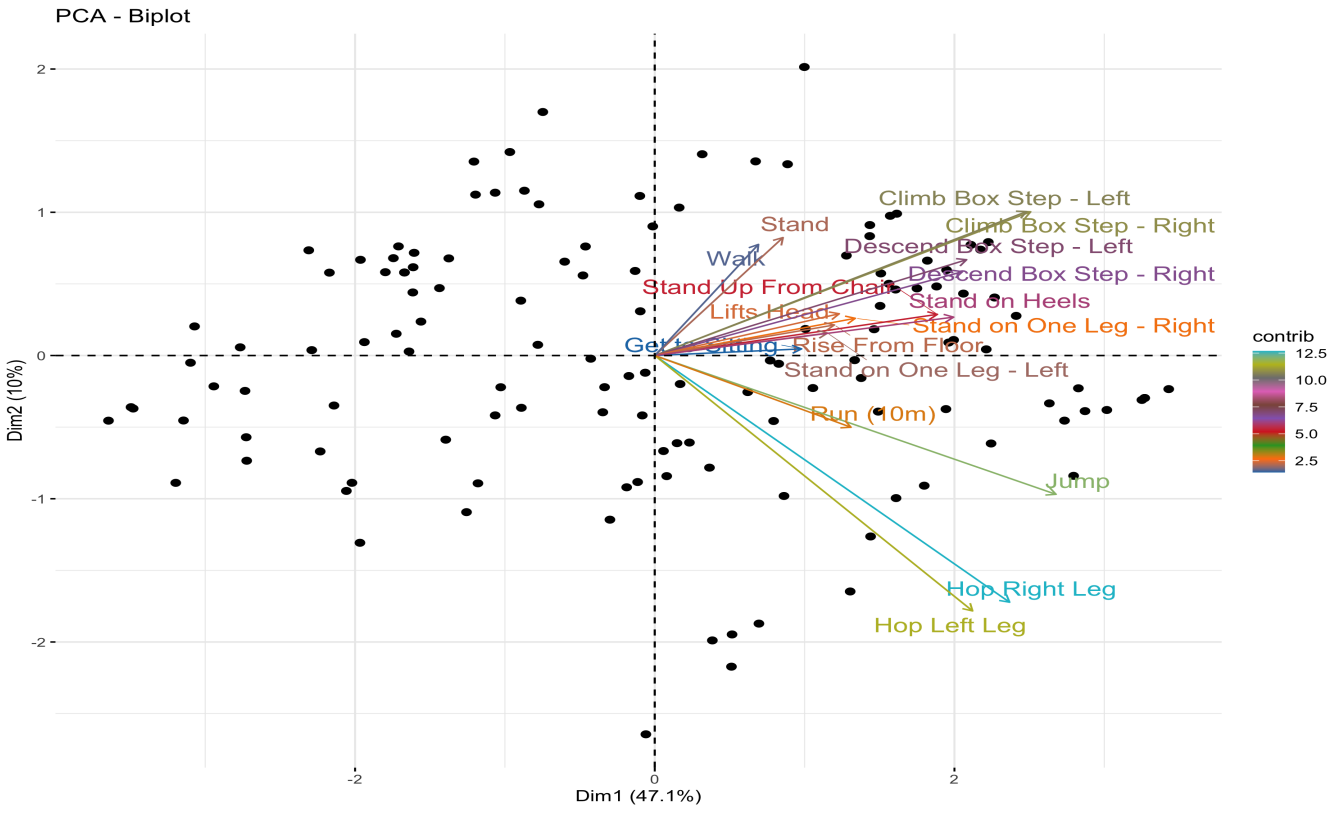

Figure 6: Diagrammatic representation of the weights of each variable onto PC1 (x-axis) and PC2 (y-axis). All items are positively weighted onto PC1, but not true for PC2.

Imperial College London

Outcome

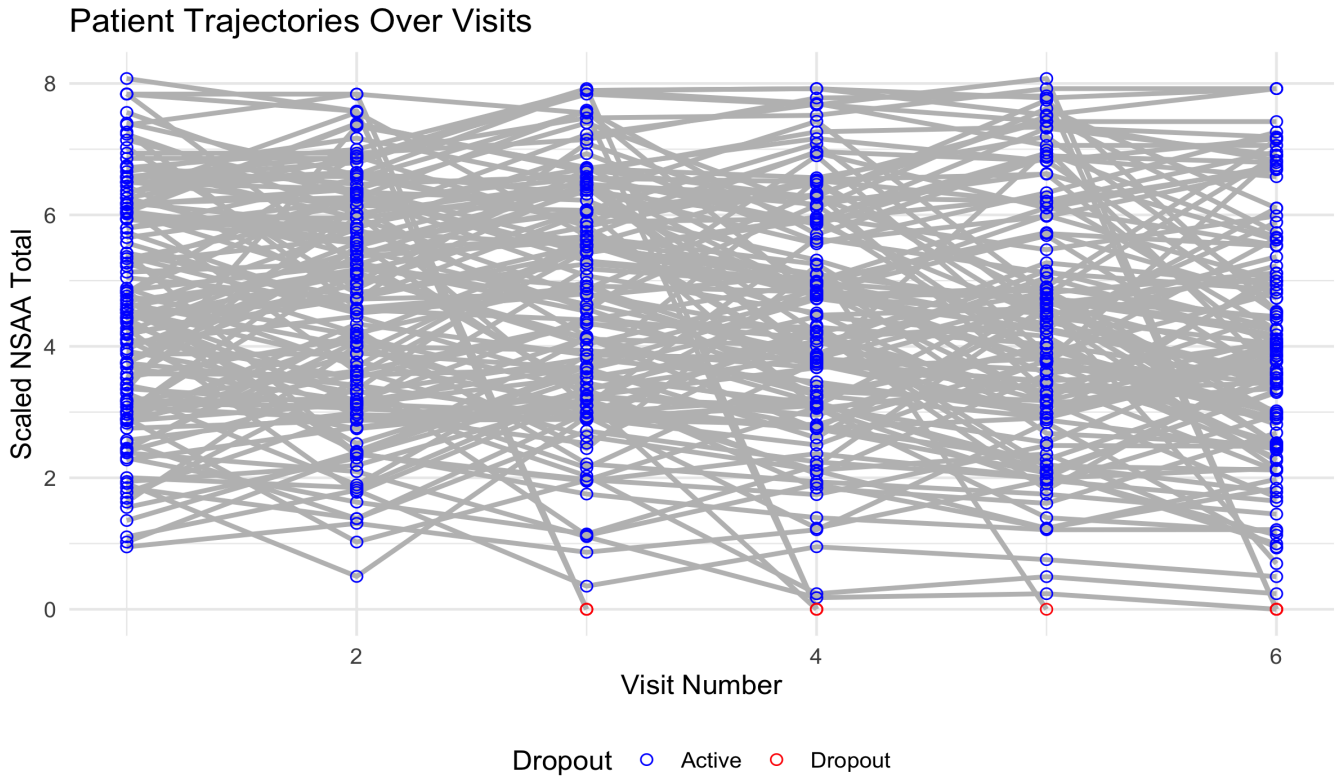

Figure 7: Scaled NSAA total is derived from projecting the marginal scores (at each visit) onto PC1 (i.e. matrix-vector multiplication). A significant reduction in cross-sectional variability. So, can detect more nuanced treatment effects. But no change in temporal variability.

Why Unstandardised? Why Baseline

- We argued that z-transformation (i.e. centre and divide by variance) was too harsh and removed the mean-variance relationships seen above - we claim this is useful.
- Also the unstandardised scores led to more time independent (stable) weights.
- A priori, we won't have future patient scores, so it is not possible to use the weights from each patient visit. Pragmatically, can only perform PCA on baseline scores to get a certain measurement.

Imperial College London

Limitations & Further Work

- Treated the scores of each NSAA item as numerical values - are they not ordinal?
- PC1 only explained $\sim 47\%$ of the variance - a unidimensional measure might not be appropriate, in fact, 7 PCs were needed to explain 80% of the variability.
- We didn't explore other less stringent variance stabilising methods.
- We performed PCA independently on each visit - could be reason for time dependency in weights (recall there was dropout) - sequentially update weights using Kalman filter?
- We need to test versatility of new measurement.

Imperial College London

Bibliography I

[1] Francesco Muntoni et al. "Categorising trajectories and individual item changes of the North Star Ambulatory Assessment in patients with Duchenne muscular dystrophy". In: *PLoS One* 14.9 (2019), e0221097.

[2] Francesco Muntoni et al. "DMD genotypes and motor function in duchenne muscular dystrophy: a multi-institution meta-analysis with implications for clinical trials". In: *Neurology* 100.15 (2023), e1540–e1554.

[3] Ronald G Victor et al. "A phase 3 randomized placebo-controlled trial of tadalafil for Duchenne muscular dystrophy". In: *Neurology* 89.17 (2017), pp. 1811–1820.

Imperial College London

Bibliography II

[4] Eric P. Hoffman, Robert H. Brown and Louis M. Kunkel. "Dystrophin: The protein product of the duchenne muscular dystrophy locus". In: *Cell* 51.6 (1987), pp. 919–928. ISSN: 0092-8674. DOI: https://doi.org/10.1016/0092-8674(87)90579-4. URL: https://www.sciencedirect.com/science/article/pii/0092867487905794.

[5] Dongsheng Duan et al. "Duchenne muscular dystrophy". In: *Nature Reviews Disease Primers* 7.1 (2021), p. 13.

[6] Quan Gao and Elizabeth M McNally. "The Dystrophin Complex: structure, function and implications for therapy". In: *Comprehensive Physiology* 5.3 (2015), p. 1223.

Bibliography III

[7] Isabelle Desguerre et al. "Clinical heterogeneity of duchenne muscular dystrophy (DMD): definition of sub-phenotypes and predictive criteria by long-term follow-up". In: *PloS one* 4.2 (2009), e4347.

[8] Giorgia Coratti et al. "Longitudinal natural history in young boys with Duchenne muscular dystrophy". In: *Neuromuscular Disorders* 29.11 (2019), pp. 857–862.

[9] Liting Tang, Shuran Shao and Chuan Wang. "Electrocardiographic features of children with Duchenne muscular dystrophy". In: *Orphanet Journal of Rare Diseases* 1 (2022), p. 320. DOI: 10.1186/s13023-022-02473-9. URL: https://doi.org/10.1186/s13023-022-02473-9.

Bibliography IV

[10] Qiliang Xiong et al. "Gait asymmetry in children with Duchenne muscular dystrophy: evaluated through kinematic synergies and muscle synergies of lower limbs". In: *BioMedical Engineering OnLine* 22.1 (2023), p. 75.

[11] Ian Joliffe. "Principal Component Analysis for Special Types of Data". In: *Principal Component Analysis*. New York, NY: Springer New York, 2002, pp. 1–6. ISBN: 978-0-387-22440-4. DOI: 10.1007/0-387-22440-8_13. URL: https://doi.org/10.1007/0-387-22440-8_13.