KU LEUVEN

FACULTY OF SCIENCE

Muscle Strength and Functional Mobility in Children with Acute Lymphoblastic Leukaemia

Solon Karapanagiotis

Supervisor: Prof. G. Verbeke

Co-supervisor: Prof. T. Troosters

Thesis presented in
fulfillment of the requirements
for the degree of Master of Science
in Statistics

Academic year 2015-2016

Contents

Preface

Longitudinal studies play a prominent role in the health sciences, as well as in public health. They are indispensable to the study of change in an outcome over time. By measuring study participants repeatedly through time, longitudinal studies allow the direct study of temporal changes within individuals and the factors that influence change. In practice, one is often confronted with multiple outcomes, all measured repeatedly over time. In such cases, depending on the research questions, one could model the joint covariance of the longitudinal outcomes. For instance, a joint modelling strategy is inevitable when the interest is in assessing the relation between some covariate and all outcomes simultaneously, in studying how the association between the various outcomes evolves over time, or in investigating the association between the evolutions of all outcomes.

Although there has been extensive research in joint models in the last two decades, they are still not widely used in practice. We believe this is primarily due to the fact that most discussions of joint model methods have been in the technical statistical literature.

The goal of this dissertation is to use existing methodology to model univariate and multivariate longitudinal data and present the methods and results in a meaningful way for both a clinically and a statistically oriented audience. The multivariate analysis will allow to answer the research questions that could not be addressed otherwise. As a result of this conceptual approach, the report is divided into two parts. Part I is oriented towards a subject-matter scientist and focusing on answering the research questions and on their clinical implications. Technical details on the statistical methodology are avoided. These are given in Part II. Thus, Part I starts with an intro-

duction on the topic and gives background information. It continues with the methodology, the results, and ends with a discussion on the findings. Conversely, Parts II focuses on statistics. The idea is to present the key concepts of the model building approach. First, a review of the notation and study design are presented. After that univariate and joint models follow. Part II ends by addressing the issue of missing data.

All the code used in the analysis is available at

https://github.com/karapa/Master-thesis/tree/master

The compilation of this dissertation would not have been possible without the guidance of prof. Geert Verbeke and Thierry Troosters. Their enlightening and support helped me the entire research period and in writing this thesis. I would never have been able to finish my project without the tremendous help of prof. Anne Uyttebroeck and Veerle Dirix and I really appreciate their patience and their kindness to me.

Abstract

Purpose: The aim of this study is to explore the time course and the predictors of muscle strength and functional mobility in children during treatment for acute lymphoblastic leukaemia (ALL). More specifically three questions are addressed: What is the trend in functional mobility and muscle strength during the treatment period for all ALL children? What are the differences, if any, between different risk group subjects with respect to muscle strength and functional mobility during the last year of treatment? What is the association between different muscle groups?

Methods: Forty-two children, aged between 4 and 15 years, were included. They were treated according to the European Organization for Research and Treatment of Cancer (EORTC) 58081 protocol, based on a Berlin-Frankfurt-Munster (BFM) scheme. Muscle strength (knee extension, ankle dorsiflexion, hand grip and pinch grip) was measured with a dynamometer and functional mobility with the 6 min walk test (6MWT) at 8 timepoints across the two-year treatment period. The trends were investigated over time in univariate and multivariate models.

Results: Children receiving treatment against ALL exhibit fluctuating trends of muscle strength and functional capacity. These trends are more severely negative in the first two months of treatment. Risk group and height are significant determinants of the functional mobility (p-value$<$0.001). Age, risk group and weight are significant determinants of the evolution in muscle strength (p-value$<$0.001). There are no significant differences, between risk group subjects with respect to functional mobility. The differences in muscle strength between average risk 1 (AR1) and average risk 2-B (AR2-B) group and very low risk (VLR) and average risk 2-T (AR2-T) are significant

(p-value<0.001). Knee extension strength is positively correlated with ankle dorsiflexion strength (mean [CI]: 0.56 [0.29, 0.75]) as well as hand grip (mean [CI]: 0.39 [0.18, 0.57]). No other significant associations were found.

Conclusions: Children receiving treatment against ALL exhibit impairments in muscle strength and functional capacity early during the course of treatment. Risk group is a determinant of these impairments.

Impact: Children may benefit from early rehabilitation that includes muscle strengthening and functional mobility components to address presenting skeletal and motor impairments.

Keywords: Acute lymphoblastic leukaemia, paediatric, muscle strength, functional mobility, multivariate longitudinal analysis.

List of Abbreviations

6MWD distance walked.

6MWT 6 min walk test.

AD ankle dorsiflexion.

ALL acute lymphocytic leukaemia.

AR average risk.

AR1 average risk 1.

AR2-T average risk 2-T.

AR2-B average risk 2-B.

BFM Berlin-Frankfurt-Munster.

EORTC European Organisation for Research and Treatment of Cancer.

HG hand grip.

KE knee extension.

LR likelihood-ratio.

ML maximum likelihood.

OLS ordinary least squares.

PG pinch grip.

REML restricted maximum likelihood.

VHR very high risk.

VLR very low risk.

Part I

1. Introduction

The percentage of the total population suffering from childhood cancer is approximately 144 to 148 cases per million [1]. Leukaemia is the most common childhood cancer, accounting for almost 1 out of 3 cases. About 3 out of 4 leukaemias among children and teens are acute lymphocytic leukaemia (ALL)[2]. Optimal use of existing antileukaemic agents and improved supportive care have improved the 5-year survival rate of ALL childhood to ~85 %[3–10].

Therefore, side effects of treatment have gained importance. Children with ALL undergo medical interventions comprising two years of chemotherapy [11]. Researchers have theorized that physical performance may be negatively affected by chemotherapy adverse effects, such as neuropathy, motor function impairment, and/or cardiopulmonary toxicity [12, 13]. Vincristine and corticosteroids (e.g. dexamethasone) can cause obesity, growth hormone deficiency and insulin resistance which are accompanied by a decline in muscle mass and by muscle weakness [14, 15]. Vincristine may also lead to axonal peripheral neuropathy. This neuropathy can provoke limitations in activity and mobility. Decreased muscle strength was found in children with ALL in the acute and medium term follow-up after finalising their treatment [16].

To our knowledge little research is available on muscle strength and functional mobility of children with ALL during the entire course of their treatment. In this study we hypothesise that the main differences between risk group children occur during the last year of the treatment (i.e., during maintenance) where the biggest deviations between treatment regimes is seen. More specifically, we hypothesise that very-high-risk children exhibit lower muscle strength and functional mobility than the other risk groups due to the one-year more intensified treatment. Also, the low and average risk group subjects are expected to differ during maintenance. AR1 and AR2-B subjects receive pulses of vincristine and corticosteroids during maintenance, which could potentially affect their physical performance.

Hence, the main research questions addressed in this study are: What is the trend in functional mobility and muscle strength in the two-year treatment

period for all ALL children? What are the differences, if any, between different risk group subjects with respect to muscle strength and functional mobility during the last year of their treatment? What is the association between different muscle groups?

1.1 Importance of the study

It is important to get a clear picture of the development of skeletal muscle abnormalities in order to guide and focus on the development of exercise training programs. Despite the positive results of exercise interventions in adult cancer patients, the evidence for benefits in childhood cancer patients is limited [1]. Studies within the population of childhood cancer patients and survivors have been initiated. However, the evolution of motor function during treatment has not been thoroughly studied. In making healthcare management decisions, participants and clinicians must be aware of the time trend of motor function in order to weigh the benefits and drawbacks of supportive care. Thus, the purpose of this study is to describe the trend of muscle strength and functional status in children with ALL, within the two-year treatment period.

2. Materials and Methods

2.1 Study population

Consecutive subjects diagnosed with ALL at the University Hospital of Leuven, at the paediatric hemato-oncology unit fulfilling inclusion criteria (age between 4 and 18 years old) are included in the study. The total population consisted of subjects diagnosed with ALL at the start of the study and children who were already on therapy before January 2013. Subjects with mental and psychomotor retardation, i.e. Down syndrome, were not included in the study because of their limited ability in cooperation and in understanding of the testing procedures. The study was approved by the Medical Ethics Committee. Parent consent and child assent, as appropriate, were obtained prior to enrolment.

2.2 Treatment

All children are treated according to the European Organisation for Research and Treatment of Cancer (EORTC) 58081 protocol, based on a Berlin-Frankfurt-Munster (BFM) scheme. Subjects were assigned to different risk groups: very low risk (VLR), average risk (AR), and very high risk (VHR). AR subjects were all children without VLR or VHR characteristics, and they were further subdivided in average risk 1 (AR1), average risk 2-B (AR2-B) and average risk 2-T (AR2-T) patient groups [17]. The treatment comprises chemotherapy for two years [11, 18, 19], based on the assigned risk group. For space considerations we do not replicate the treatment regime here but refer to [11, 17].

2.3 Design

Subjects were tested at eight different timepoints ($T_0 - T_7$) across a period of two years after diagnosis (Figure 1). The timepoints were chosen to represent important phases of the treatment: at the time of diagnosis (T_0), at the end of induction (T_1 = week 5), at the end of consolidation (T_2 = week 10 to 12), after interval therapy (T_3 = week 22), after reinduction (T_4=week 28 to 30) and during maintenance ($T_5 - T_7$ = week 52, 78, 104). Depending on the health status of the patient testing sessions could be postponed for

several days.

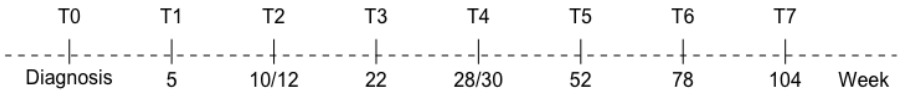


Figure 1: Timepoints when subjects were tested

2.4 Muscle strength

2.4.1 Lower extremity strength

A hand-held type MicroFET 2 (Hoggan Health Industries, United States) was used to measure isometric ankle (distal lower extremity) dorsiflexion and knee extension (proximal lower extremity) strength. The dynamometer was calibrated prior to the tests using a 2kg weight. Testing was standardised according to the Bohannon protocol [20]. For knee extension, an isometric "make" test was used in which the child had to stretch the knee with maximal force while the investigator kept the dynamometer steady. For dorsiflexion, the child had to hold his/her foot at rest while the investigator pushed back in the opposite direction. Among children between 4 and 17 years old these strength measures have intra- and inter-rater reliability and intra-class correlation coefficients ranging 0.92-0.99 and 0.85-0.97, respectively [16]. Each test was repeated three times and the peak force of the dominant lower limb was used in the analysis.

2.4.2 Grip and Pinch strength

Hand grip strength was assessed using a Jamar hydraulic hand dynamometer model 5030J1 (Patterson Medical Holdings, Inc., United States) on the dominant hand side. The arm position was determined according Joyce et al. [21]. The Jamar dynamometer provides valid measurments of hand grip strength across subjects [22]. A lateral pinch grip test was assessed on the dominant side to determine thenar strength using the Jamar hydraulic pinch gauge model 749 805 (Sammons Preston, Bolingbrok)[23]. The ranges

of intra- and inter-rater reliability for grip strength and pinch grip are 0.92-0.97 and 0.80-0.95 respectively [23–25]. Three measurements were taken and the peak was used in the analysis. Verbal encouragement was given.

2.5 Functional mobility

The 6 min walk test (6MWT) was used to assess functional mobility. Since it is self-paced, this test represents fitness for usual daily activities. The child was instructed to walk as far as possible in a 6 min time period (not running or jogging) along the 25 m course; the distance walked (6MWD) was recorded. Children were allowed to stop, slow down and rest during the test if necessary. Encouragement was standardised to short sentences like 'you are doing a good job' and 'you are almost there'. The 6MWT has been used to assess performance in ill children, including oncologic patients [26].

2.6 Statistical methods

Frequencies and percentages were calculated to describe the demographics of the study participants. Continuous variables are expressed as mean (95 % CI) unless stated otherwise. Functional mobility is assessed in a marginal multivariate model adjusted for risk group, height, age, and sex. Muscle strength (i.e. ankle dorsiflexion, knee extension, hand grip and pinch grip) is assessed in a joint random effects model adjusted for risk group, weight, age, and sex. Since the first five timepoints all children receive similar therapeutic regime we use a common risk group effect for these timepoints. The joint model contains four random effects (intercepts) to capture the association between the four outcomes. This choice is based on exploratory techniques of the mean, variance and association structure. We further relax the conditional independence assumption by allowing the error components of the four outcomes to be correlated. In both models, estimates are obtained via restricted maximum likelihood. The marginal and random effects models adjust for the correlation between repeated observations measured in the same subject and can handle longitudinal data on subjects with a varying number of observations. Moreover, the joint model allows the estimation of the associations between the four outcomes. All performed measurements are taken into account. We first fit models including all covariates, allowing for interactions with time, treated as categorical covariate. Then we per-

form backward variable selection based on Akaike information criteria and likelihood ratio tests. We present the results graphically using predicted values from the models, controlling for covariates. Confidence intervals for the association estimates are calculated using the delta method for approximating standard errors. We allow the probability that a response is missing to depend arbitrarily on observed values of the response at other times, but not additionally on the unobserved response itself [27]. Statistical significance is set a priori at alpha $< .05$. All analyses were completed using SAS software Version 9.2.

3. Results

3.1 Study population

Between 2013 and 2015, 42 children were included in the study. To be included in the analysis at least one testing should have been completed. Participants had a median age of 5.64 (range 2-15) years at diagnosis. Twenty-four (57.1%) are male. Other patient characteristics are shown in Table 1. The median number of measurements per subject was 7 (range 3-8).

Table 1: Characteristics of study participants

	n (%)
Sex	
Male	24 (57.14)
Female	18 (42.86)
Risk-group	
AR1	16 (38.10)
AR2-B	3 (7.14)
AR2-T	15 (35.71)
VHR	5 (11.90)
VLR	3 (7.14)
Age*, median (range)	5.77 (4-15)

* refers to the age (years) at study entry

3.2 Functional mobility

Risk group and height are significant determinants of the functional mobility (p-value<0.001). No significant effect of age (p-value=0.44), or gender (p-value=0.35) are observed. In Fig. 2, the patterns of 6MWD over timepoints for a child at the 50th percentile of height, by risk group, are presented. When controlling for height, there is no significant difference between the VHR group and all the others at T_5 (p-value=0.13). Similarly, there is no significant difference between the AR1 and AR2-B versus VLR and AR2-T groups at T_5, T_6, T_7 (p-value=0.20).

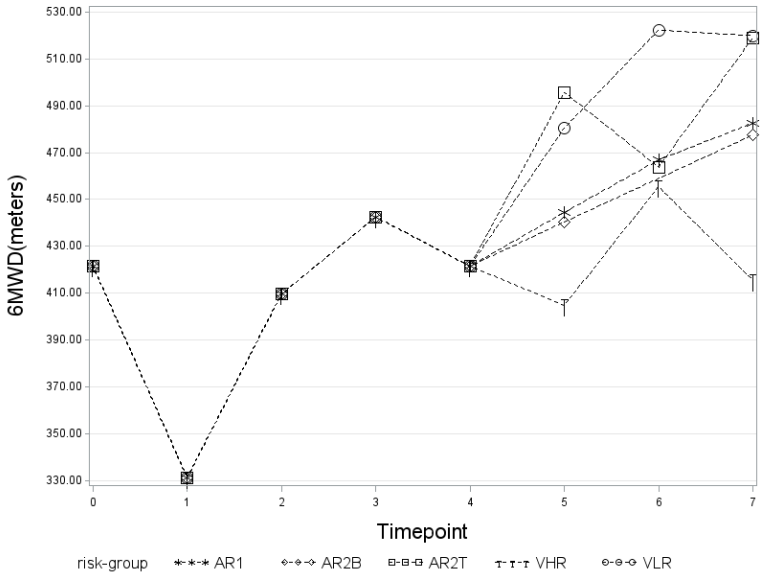


Figure 2: Estimated mean 6MWD (meters) by risk-group.

3.3 Muscle strength

Age, risk group and weight are significant determinants of the evolution in muscle strength (p-value<0.001). No significant effect of gender is shown (p-value=0.29). Fig. 3 shows the time trend of each muscle group for a child at the 50th percentile of age and weight, by risk group. After controlling for age and weight, there is no significant difference between the VHR group and all the others at T_5 (p-value=0.37). The differences in muscle strength between AR1 and AR2-B group and VLR and AR2-T group at T_5, T_6, T_7 are significant (p-value<0.001).

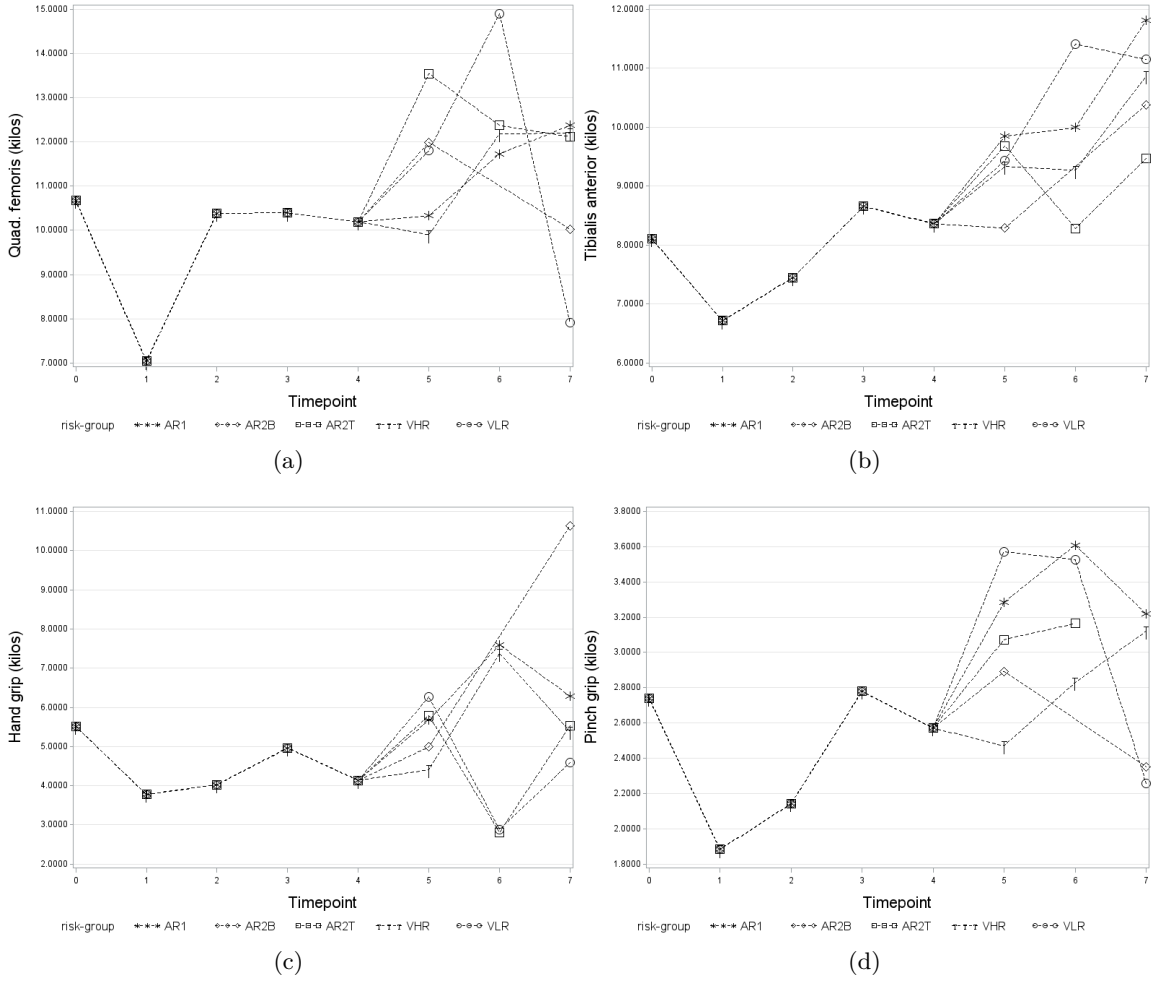


Figure 3: Estimated mean muscle strength (kilos) by risk-group; (a) quadriceps femoris; (b) tibialis anterior; (c) hand grip; (d) pinch grip.

3.4 Associations

Table 2 shows the association between responses measured at the same timepoint. Knee extension strength is positively correlated with ankle dorsiflexion (r=0.56 [0.29, 0.75]) as well as hand grip (r=0.39 [0.18, 0.57]). No other significant correlations are found.

Table 2: Estimated correlation (r) for the same timepoint with 95% CI

	r	[95% CI]
AD,KE	0.56	[0.29, 0.75]
KE,HG	0.39	[0.18, 0.57]
KE,PG	0.30	[-0.20, 0.68]
AD,HG	0.05	[-0.07, 0.16]
AD,PG	0.30	[-0.20, 0.67]
HG,PG	0.43	[-0.25, 0.83]

Abbreviations: ankle dorsiflexion (AD); knee extension (KE); hand grip (HG); pinch grip (PG)

Table 2 presents the estimated association between responses measured at different timepoints. Once more only the associations between ankle dorsiflexion and knee extension (r=0.29 [0.10, 0.46]) and the latter with hand grip (r=0.24 [0.06, 0.40]) reach statistical significance.

Table 3: Estimated correlation (r) for different timepoints with 95% CI

	r	[95% CI]
AD,KE	0.29	[0.10, 0.46]
KE,HG	0.24	[0.06, 0.40]
KE,PG	0.22	[-0.24, 0.60]
AD,HG	-0.08	[-0.20, 0.05]
AD,PG	0.21	[-0.24, 0.60]
HG,PG	0.27	[-0.29, 0.68]

Abbreviations: ankle dorsiflexion (AD); knee extension (KE); hand grip (HG); pinch grip (PG)

4. Discussion

This is the first cohort study investigating the two-year trends and predictors of functional mobility and muscle strength over time in children with ALL. This study shows children with ALL exhibit a drop in functional mobility and muscle strength after the induction phase of the treatment followed by a hectic recovery. We have not compared them with healthy children but in other studies lower measures of physiologic and functional capacity during and after cancer treatment have been reported when compared with healthy controls [1, 16, 28]. Hence, physical exercise training interventions have been proposed.

Early results for exercise training are promising [29] but not yet convincing [1]. Hang et. al. [29] provide a few limitations of the existing literature, such as the small number of studies and the insufficient sample sizes. We further believe the treatment modality reflected by the risk group of the patient should be take into account when prescribing exercise training. Having that in mind our study concentrates on the documentation of the trajectory of development of these neuromuscular problems during treatment between different risk group patients.

We first hypothesised that AR1 and AR2-B subjects perform worse than AR2-T and VLR subjects. Indeed, AR1 and AR2-B children are found to have lower muscle strength than AR2-T and VLR. The former risk groups receive pulses of vincristine and corticosteroids such as dexamethasone during maintenance [17]. Vincristine can cause motor and autonomic neuropathy [30, 31]. Musculoskeletal abnormalities including neuropathic-induced contractures have been reported in children previously treated for leukaemia with vincristine [32]. Retrospective analyses have also linked the use of methotrexate [33] and asparaginase [34] with the development of more severe persistent musculoskeletal limitations in survivors of ALL. During administration of corticosteroids, especially prolonged courses or higher doses, skeletal muscle atrophy can be expected in 40-60% of patients [35]. On the other hand, no difference is found regarding functional mobility. The effect of corticosteroids may not be prominent enough to manifest during the 6MWT. This is due to the different aims of the two testing procedures.

Strength is a measure of the ability of a muscle to develop tension and exert a force on a bony lever [36]. In this study we measured the maximal isometric force because it is theoretically simple and experimentally well controllable. Nevertheless, everyday activities necessitate neither maximal nor isometric force. The ability to appropriately and repeatedly recruit muscles is more fundamental to physical performance. The 6MWT aims to capture this functional aspect of physical performance. It assesses the level of functional exercise capacity (i.e., the ability to engage in physically demanding activities of daily living)[37].

We further hypothesised that VHR subjects would exhibit lower physical performance than the other risk groups due to the intensification of their treatment after the consolidation phase up to one year in the treatment [11]. Neither muscle strength nor functional capacity differs between VHR subjects and the other risk groups. The absence of statistical significance can be attributed to the small sample size (i.e., 5) of the VHR subjects.

It should be noticed that the tests referring to muscle strength are joint, in the sense that they relate to all outcomes simultaneously. This is the main advantage of modelling multiple outcomes together. Namely, we can have overall tests of the muscle performance.

Furthermore, the joint model allowed us to estimate the associations between muscle groups during the two-year period. The associations are mostly of moderate to small magnitude for all muscle groups. As expected, measurements at the same timepoint are, in general, more strongly associated than measurements at different timepoints. The only significant ones are between knee extension (i.e., quadriceps femoris) and ankle dorsiflexion (i.e., tibialis anterior) as well as the former with hand grip. These estimates are positive, implying that these measures act in concert, that is, a higher quadriceps force implies a higher tibialis anterior and hand grip force. This contains valuable information to the clinician since the measurement of one muscle can act as a surrogate measurement. If, for example, the measurement of the quadriceps femoris strength at any timepoint is of interest but cannot be obtained then the clinician can use the measurement of the tibialis anterior or hand grip at the same or different timepoint as a substitute to get an

indication of the quadriceps' force.

Further, it should be noticed that not all of the associations are different from zero which implies different effect of the treatment and/or of the disease on different muscle groups. For example, corticosteroids can cause steroid myopathy which is more pronounced in fast-twitch fibers (type IIx and IIb) with less or no impact on type I fibers [38, 39]. Therefore, fast-twitch glycolytic muscles (i.e., tibialis anterior, quadricpes femoris) are more susceptible than oxidative muscles to glucocorticoid-induced muscle atrophy. Hence the positive association between quadriceps femoris and tibialis anterior. Of surprise may be the lack of association between hand grip and pinch grip. A possible explanation may be the fact that hand grip requires synergistic function of intrinsic and extrinsic muscles of the hand. On the contrary, lateral pinch grip rely more on the thenar muscles. The latter are considered more oxidative muscles than the former [40]. Moreover, the distribution of muscle fibers between the antebrachium (i.e., the extrinsic muscles of the hand), and leg-thigh muscles is quite similar [40] which could explain the positive association between quadriceps femoris and hand grip.

5. Conclusions

Children receiving treatment for ALL exhibit fluctuating trends of muscle strength and functional capacity. These trends are more severely negative in the first two months of treatment. Moreover, the risk group, and consequently the treatment regime of the children determine muscle strength and functional mobility. These findings suggest that exercise training interventions initiated early on during the course of treatment for ALL need to be tested, and the risk group and the phase of the treatment should be taken into consideration. Further, the exercise interventions should include muscle strengthening and functional mobility components to address presenting skeletal and motor impairments. Additionaly, the measurements of a few muscle groups (quadriceps femoris, tibialis anterior, hand grip) can be used as surrogate to the measurement of a desired but unobtainable muscle group.

Part II

Broadly, the structure of this part is as follows: some notation is introduced (Section 1) and the design of the study is briefly revisited (Section 2). The modelling approach of the functional mobility (Section 3) and muscle strength (Section 4) are presented. Section 5 addresses the issue of missing data.

1. Notation

First we introduce some notation we are going to use throughout the document. Let Y_{ij} denote the response variable for the i^{th} individual ($i = 1,\ldots,42$) measured at time t_j, ($j = 1,\ldots,n_i$). The subscript i in n_i allows for some subjects not to be measured at all $n = 8$ occasions (i.e n_i denotes the number of observed responses on the i^{th} subject, where $n_i \leq n$.). Let also $\boldsymbol{Y}_i$ be the n_i-dimensional vector of all repeated measurements for the i^{th} subject, that is, $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{in_i})$. Associated with each response, $\boldsymbol{Y}_i$, there is a $n_i \times p$ matrix of covariates, X_i.

We follow the convention to denote a random variable by an upper-case letter (e.g. Y_{ij} is the response variable for the i^{th} individual at the j^{th} occasion) and the realised value of a random variable by the corresponding lower-case letter (e.g. y_{ij} denotes the realised value of Y_{ij}).

2. Design

Briefly, the research project is an ongoing observational study, which started in 2013. Participants are children diagnosed with ALL. The 42 children were followed from diagnosis until approximately 2 years into their treatment. By design measurements were taken at 8 unequally spaced but fixed timepoints. Due to missing visits, the number of measurements per subject varies from 3 to 8. The primary outcomes are the muscle strength (kilos) and functional mobility (meters) during the course of their treatment. Muscle strength is measured at four different muscle sites: knee extension, ankle dorsiflexion, hand grip and pinch grip. Functional mobility was assessed through 6MWD. Age (years), risk group (VLR, VHR, AR1, AR2-B and AR2-T), height (centimetres), weight (kilos), and gender (female, male) are used as covariates. For more details see Materials and Methods (Section 2 Part I).

2.1 Research questions

The main research questions and their motivation are presented in the Introduction (Section 1 Part I). To fix ideas we repeat them here:

- What is the trend in functional mobility and muscle strength in the two-year treatment period for all ALL children?
- What are the differences, if any, between different risk group subjects with respect to muscle strength and functional mobility during the last year of their treatment?
- What is the association between muscle performance of different muscle groups?

3. Univariate Modelling: Functional Mobility

We start by modelling the functional mobility, which is assessed by the 6MWD. The distance covered (meters) is our response variable. From organisational viewpoint we divide the model-fitting process into four stages [41]:

1. formulation - choosing the general form of the model;
2. estimation - attaching numerical values to parameters;
3. inference - calculating confidence intervals and testing hypotheses about parameters of direct interest;
4. diagnostics - checking if the model fits the data.

3.1 Formulation

Formulation of a model is essentially a continuation of the exploratory data analysis, but directed towards the specific aspects of the data which our model aims to describe. We mainly use graphics, to shed light on modelling decisions. The focus of attention is the mean and covariance structure of the data.

3.2 Mean structure

The individual profiles and the mean profile with standard errors (SE), are displayed in Figure 4. This graph makes apparent a number of important patterns. First, there is a considerable drop of the performance after the first timepoint. Second, the trajectory after the first timepoint is generally non-linear, increasing, with a lot of fluctuations. Third, it is clear that not all children are observed at every timepoint which may explain the wider SE values towards the end of the study.

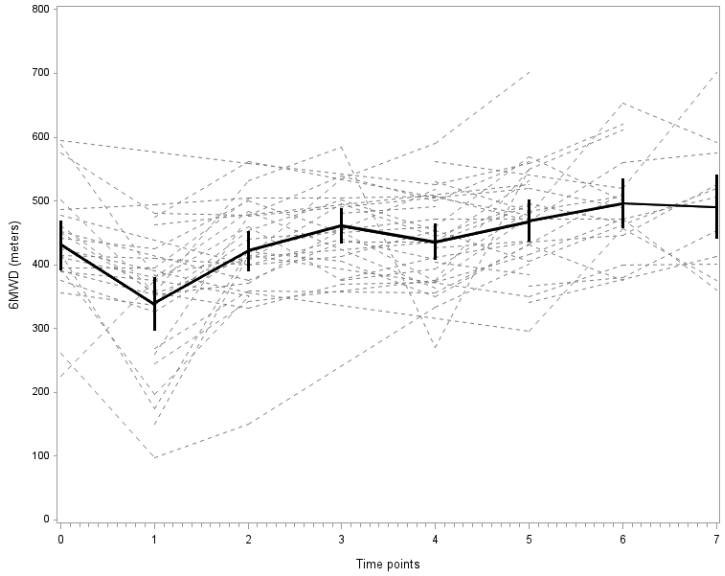


Figure 4: Individual profiles (dashed lines), Mean (SE) profile (solid line).

3.3 Covariance structure

The same plot as before but now with standard deviations (SD) per time-point is given in Figure 5. Sample SD is an indicator of within-time across-subject variability. We see that the SD remains approximately constant across timepoints.

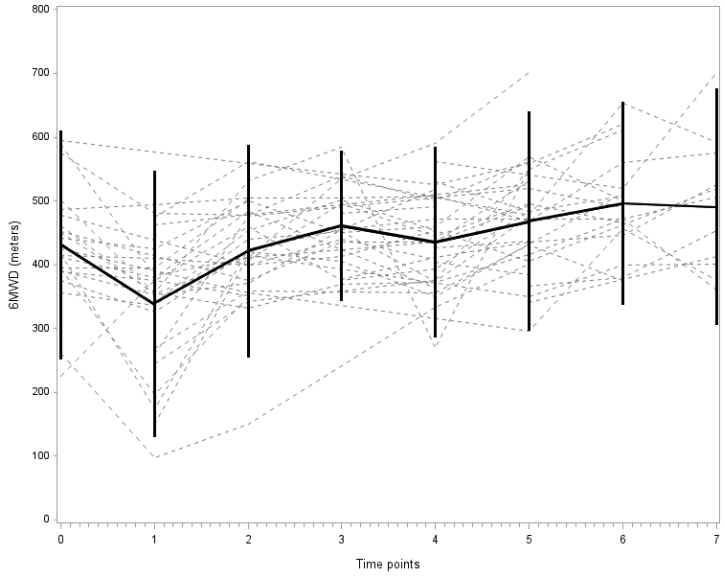


Figure 5: Individual profiles (dashed lines), Mean (SD) profile (solid line).

In order to have a more in depth picture for the covariance structure, we remove the effects of explanatory variables. We use residuals obtained by subtracting from each measurement the ordinary least squares (OLS) estimate of the corresponding mean response. Hence, we first regress the response y_{ij} on the explanatory variables x_{ij} to obtain residuals, $r_{ij} = y_{ij} - x'_{ij}\hat{\beta}$.

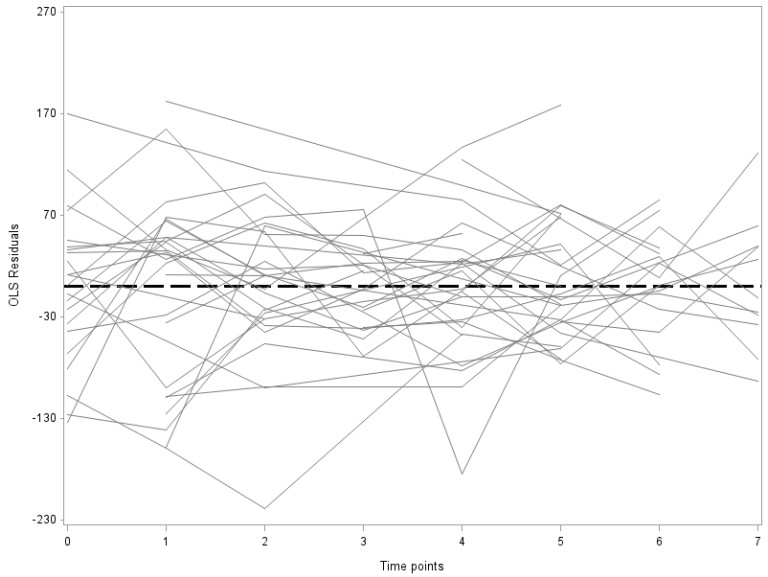


Figure 6: Ordinary least squares (OLS) residual profiles.

There seems to be no systematic structure remaining in the residual profiles (Figure 6). The residuals seem to exhibit constant variability.

Given that the variance function is assumed to be constant, a function that describes the association among repeated values is the variogram. For a stochastic process $Y(t)$, the variogram is defined as

$$\gamma(u) = \frac{1}{2}E\Big[[Y(t) - Y(t-u)]^2\Big], u \geq 0.$$

If $Y(t)$ is stationary, the variogram is directly related to the serial correlation function, $\rho(u)$, by

$$\gamma(u) = \sigma^2[1 - \rho(u)],$$

where σ^2 is the variance of $Y(t)$. In our context, the empirical counterpart of the variogram, which we call the sample variogram, can be calculated from observed half-squared-differences between pairs of residuals,

$$\nu_{ijk} = \frac{1}{2}\big[(r_{ij} - r_{ik})^2\big],$$

and the corresponding time-differences,

$$u_{ijk} = t_{ij} - t_{ik},$$

for all $i = 1, ..., 42$ and for all $j \neq k$. By letting $\hat{\gamma}(u)$ be the average of all of the ν_{ijk} corresponding to that particular value of u we can plot in a scatter plot of all half squared differences ν_{ijk} versus the corresponding time differences $u_{ijk} = |t_{ij} - t_{ik}|$. The variance, σ^2, is estimated as the average of all half-squared-differences $\frac{1}{2}(y_{ij} - y_{lk})$, with $i \neq l$. The estimate of $\gamma(u)$ can now be used for deciding whether or not our model should include serial correlation, and if so, for selecting an appropriate function $\rho(u)$, since

$$\hat{\rho}(u) = \frac{1 - \hat{\gamma}(u)}{\hat{\sigma}^2}.$$

An estimate of the variogram for the data is pictured in Figure 7. The diagram shows both the basic quantities (ν_{ijk}, u_{ijk}) and a smooth estimate of $\gamma(u)$ which has been produced using lowess. Also, to accentuate the shape of the smooth estimate, we have truncated the vertical axis at 100000. The horizontal line is the variogram-based estimate of the variance (σ^2). The empirical variogram exhibits a smooth decrease with increasing lag, implying increasing correlation as observations are separated in time. This is not what one would expect since the correlation usually decreases for observations further apart. Moreover, since the trend is not very steep and because it is opposite to what one would expect we assume constant correlation, implying compound symmetry as covariance structure. We further try to extend this assumption by fitting a heterogeneous compound symmetry structure. This is a simple extension where the variances along the diagonal of the matrix do not have to be the same. The heterogeneous compound symmetry increases the loglikelihood from -2005.0 to -1986.2 . A likelihood-ratio (LR) test ($\chi^2 = 18.8; df = 7$) rejects the constant variance assumption (p-value=0.009). This is in contrast to the earlier findings from the variogram. This is because the variogram is based on very strict assumptions about the variance function, which may not be realistic. Namely, it is applicable when the variance function can be assumed to be constant.

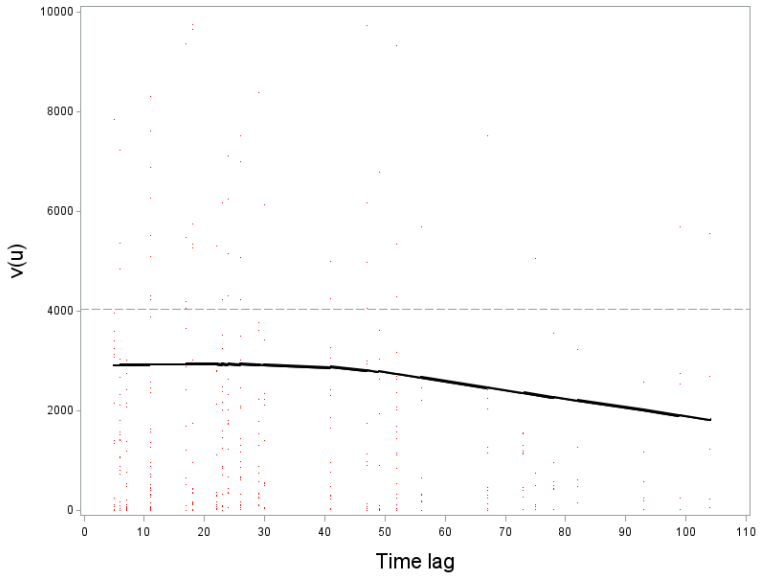


Figure 7: Sample variogram of residuals. Horizontal line estimates process variance.

3.4 Estimation

Our objective is to estimate the parameters in the model. We use a marginal model in which the regression of the response on explanatory variables is modelled separately from within-person correlation. In the regression we model the marginal expectation, $E(Y_{ij})$, as a function of the explanatory variables. The general form is

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + \epsilon_i, \tag{1}$$

with $\boldsymbol{Y}_i$ the vector of responses of the 6MWT, X_i the matrix of covariates, $\boldsymbol{\beta}$ the vector of regression parameters and ϵ_i the vector of error components, $\epsilon_i \sim N(0, \Sigma_i)$ with Σ_i a $n_i \times n_i$ matrix. The X matrix contains time (modelled as categorical variable), age, risk group, height, gender, and their interactions with time. This is our initial model.

Assuming heterogeneous compound symmetry as covariance structure, Σ_i has elements

$$\sigma_i\sigma_j[\rho I(i \neq j) + I(i = j)],$$

where I is an indicator variable that equals 1 when the argument inside the

parenthesis holds and 0 otherwise. We favour restricted maximum likelihood (REML) estimates over classical maximum likelihood (ML) estimates to reduce the bias in the estimation of the variance components. For more details and justification see Verbeke and Molenberghs, 2009 [42].

3.5 Inference

We use LR test and approximate F-Tests to test hypotheses about $\boldsymbol{\beta}$. The LR test goes as follows,

$$-2ln(\lambda) = -2ln\left[\frac{L_{ML}(\hat{\boldsymbol{\beta}}_{ML,0})}{L_{ML}(\hat{\boldsymbol{\beta}}_{ML})}\right],$$

where $\hat{\boldsymbol{\beta}}_{ML,0}$ and $\hat{\boldsymbol{\beta}}_{ML}$ are the ML estimates obtained from maximizing L_{ML} over the null hypothesis (H_0) and the alternative (H_A), respectively. Fitting our initial model (1) without age and gender increases the loglikelihood from -1986.2 to -2011.9. The LR ($\chi^2 = 25.7; df = 24$) does not reject the simpler model (p-value=0.369). It should be emphasised that the result is obtained after models are fitted using ML estimation since it is not valid under REML estimation.

Here we present the parametrisation of the final model:

$$Y_{i0} = \beta_{1,0} + \beta_{2,0}height + \epsilon_{i0}$$

$$\vdots$$

$$Y_{i4} = \beta_{1,4} + \beta_{2,4}height + \epsilon_{i4}$$

$$Y_{i5} = \beta_{1,5}AR1+\beta_{2,5}AR2T+\beta_{3,5}AR2B+\beta_{4,5}VLR+\beta_{5,5}VHR+\beta_{6,5}height+\epsilon_{i5}$$

$$\vdots$$

$$Y_{i7} = \beta_{1,7}AR1+\beta_{2,7}AR2T+\beta_{3,7}AR2B+\beta_{4,7}VLR+\beta_{5,7}VHR+\beta_{6,7}height+\epsilon_{i7}$$

The subscripts for all parameters consist of two numbers. The first number refers to the order in which the parameter appears in the model. The second number refers to the timepoint ($0 = T_0, 1 = T_1$, etc). For the first 5 timepoints there is a common risk group effect ($\beta_{1,0} - \beta_{1,4}$) since all subjects are assigned to the same risk group.

We opt to answer the first research question visually i.e., using graphs with the predicted values from the model controlling for covariates (see Results (Section 3 Part I): Fig. 2). The second research question can be divided into two parts. Do AR1 and AR2-B subjects perform differently than AR2-T and VLR subjects at T_5? And, do VHR subjects perform differently than the other risk groups at T_5, T_6, T_7? For motivation of the questions we refer to the Introduction (Section 1 Part I). The two questions correspond to the following hypotheses:

$$H_0 : \begin{cases} \beta_{1,5} = \beta_{5,5} \\ \beta_{2,5} = \beta_{5,5} \\ \beta_{3,5} = \beta_{5,5} \\ \beta_{4,5} = \beta_{5,5} \end{cases} \quad (2)$$

$$H_0' : \begin{cases} \beta_{1,5} + \beta_{3,5} = \beta_{4,5} + \beta_{2,5} \\ \beta_{1,6} + \beta_{3,6} = \beta_{4,6} + \beta_{2,6} \\ \beta_{1,7} + \beta_{3,7} = \beta_{4,7} + \beta_{2,7} \end{cases} \quad (3)$$

Equations 2 and 3 are of the form

$$H_0 : L\boldsymbol{\beta} = 0, \text{versus } H_A : L\boldsymbol{\beta} \neq 0.$$

Testing these hypotheses is based on an F-approximation to the distribution of

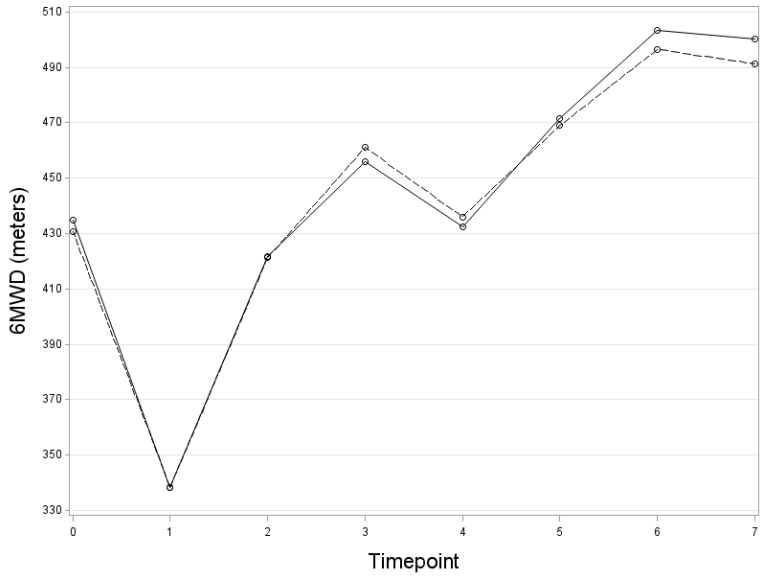
$$F = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' L' \Big[L\big(\sum_{i=1}^{N} X_i' \Sigma_i^{-1} X_i \big)^{-1} L' \Big]^{-1} L(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{rank(L)}.$$

The numerator degrees of freedom equals $rank(L)$. The denominator degrees of freedom are calculated based on the Kenward and Roger approximation [43]. For the results we refer again to the Results (Section 3 Part I).

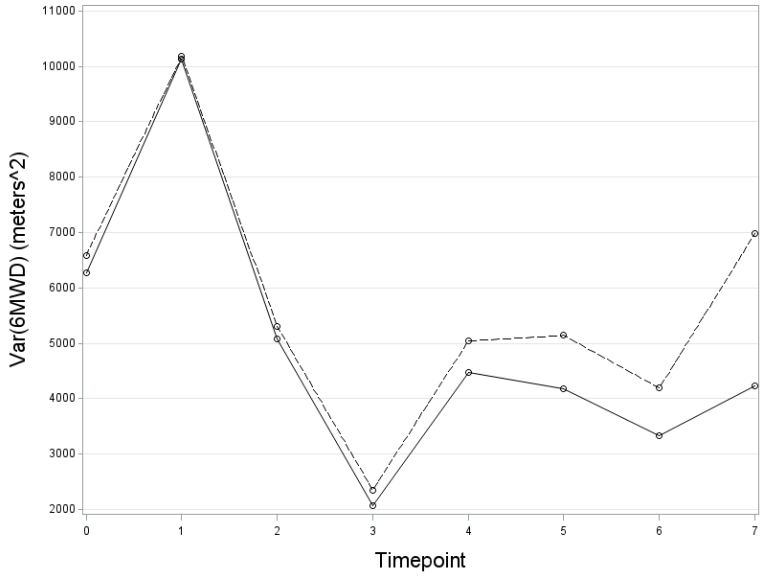
3.6 Diagnostics

The model fitting process is concluded by diagnostic checking of the model against the data process. The goal is to highlight any systematic discrepancies between the fitted model and the data. Since we are essentially modelling the mean and covariance structure of the data, we choose to su-

perimpose the fitted mean response profiles on a time-plot of the average observed response (Figure 8(a)) and to superimpose the fitted variance function on a plot with the observed one (Figure 8(b)). At first sight, the fit to the mean response profiles and the variance function appear quite satisfactory. The apparent lack of fit at a few timepoints towards the end of the study can be attributed to the missing observations. Overall, we could argue that the model describes adequately the observed data. To be noted, the observed variance function has been calculated using OLS residuals, obtained by fitting the final model, and by assuming independence of all observations.



(a)



(b)

Figure 8: Comparison between observed (solid lines) and fitted (dashed lines) mean and variance response profiles. (a) mean response profiles; (b) variance response profiles.

4. Joint Modelling: Muscle Strength

In order to avoid superfluous repetition the joint model is presented more briefly. The steps of model formulation, estimation, inference and diagnostics follow the same principles as already described. Here, we focus on important deviations from the above presented process.

In the context of jointly modelling let $Y_{ij}^{KE}, Y_{ij}^{AD}, Y_{ij}^{HG}, Y_{ij}^{PG}$ denote the knee extension, ankle dorsiflexion, hand grip and pinch grip strength for subject i at time j. Each response is described using the linear mixed-effects model

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \epsilon_i$$

where X_i and Z_i are matrices of the covariates, $\boldsymbol{\beta}$ is the vector of fixed effects, $\boldsymbol{b}_i$ are the random effects, and the ϵ_i represent the usual error components. In total there are four random effects (intercepts) which jointly follow a zero-mean normal distribution with unstructured covariance matrix which we denote by D. Further, the error components are considered correlated and follow a four-dimensional zero-mean normal distribution with unstructured covariance matrix Σ. They practically represent the fact that the $(Y_{ij}^{KE}, Y_{ij}^{AD}, Y_{ij}^{HG}, Y_{ij}^{PG})$ measures are not perfect measures of the muscle strength, but rather, in some sense, an approximation to the true quantity. Since the $(Y_{ij}^{KE}, Y_{ij}^{AD}, Y_{ij}^{HG}, Y_{ij}^{PG})$ are measures on the same individual conducted at the same timepoints, they are correlated conditional on X_i. The matrix X contains time, risk group, weight, age, and their interactions with time. These are the only significant terms after backward elimination as described in Section 3.4. We also test for uncorrelated error components, which decreases the loglikelihood from -2570.2 to -2637.3. The LR ($\chi^2 = 67.1; df = 6$) rejects the uncorrelated error components assumption (p-value<0.001).

To fix ideas we present the linear random-effects model for ankle dorsiflexion:

$$Y_{i0}^{AD} = \beta_{1,0}^{AD} + \beta_{2,0}^{AD} weight + \beta_{3,0}^{AD} age + b_i^{AD} + \epsilon_{i0}^{AD}$$

$$\vdots$$

$$Y_{i4}^{AD} = \beta_{1,4}^{AD} + \beta_{2,4}^{AD} weight + \beta_{3,4}^{AD} age + b_i^{AD} + \epsilon_{i0}^{AD}$$

$$Y_{i5}^{AD} = \beta_{1,5}^{AD} AR1 + \beta_{2,5}^{AD} AR2T + \beta_{3,5}^{AD} AR2B + \beta_{4,5}^{AD} VLR + \beta_{5,5}^{AD} VHR + \beta_{6,5}^{AD} weight + \beta_{7,5}^{AD} age + b_i^{AD} + \epsilon_{i5}^{AD}$$

$$\vdots$$

$$Y_{i7}^{AD} = \beta_{1,7}^{AD} AR1 + \beta_{2,7}^{AD} AR2T + \beta_{3,7}^{AD} AR2B + \beta_{4,7}^{AD} VLR + \beta_{5,7}^{AD} VHR + \beta_{6,7}^{AD} weight + \beta_{7,7}^{AD} age + b_i^{AD} + \epsilon_{i7}^{AD},$$

where b_i^{AD} represents the random intercept. Similar equations for the other responses can be obtained. Hypotheses similar to equations 2 and 3 are tested for all outcomes simultaneously. For the results see Results (Section 3 Part I).

The answer to the question "what is the association between muscle performance of different muscle groups" typically can be derived from the covariance matrix of the random effects. Let Y^k and Y^l $(k \neq l)$ denote any 2 of the 4 outcomes. Also assume they are measured at the same timepoint $(t = t')$. Then their correlation is given by

$$Corr(Y_i^k(t), Y_i^l(t)) = \frac{Cov(b_i^k, b_i^l) + Cov(\epsilon_i^k, \epsilon_i^l)}{\sqrt{Var(b_i^k) + Var(\epsilon_i^k)}\sqrt{Var(b_i^l) + Var(\epsilon_i^l)}} \quad (4)$$

For any two outcomes measured at different timepoints $(t \neq t')$ $Cov(\epsilon_i^k, \epsilon_i^l)$ does not contribute to the calculation. Hence equation 4 becomes

$$Corr(Y_i^k(t), Y_i^l(t)) = \frac{Cov(b_i^k, b_i^l)}{\sqrt{Var(b_i^k) + Var(\epsilon_i^k)}\sqrt{Var(b_i^l) + Var(\epsilon_i^l)}} \quad (5)$$

Using the delta rule [44] confidence bounds are obtained for both 4 and 5. The results are presented in Tables 2 and 3 (Results, Section 3 Part I).

4.1 Diagnostics

Diagnostics similar to those in Section 3.6 have been performed. Furthermore, to verify to what extent the model implies a realistic association structure we compare the marginal correlations implied by the model with the observed marginal correlations. Marginal denotes the fact the random ef-

fects have been integrated out. Since the time intervals between successive measurements are the same between subjects, the observed marginal correlation is calculated as the association at each time point. Table 4 shows the so-obtained observed marginal correlations compared with the marginal correlations implied by the random effects model. We assume the measurements have been taken at the same timepoint. For measurements taken at different timepoints Table 5 presents the average implied and observed correlations over time lags (i.e, lag 1 to 7). It can be clearly seen that all of the observed marginal correlations lie inside the confidence bands of the implied ones. Hence, we can say that the model is implying a realistic association structure. Note that the observed correlations have been calculated using OLS residuals, obtained by fitting the joint model, but by assuming independence of all observations. These residuals can be used since it follows from the theory of generalized estimating equations that the OLS estimator is consistent [45].

Table 4: Observed and implied correlation (r) for the same timepoint with 95% CI

	Observed r	Implied r	[95% CI]
AD,KE	0.61	0.56	[0.29, 0.75]
KE,HG	0.34	0.39	[0.18, 0.57]
KE,PG	0.18	0.30	[-0.20, 0.68]
AD,HG	0.08	0.05	[-0.07, 0.16]
AD,PG	0.24	0.30	[-0.20, 0.67]
HG,PG	0.46	0.43	[-0.25, 0.83]

Abbreviations: ankle dorsiflexion (AD); knee extension (KE); hand grip (HG); pinch grip (PG)

Table 5: Observed and implied correlation (r) for different timepoints with 95% CI

	Observed r	Implied r	[95% CI]
AD,KE	0.42	0.29	[0.10, 0.46]
KE,HG	0.20	0.24	[0.06, 0.40]
KE,PG	0.20	0.22	[-0.24, 0.60]
AD,HG	0.04	-0.08	[-0.20, 0.05]
AD,PG	0.22	0.21	[-0.24, 0.60]
HG,PG	0.30	0.27	[-0.29, 0.68]

Abbreviations: ankle dorsiflexion (AD); knee extension (KE); hand grip (HG); pinch grip (PG)

5. Missing data

Missing values arose during the data collection process. They are defined as intended measurements which could not be obtained. When there are missing data, the validity of any method of analysis requires that certain assumptions about the missing data mechanism, be tenable. The key issue is whether the reasons for missingness are related to the outcome of interest [46]. Our analysis was based on the missing at random (MAR) mechanism. This assumes that the probability that responses are missing depends on the set of observed responses, but is further unrelated to the specific missing values. Under MAR a likelihood-based analyses yields valid results provided we have correctly specified the model for $f(\boldsymbol{Y}_i|X_i)$. For more details see Molenberghs et al, 2014 [46].

Since the MAR assumption cannot be verified from the data at hand, the standard approach is to assess robustness of inferences to departures from assumptions via sensitivity analyses. Under the previous definition of missing data Table 6 shows that the discrepancy between the observed and intended measurements is relatively small, ranging from 0 to 9 per timepoint. The investigation of the impact of missing data on the final results is beyond the scope of this paper.

Table 6: Observed and missing values per response per timepoint

Timepoint	Observed	Missing
	KE AD HG PG 6MWD	KE AD HG PG 6MWD
T_0	25 - 25 - 21 - 22 - 22	0 - 0 - 2 - 3 - 3
T_1	28 - 28 - 22 - 25 - 25	2 - 2 - 6 - 4 - 1
T_2	29 - 29 - 26 - 28 - 29	1 - 1 - 4 - 2 - 1
T_3	19 - 19 - 18 - 16 - 19	6 - 6 - 7 - 9 - 6
T_4	27 - 27 - 26 - 27 - 27	0 - 0 - 1 - 0 - 0
T_5	26 - 26 - 25 - 26 - 26	2 - 2 - 3 - 2 - 2
T_6	18 - 18 - 18 - 18 - 17	1 - 1 - 1 - 1 - 2
T_7	13 - 13 - 13 - 13 - 13	1 - 1 - 1 - 1 - 1

Abbreviations: ankle dorsiflexion (AD); knee extension (KE); hand grip (HG); pinch grip (PG); distance walked (6MWD)

Bibliography

[1] K. I. Braam, P. van der Torre, T. Takken, M. A. Veening, E. van Dulmen-den Broeder, and G. J. Kaspers, "Physical exercise training interventions for children and young adults during and after treatment for childhood cancer," *The Cochrane Library*, 2016.

[2] C. A. Stiller, "Epidemiology and genetics of childhood cancer," *Oncogene*, vol. 23, no. 38, pp. 6429–6444, 2004.

[3] V. Conter, M. Arico, G. Basso, A. Biondi, E. Barisone, C. Messina, R. Parasole, G. De Rossi, F. Locatelli, A. Pession, *et al.*, "Long-term results of the italian association of pediatric hematology and oncology (aieop) studies 82, 87, 88, 91 and 95 for childhood acute lymphoblastic leukemia," *Leukemia*, vol. 24, no. 2, pp. 255–264, 2010.

[4] A. Möricke, M. Zimmermann, A. Reiter, G. Henze, A. Schrauder, H. Gadner, W. Ludwig, J. Ritter, J. Harbott, G. Mann, *et al.*, "Long-term results of five consecutive trials in childhood acute lymphoblastic leukemia performed by the all-bfm study group from 1981 to 2000," *Leukemia*, vol. 24, no. 2, pp. 265–284, 2010.

[5] S. P. Hunger, X. Lu, M. Devidas, B. M. Camitta, P. S. Gaynon, N. J. Winick, G. H. Reaman, and W. L. Carroll, "Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group," *Journal of Clinical Oncology*, pp. JCO–2011, 2012.

[6] A. J. Veerman, W. A. Kamps, H. van den Berg, E. van den Berg, J. P. Bökkerink, M. C. Bruin, M. M. van den Heuvel-Eibrink, C. M. Korbijn, E. T. Korthof, K. van der Pal, *et al.*, "Dexamethasone-based therapy for childhood acute lymphoblastic leukaemia: results of the prospective

dutch childhood oncology group (dcog) protocol all-9 (1997–2004)," *The lancet oncology*, vol. 10, no. 10, pp. 957–966, 2009.

[7] K. Schmiegelow, E. Forestier, M. Hellebostad, M. Heyman, J. Kristinsson, S. Söderhäll, and M. Taskinen, "Long-term results of nopho all-92 and all-2000 studies of childhood acute lymphoblastic leukemia," *Leukemia*, vol. 24, no. 2, pp. 345–354, 2010.

[8] C. Mitchell, S. Richards, C. J. Harrison, and T. Eden, "Long-term follow-up of the united kingdom medical research council protocols for childhood acute lymphoblastic leukaemia, 1980–2001," *Leukemia*, vol. 24, no. 2, pp. 406–418, 2010.

[9] A. Kolenova, E. Kaiserova, M. Makohusova, Z. Subova, E. Bubanska, I. Oravkinova, P. Svec, O. Fabri, S. Hunger, S. Sufliarska, *et al.*, "Improved outcome for children and adolescent with acute lymphoblastic leukemia in the first decade of the 21st century: A report from the slovak republic.," *Neoplasma*, vol. 62, no. 5, pp. 812–820, 2014.

[10] M. A. Smith, N. L. Seibel, S. F. Altekruse, L. A. Ries, D. L. Melbert, M. O'Leary, F. O. Smith, and G. H. Reaman, "Outcomes for children and adolescents with cancer: challenges for the twenty-first century," *Journal of Clinical Oncology*, vol. 28, no. 15, pp. 2625–2634, 2010.

[11] C. Domenech, S. Suciu, B. De Moerloose, F. Mazingue, G. Plat, A. Ferster, A. Uyttebroeck, N. Sirvent, P. Lutz, K. Yakouben, *et al.*, "Dexamethasone (6 mg/m2/day) and prednisolone (60 mg/m2/day) were equally effective as induction therapy for childhood acute lymphoblastic leukemia in the eortc clg 58951 randomized trial," *haematologica*, vol. 99, no. 7, pp. 1220–1227, 2014.

[12] A. Hartman, W. Hop, T. Takken, R. Pieters, and M. van den Heuvel-Eibrink, "Motor performance and functional exercise capacity in survivors of pediatric acute lymphoblastic leukemia," *Pediatric blood & cancer*, vol. 60, no. 3, pp. 494–499, 2013.

[13] S. Y. Tan, B. K. Poh, H. X. Chong, M. N. Ismail, J. Rahman, A. L. Zarina, A. R. E. Juraida, A. Tahir, A. T. Ruzita, R. Roslee, *et al.*, "Physical activity of pediatric patients with acute leukemia undergoing

induction or consolidation chemotherapy," *Leukemia research*, vol. 37, no. 1, pp. 14–20, 2013.

[14] K. K. Ness, S. H. Armenian, N. Kadan-Lottick, and J. G. Gurney, "Adverse effects of treatment in childhood acute lymphoblastic leukemia: general overview and implications for long-term cardiac health," *Expert review of hematology*, vol. 4, no. 2, pp. 185–197, 2011.

[15] K. K. Ness, K. S. Baker, D. R. Dengel, N. Youngren, S. Sibley, A. C. Mertens, and J. G. Gurney, "Body composition, muscle strength deficits and mobility limitations in adult survivors of childhood acute lymphoblastic leukemia," *Pediatric blood & cancer*, vol. 49, no. 7, pp. 975–981, 2007.

[16] V. Gocha Marchese, L. A. Chiarello, and B. J. Lange, "Strength and functional mobility in children with acute lymphoblastic leukemia," *Medical and pediatric oncology*, vol. 40, no. 4, pp. 230–232, 2003.

[17] B. De Moerloose, S. Suciu, Y. Bertrand, F. Mazingue, A. Robert, A. Uyttebroeck, K. Yakouben, A. Ferster, G. Margueritte, P. Lutz, *et al.*, "Improved outcome with pulses of vincristine and corticosteroids in continuation therapy of children with average risk acute lymphoblastic leukemia (all) and lymphoblastic non-hodgkin lymphoma (nhl): report of the eortc randomized phase iii trial 58951," *Blood*, pp. 36–44, 2010.

[18] C.-H. Pui, C. G. Mullighan, W. E. Evans, and M. V. Relling, "Pediatric acute lymphoblastic leukemia: where are we going and how do we get there?," *Blood*, vol. 120, no. 6, pp. 1165–1174, 2012.

[19] A. Uyttebroeck, S. Suciu, G. Laureys, A. Robert, H. Pacquement, A. Ferster, G. Marguerite, F. Mazingue, M. Renard, P. Lutz, *et al.*, "Treatment of childhood t-cell lymphoblastic lymphoma according to the strategy for acute lymphoblastic leukaemia, without radiotherapy: long term results of the eortc clg 58881 trial," *European journal of cancer*, vol. 44, no. 6, pp. 840–846, 2008.

[20] R. W. Bohannon, "The clinical measurement of strength," *Clinical Rehabilitation*, vol. 1, no. 1, pp. 5–16, 1987.

[21] E. D. Joyce, V. G. Nolan, K. K. Ness, R. J. Ferry, L. L. Robison, C.-H. Pui, M. M. Hudson, and S. C. Kaste, "Association of muscle strength and bone mineral density in adult survivors of childhood acute lymphoblastic leukemia," *Archives of physical medicine and rehabilitation*, vol. 92, no. 6, pp. 873–879, 2011.

[22] W. A. van den Beld, W. A. van den Beld, G. A. van der Sanden, R. C. Sengers, A. L. Verbeek, and F. J. Gabreëls, "Validity and reproducibility of the jamar dynamometer in children aged 4–11 years," *Disability and rehabilitation*, vol. 28, no. 21, pp. 1303–1309, 2006.

[23] V. Mathiowetz, D. M. Wiemer, and S. M. Federman, "Grip and pinch strength: norms for 6-to 19-year-olds," *American Journal of Occupational Therapy*, vol. 40, no. 10, pp. 705–711, 1986.

[24] L. Merlini, E. S. Mazzone, A. Solari, and L. Morandi, "Reliability of hand-held dynamometry in spinal muscular atrophy," *Muscle & nerve*, vol. 26, no. 1, pp. 64–70, 2002.

[25] C. Häger-Ross and B. Rösblad, "Norms for grip strength in children aged 4–16 years," *Acta Paediatrica*, vol. 91, no. 6, pp. 617–625, 2002.

[26] M. C. Hooke, A. W. Garwick, and J. P. Neglia, "Assessment of physical performance using the 6-minute walk test in children receiving treatment for cancer," *Cancer nursing*, vol. 36, no. 5, pp. E9–E16, 2013.

[27] D. B. Rubin, "Inference and missing data," *Biometrika*, vol. 63, no. 3, pp. 581–592, 1976.

[28] A. F. San Juan, C. Chamorro-Vina, J.-L. Maté-Muñoz, M. F. del Valle, C. Cardona, M. Hernández, L. Madero, M. Pérez, M. Ramírez, and A. Lucia, "Functional capacity of children with leukemia," *International journal of sports medicine*, vol. 29, no. 02, pp. 163–167, 2008.

[29] T.-T. Huang and K. K. Ness, "Exercise interventions in children with cancer: a review," *International journal of pediatrics*, vol. 2011, pp. 1–11, 2011.

[30] A. Moore and R. Pinkerton, "Vincristine: can its therapeutic index be enhanced?," *Pediatric blood & cancer*, vol. 53, no. 7, pp. 1180–1187, 2009.

[31] D. Balayssac, J. Ferrier, J. Descoeur, B. Ling, D. Pezet, A. Eschalier, and N. Authier, "Chemotherapy-induced peripheral neuropathies: from clinical relevance to preclinical evidence," *Expert opinion on drug safety*, vol. 10, no. 3, pp. 407–417, 2011.

[32] J. R. Ryan and A. Emami, "Vincristine neurotoxicity with residual equinocavus deformity in children with acute leukemia," *Cancer*, vol. 51, no. 3, pp. 423–425, 1983.

[33] K. K. Ness, M. M. Hudson, C.-H. Pui, D. M. Green, K. R. Krull, T. T. Huang, L. L. Robison, and E. B. Morris, "Neuromuscular impairments in adult survivors of childhood acute lymphoblastic leukemia," *Cancer*, vol. 118, no. 3, pp. 828–838, 2012.

[34] L. Hovi, P. Era, J. Rautonen, and M. A. Siimes, "Impaired muscle strength in female adolescents and young adults surviving leukemia in childhood," *CANCER-PHILADELPHIA-*, vol. 72, pp. 276–276, 1993.

[35] H. Lee, B. Oran, R. Saliba, D. Couriel, K. Shin, P. Massey, J. Neumann, M. De Lima, R. Champlin, and S. Giralt, "Steroid myopathy in patients with acute graft-versus-host disease treated with high-dose steroid therapy," *Bone marrow transplantation*, vol. 38, no. 4, pp. 299–303, 2006.

[36] K. H. Kroemer, "Assessment of human muscle strength for engineering purposes: a review of the basics," *Ergonomics*, vol. 42, no. 1, pp. 74–93, 1999.

[37] E. Finch, D. Brooks, P. W. Stratford, P. W. Statford, and N. E. Mayo, *Physical rehabilitation outcome measures: a guide to enhanced clinical decision making*. BC Decker Hamilton, 2002.

[38] O. Schakman, H. Gilson, and J.-P. Thissen, "Mechanisms of glucocorticoid-induced myopathy," *Journal of Endocrinology*, vol. 197, no. 1, pp. 1–10, 2008.

[39] O. Schakman, S. Kalista, C. Barbé, A. Loumaye, and J.-P. Thissen, "Glucocorticoid-induced skeletal muscle atrophy," *The international journal of biochemistry & cell biology*, vol. 45, no. 10, pp. 2163–2172, 2013.

[40] T. F. Tirrell, M. S. Cook, J. A. Carr, E. Lin, S. R. Ward, and R. L. Lieber, "Human skeletal muscle biochemical diversity," *Journal of Experimental Biology*, vol. 215, no. 15, pp. 2551–2559, 2012.

[41] P. Diggle, P. Heagerty, K.-Y. Liang, and S. Zeger, *Analysis of longitudinal data*. OUP Oxford, 2013.

[42] G. Verbeke and G. Molenberghs, *Linear mixed models for longitudinal data*. Springer Science & Business Media, 2009.

[43] M. G. Kenward and J. H. Roger, "Small sample inference for fixed effects from restricted maximum likelihood," *Biometrics*, pp. 983–997, 1997.

[44] R. J. Serfling, *Approximation theorems of mathematical statistics*, vol. 162. John Wiley & Sons, 2009.

[45] K.-Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.

[46] G. Molenberghs, G. Fitzmaurice, M. G. Kenward, A. Tsiatis, and G. Verbeke, *Handbook of Missing Data Methodology*. CRC Press, 2014.

AFDELING
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be