University of Applied Sciences
Department Life Science
Master of Health Sciences

Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Master Thesis

"Psychometric Performance of the QoLISSY Questionnaire in a Randomized Clinical Trial for Growth Hormone Treatment in Short Stature Youth in the United States and Chile"

Date of submission: 26.03.2018
Submitted by: Richelle Valdez [redacted]
First Supervisor (HAW): Prof. Christine Färber
Second Supervisor (UKE): Janika Blömeke (MSc. Health Sciences)

Preface

The master thesis in your hands, titled "Psychometric Performance of the QoLISSY Questionnaire in a Randomized Clinical Trial for Growth Hormone Treatment in Short Stature Youth in the United States and Chile," was written as a capstone for my master's degree in health sciences at the Hamburg University of Applied Sciences (HAW). I started the process of writing this thesis in May 2017 and finished writing it in March 2018.

The topic of my master thesis was formulated in coordination with my supervisors at the University Medical Center Hamburg-Eppendorf (UKE), where I had done an internship, and my supervisor at the HAW. At that time, I wanted to have a thesis topic that would allow me to delve more deeply into the area of psychometric testing and statistical analyses — luckily, my supervisors at the UKE had suitable data from an interesting study done a few years ago in cooperation with a partner institution in the USA, and had graciously offered me the opportunity to write my thesis using this data.

Although writing this thesis was (at times) difficult, doing so has allowed me to learn more deeply about psychometric theory, testing, and evaluation. Luckily, I had the support from my supervisors, C. Färber and J. Blömeke to help guide me through this process.

For this reason, I would like to thank C. Färber and J. Blömeke for their excellent support, for being so patient with me, for always being willing to answer questions I had, for taking time to read draft material I would send, and for checking up on me every once in a while, to see how I was fairing during the writing process. I would also like to extend my gratitude to M. Bullinger and N. Mauras, without whose permission to use the data, I would not have been able to conduct my analyses.

To all my other colleagues at the UKE: thank you for also checking up on me and offering advice. There were many times when I struggled with writer's block, but your encouragement kept me motivated. I would also like to thank E. Morino for supporting me throughout this process, and for meticulously proofreading and giving thoughtful comments on my thesis. Lastly, I'd like to thank my parents for their undying support; without you, I would not have this wonderful opportunity to indulge my scientific interests.

I hope that you enjoy reading this thesis as much as I enjoyed writing it!

Richelle Valdez
Hamburg, Germany
March 24, 2018

Abstract

Background: Short stature in pediatric populations is associated with negative impacts on psychosocial well-being, higher occurrences of bullying, social isolation and stigmatization, and other negative health outcomes. To facilitate the improvement of the quality of life of this patient group, the Quality of Life in Short Statue Youth (QoLISSY) questionnaire was developed, which assesses the health-related quality of life (HrQoL) of short-statured children and adolescents via self- and proxy reports. This thesis aims to evaluate the psychometric performance of the QoLISSY instrument, using data from a previous study in which a randomized open label comparator trial was conducted that compared treatment of idiopathic short stature (ISS) in adolescent males with aromatase inhibitors (AI), growth hormones (GH), and a combination treatment of both (AI/GH).

Methods: In total, 76 boys diagnosed with ISS (12 to 18 years) and their parents were recruited. Patients were treatment naïve and randomized into a treatment type (AI, GH, or AI/GH). In addition to clinical variables, HrQoL was assessed using the QoLISSY and KIDSCREEN questionnaires before and after 12, 24, and 36 months of treatment. Descriptive statistics, content validity, construct validity, internal consistency, responsiveness, and parent-child score agreement were analyzed.

Results: The QoLISSY instrument shows good internal consistency, convergent validity, inter-scale correlations, and content validity. This is also true for most scales in terms of skewness, floor and ceiling effects, and divergent validity. The results of this thesis suggest that the QoLISSY instrument can detect significant changes of HrQoL between baseline and 24-months. The combination therapy (GH and AI) group reported higher HrQoL in all scales than the other two treatment types for both the child and parent report.

Conclusion: Results support that the QoLISSY is a psychometrically-sound instrument that can be used to track HrQoL changes over time, explore the experiences associated with short stature (and its treatment) through both the perspectives of the patients and their parents, and to highlight areas in life of short-statured children and adolescents that can be improved through intervention.

Keywords: health-related quality of life, idiopathic short stature, patient-reported outcomes, aromatase inhibitors, growth hormone, randomized open label comparator trial.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Aromatase inhibitors (treatment)
ANOVA	Analysis of variances
AUC	Receiver operating characteristics curve
GH	Growth hormone (treatment)
AI/GH	Combination therapy of aromatase inhibitors and growth hormone
GHD	Growth hormone deficiency
HrQoL	Health-related quality of life
ISS	Idiopathic short stature
PRO	Patient-reported outcomes
QoL	Quality of life
QoLISSY	Quality of Life in Short Stature Youth
SD	Standard deviation
SDS	Standard deviation score
WHO	World Health Organization

1 Introduction

Although clinical endpoints such as blood pressure, pulse, radiography or other laboratory tests are important to describe health (Higginson & Carr, 2001), quality of life (QoL) is becoming another very important outcome to document in clinical medicine (Brutt et al., 2009; Bullinger et al., 2013). Because QoL is a multi-faceted concept that can be defined differently by everyone and for any discipline, measuring QoL is methodologically challenging (US CDCP, 2016). The World Health Organization (WHO) has proposed the definition of QoL to be an "individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards, and concerns" (World Health Organization, 1997). Another definition further specifies different aspects that contribute to the QoL of people such as an individual's health, job, housing situation, school setting, neighborhood, culture, and spirituality (US CDCP, 2016).

This thesis' subject matter revolves around a concept quite similar (and yet different) to QoL: health-related quality of life (HrQoL). HrQoL is defined as the "subjective perception of health and includes aspects of well-being and functioning in physical, emotional, mental, and social domains" (Brutt et al., 2009). It does not specifically refer to the health status of individuals who have illnesses that impair everyday functioning or cause symptoms (Kaplan & Ries, 2007). HrQoL is influenced by individual factors like physical and mental health, health risks, functional status, social support, and socioeconomic status, as well by community factors like conditions, and policies that influence a populations' health perceptions and functional status (US CDCP, 2016).

Research about disease/health and its effects on the quality of life have become increasingly important. Kaplan and Ries argue that illnesses shorten life expectancy and may cause dysfunctions and symptoms that lead to disabilities or difficulties in an individual's daily life. By measuring HrQoL, one can quantify the impact of an illness or condition, compare these impacts with impacts of other diseases, evaluate health changes due to intervention or the progression of a disease (Kaplan & Ries, 2007). HrQoL is also noted to be an important and valid component to public health surveillance that indicates unmet needs and intervention outcomes, and is an even more powerful predictor of mortality or morbidity than most objective measures (DeSalvo, Bloser, Reynolds, He, & Muntner, 2006). HrQoL research and measurements allow the scientific quantification of the impact of health on quality of life, and help determine the burden of diseases,

injuries, and disabilities. In doing so they are also a mode of showing the progress of a nation's health objectives (Bullinger, 2002; US CDCP, 2016; World Health Organization, 1997).

On the healthcare giving side, increased physician knowledge of patients' HrQoL can improve the clinician-patient communication and relationship (Higginson & Carr, 2001; World Health Organization, 1997), make the work of the doctor more meaningful while providing a sense of a more comprehensive and meaningful healthcare experience to the patient (World Health Organization, 1997). It can also help the screening of hidden problems, identifying preferences, prioritizing problems, and training new staff (Higginson & Carr, 2001). Investigating the QoL of populations is not only useful for health economic (cost benefit) analyses (Brutt et al., 2009; Bullinger, 2002), but also for identifying burdens or risks of patients (Ravens-Sieberer, Erhart, Wille, & Bullinger, 2008; US CDCP, 2016), for generating ideas for creating relevant interventions (Ravens-Sieberer, Ellert, & Erhart, 2007; US CDCP, 2016), and for providing direction in changing and developing health policies, allocating resources, developing strategic plans, and developing community interventions (US CDCP, 2016; World Health Organization, 1997).

Although there are many instruments already created to measure HrQoL, there is still a lack of disease-specific HrQoL measurements, which are instruments that ask relevant, disease-specific questions that a generic instrument does not cover. Furthermore, many HrQoL assessments for pediatric populations are also lacking – one such population is short stature children. Because of this lack of short-stature specific-HrQoL instruments, the Quality of Life in Short Stature Youth (QoLISSY) instrument was developed and translated into a variety of languages. Because the QoLISSY has recently been used in a randomized three-arm open label comparator study that tests the efficacy of a new treatment option in comparison to older ones in short statured adolescent boys, this thesis strives to analyze the psychometric performance of this new measure and its usability in accurately assessing HrQoL in intervention studies.

The following subsections will provide a more comprehensive explanation of HrQoL measurements, specific challenges that these measurements must face, developments of HrQoL assessment in pediatric research, as well as more background information about short stature (ex. causes, treatment, literature about how short stature affects HrQoL, etc) and the randomized three-arm open label comparator that this thesis is based on.

1.1 Health-Related Quality of Life: Measurement

HrQoL can be measured by a variety of instruments, which can be classified either as generic or condition-specific, both of which are developed under the idea that HrQoL is multidimensional and that information is patient-derived (Bullinger, 2002). One method of measuring HrQoL is the use of patient-reported outcome (PRO) instruments.

PRO instruments are a form of patient-centered assessment that is defined as a direct report from the patient about his/her health condition and not the interpretation of the patient's condition by a third party, such as a healthcare provider (Schepers, Haverman, Zadeh, Grootenhuis, & Wiener, 2016). In this way, PRO instruments can provide researchers the opportunity to incorporate the patients' perspective on clinical care, research, and clinical trial. Although physical, physiological, or biochemical data can be measured objectively by medical technology, PRO instruments can provide information that cannot be measured without directly asking the patient. These include various symptoms that are not obvious or cannot be observed by observers (ex. depression, headaches, sleep disturbances, etc.), the frequency and severity of symptoms, as well as feedback about treatment measures (ex. satisfaction of treatment) and the impact of the disease on the patient's daily life (Deshpande, Rajan, Sudeepthi, & Abdul Nazir, 2011).

PRO instruments such as the SF-36, SIP, QWB, EQ-5D, and the HUI were all developed to be used across a wide range of populations and interventions to attempt to describe and quantify the multi-faceted concept, QoL (Coons, Rao, Keininger, & Hays, 2000).

Table 1 summarizes the HrQoL measures that were reviewed by Coons et al., as well as information about what HrQoL concepts each instrument intends to measure.

Table 1: Generic health-related quality of life (HrQoL) instruments (Coons et al., 2000).

HrQoL Instrument	**QoL concepts measured**	**Developer(s)**
Medical Outcomes Study 36-Item Short Form (SF-36) Health Survey	Physical functioning, role limitations due to physical problems, bodily pain, general health perceptions, vitality, social functioning, role limitations due to emotional problems, mental health, health transition	RAND (1992)
Nottingham Health Profile (NHP)	Energy level, emotional reactions, physical mobility pain, social isolation, sleep	Hunt et al. (1986)
Sickness Impact Profile (SIP)	Physical dimensions (ambulation, mobility, body care and movement), pyschosocial dimensions (communication, alertness behavior, social interaction), independent categories (sleep and rest, eating, work, home management, recreation and past times)	Gibson, B. & Bergner, M. (1981)
Dartmouth COOP (Primary Care Cooperative Information Project) Charts	Physical fitness, daily activities, social activities, quality of life, overall health, change in health, pain, emotional status, social support	E. Nelson et al., and the Dartmouth Primary Care Cooperative Information Project (COOP Project) (1990)
Quality of Well-Being Scale (QWB)	Mobility, physical activity, social activity, symptoms or problems	Kaplan, R., Anderson, P., and Ganiats, T.G. (1970)
Health Utilities Index (HUI)	Vision, hearing, speech, ambulation, dexterity, emotion, cognition, pain	Torrance et al. (1998)
EuroQoL Instrument (EQ-5D)	Self-care, usual activity, mobility, pain or discomfort, anxiety or depression	EuroQol Group (1990)

1.1.1 Advancements and Challenges in HrQoL PRO Measurement

Advancement in PRO research has grown in recent years. For example, the PROMIS (Patient-Reported Outcomes Measurement Information System) project is developing standardized item banks to measure PROs, such as key symptoms and health concepts like HrQoL that are applicable to many chronic conditions, in the effort to unify the field of PRO measurement across clinical research (Alonso et al., 2013; Cella et al., 2007).

While a generic HrQoL PRO instrument can be used with any population, this type of instrument may not adequately capture the specific burdens and the experiences of having a certain condition (Versteegh, Leunis, Uyl-de Groot, & Stolk, 2012). Therefore, condition-specific instruments are used to measure the HrQoL of patients with a specific condition (Kaplan & Ries, 2007), which can evaluate subtle changes and differences (Price et al., 2009).

Because many HrQoL instruments have been developed in English (North American) (Price et al., 2009), few HrQoL instruments have been culturally validated and translated for international use in different languages/cultures, which makes cross-cultural comparisons challenging (Bullinger et al., 2013). The translation of HrQoL tools is especially important for HrQoL research in populations with rare diseases – because of these diseases' rarity, cohort sizes are often small and international collaboration is pivotal to conduct research that produces significant results (Price et al., 2009).

1.2 Health-Related Quality of Life: Recent Development in Pediatric Research

HrQoL has also become an important outcome measurement in pediatric clinical research, in which physical, emotional, and social domains of well-being and functioning of the child is investigated using the child's perspective, parents' perspectives (or another proxy such as nurses or doctors), or a combination of both (Matza, Swensen, Flood, Secnik, & Leidy, 2004; Solans et al., 2008). This growth of interest in pediatric HrQoL is reported to be partly due to the lack of condition-specific HrQoL instruments that take into account the parents' and patients' perspectives of experiencing illness (Bullinger et al., 2013).

Despite the growing amount of pediatric HrQoL instruments (generic and condition-specific), unique methodological problems to measuring HrQoL in children and adolescents have been encountered. In the past, adults' (such as parents, doctors, caretakers, and other so-called proxy

sources) accounts about young patients' experiences with illness often went unchallenged and were assumed to be accurate descriptions of the children's disease experiences. (Eiser & Morse, 2001). This assumption may hold true for very young patients, who most likely will not understand or complete questionnaires for themselves due to not yet fully developed cognitive and linguistic abilities. However just using information from proxy sources about older children's HrQoL (who can report their HrQoL experiences themselves) may result in an incomplete assessment of the children's HrQoL (Eiser & Morse, 2001). At the same time, proxy reports can also be seen as a complementary source of information to child reports, and it is for this reason that the standard practice of creating new HrQoL instruments also includes determining the correlations between child and proxy reports of HrQoL—if these correlations are poor, then the HrQoL instrument is seen as inadequate. However, this argument is not without fault. Proxies and their children may very well not agree on many issues (Eiser & Morse, 2001).

These differences in perspectives are further complicated when using either generic or condition-specific HrQoL instruments with children and adolescents. Quitmann et al. (2016) aimed to explore the levels of agreement and disagreement between short-statured children and parent reports of generic and condition-specific HrQoL and found that parent and child agreement on reported HrQoL was strongly correlated and that parents tended to report their children's HrQoL to be lower than how their children report their own HrQoL in both generic and condition-specific instruments, with condition-specific instruments having fewer discrepancies in reported HrQoL (Julia Quitmann, Rohenkohl, Sommer, Bullinger, & Silva, 2016).

1.3 Theoretical Background of Short Stature

In order to understand the mechanics of how short stature is caused and medically treated, understanding how normal bone growth occurs (and possible ways this growth can be delayed/inhibited entirely) is crucial.

Normal bone growth during development is characterized by endochondral ossification, which is the formation of bone within cartilage. This process begins when a cartilage model of the bone is formed from differentiated mesenchymal cells (precursors to connective tissue cells) called chondrocytes. Over time, this cartilage model is gradually replaced by bone (this process is called ossification), starting from the shaft towards the ends of the bone (epiphyses), eventually leading to all the cartilage being replaced by bone, except for one small region called the epiphyseal growth

plate. Growth continues if the epiphyseal growth plate exists, where bone can continue to develop from the continual proliferation of cartilage in this area (Nilsson et al., 2014). Figure 1 illustrates the ossification process as cartilage is replaced by bone matter, as well as the process of bone growth.

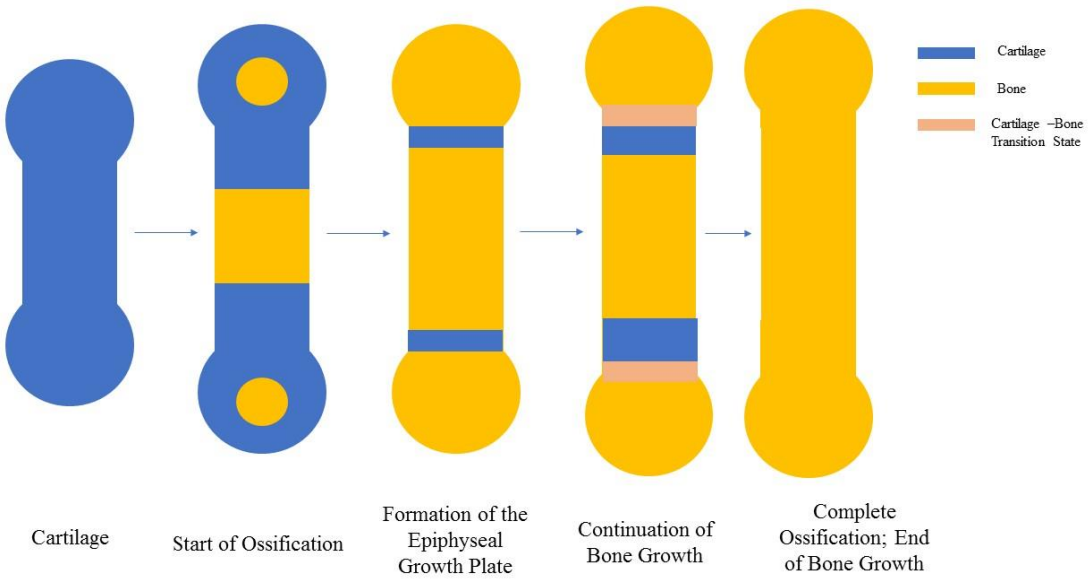

Figure 1: Process of bone growth

This continued growth is stimulated by an endocrine regulation system involving growth hormone (GH), in which the hypothalamus signals the body to release GH, which stimulates the production of insulin-like growth factor (IGF-I), which in turn signals the proliferation of chondrocyte cells in the epiphyseal growth plate (Nilsson et al., 2014). Recent studies have also found that estrogen plays a role in bone growth (Borjesson, Lagerquist, Windahl, & Ohlsson, 2013) in females as well as males, in which androgen is converted into estrogen by a special class of enzymes called aromatase (Hess, 2003). These combined interactions lead to proliferation of cartilage and eventual bone development. During late adolescence, hormones (estrogen and androgen) cause the epiphyseal growth plate to close, resulting in total bone formation and the end of bone growth (Borjesson et al., 2013; Mackie, Ahmed, Tatarczuch, Chen, & Mirams, 2008). This formation of bone of the epiphyseal growth plate is called epiphyseal fusion (Emons, Chagin,

Savendahl, Karperien, & Wit, 2011). Figure 2 illustrates a simplified summary of the regulation of bone growth in males and females.

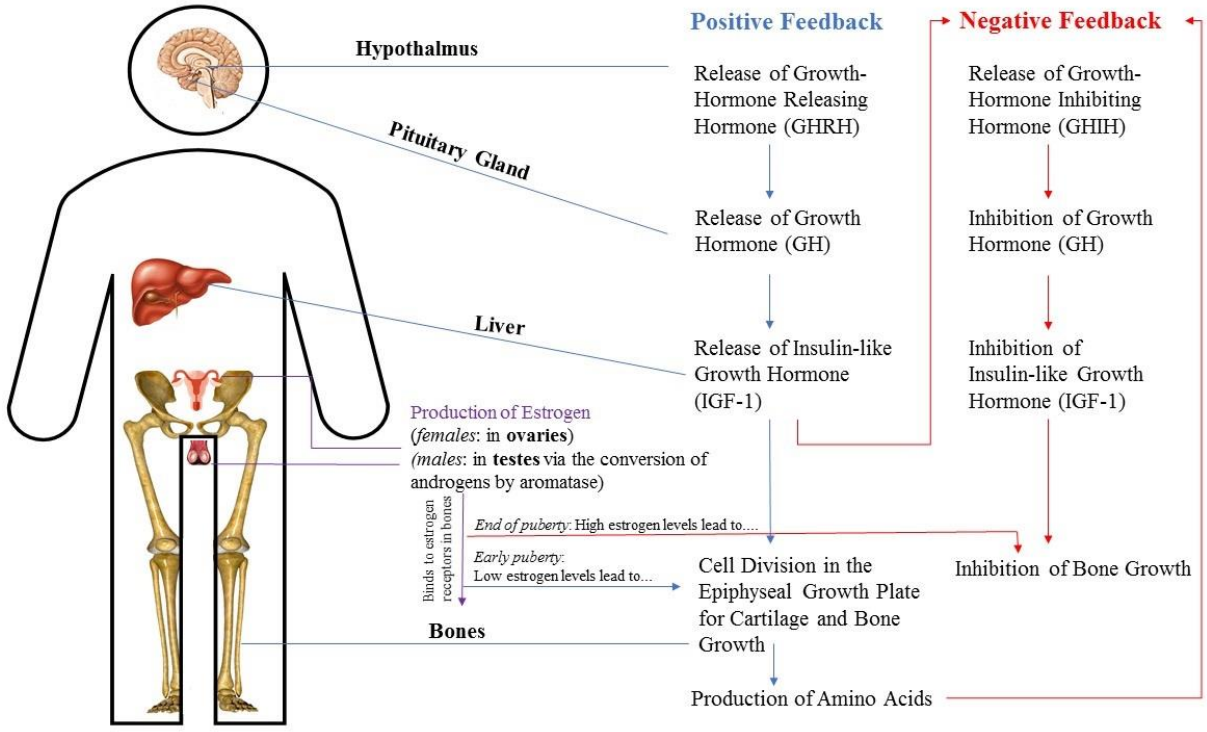

Figure 2: Regulation of bone and cartilage growth

Bone maturation and growth may be impeded through a variety of factors such as genetics, poor nutrition, or may be directly or indirectly delayed by a disease or disorder, which may result in short stature (Waqar Rabbani, Imran Khan, Bila Afzal, & Rabbani, 2013). Normal variation of growth that leads to short stature includes idiopathic short stature (ISS) (Cohen et al., 2008). Short stature is defined in ISS as being 2 standard deviations below the mean height of the corresponding sex and age group, in the absence of any apparent cause (Cohen et al., 2008; GH Research Society, 2000). However abnormal variation of growth that leads to short stature could also be due to pathological causes, such as having a growth hormone deficiency (GHD).

1.3.1 Treatment of Short Stature

Treatment options have become available for short stature cases of endocrine origin, such as GHD, and non-endocrine origin such as ISS. However, treatment has only been approved in a specific number of countries, one being the US (Grimberg et al., 2016). Treatment options that have been used include growth hormone replacement treatment (GH), which has been shown to improve growth in both GHD and ISS patients (Dahlgren, 2011; Ranke et al., 2007; Richmond & Rogol, 2016), but the efficacy of GH has been disputed because of conflicting factors such as the possible confounding impact of normal pubertal development during treatment and the severity of GHD (Richmond & Rogol, 2016).

Aromatase inhibitors (AI) have also been described to be a possible treatment option for short stature (Cohen et al., 2008; Ferris & Geffner, 2017; Geffner, 2009; Hero, 2016; McGrath & O'Grady, 2015). The mechanism of aromatase inhibitors essentially results in delaying epiphyseal fusion and increasing the period of growth. This is done by inhibiting the aromatase enzyme, which is responsible for converting androgen (hormones that are involved with the development of male characteristics) into estrogen (see Figure 2 for the role of aromatase/estrogen in the regulation of bone growth). The presence of estrogen has been shown to accelerate the process of epiphyseal fusion (Borjesson et al., 2013; Eshet et al., 2004; Hess, 2003). Because estrogen plays an integral role in pubertal development in females, the use of AI is not recommended to increase adult height in teenage girls (McGrath & O'Grady, 2015).

A recent RCT has also found that combination therapy (GH + AI) over the course of 24 months increases height potential in ISS patients in the US and in Chile compared to ISS patients who had undergone GH-only treatment or AI-only treatment to treat short stature (Mauras et al., 2016).

1.4 Quality of Life in Short Stature Youth and Measurement

Not only has research been directed towards creating possible treatment options for short stature children, but also towards the HrQoL experience of this population. Short stature has been documented in pediatric populations to be associated with negative impacts on children's psychosocial well-being (Attanasio, Shavrikova, Blum, & Shalet, 2005; Blum et al., 2003), occurrences of bullying, social isolation and stigmatization (Brutt et al., 2009; Voss & Mulligan, 2000), lower social competences and an increase in behavioral problems (Wit et al., 2008), and a higher likelihood of developing depression (Abe et al., 2009). These outcomes may be especially

profound for short males than it is for females because of the stereotype that tall stature is more important for boys than girls (David E. Sandberg, Bukowski, Fung, & Noll, 2004). This stereotype may be one of the reasons why short stature boys are referred to pediatric endocrinologists for evaluation and treatment more often than short stature girls are (August et al., 1990; D. E. Sandberg, Brook, & Campos, 1994).

The measurement of HrQoL of short stature youth has used generic instruments like the PedsQL Generic Core Scales (Stephen et al., 2011; Wu, Li, & Gao, 2013), the KIDSCREEN and DISABKIDS questionnaires (Neuza Silva, Bullinger, Quitmann, Ravens-Sieberer, & Rohenkohl, 2013), the KINDL questionnaire (Geisler et al., 2012), and the Self-Perception Profile and Youth Self Report (D. E. Sandberg et al., 1994).

A few short-stature specific HrQoL instruments have been created, such as the Idiopathic Short Stature QoL (ISSQOL) Questionnaire, the TACQOL-S questionnaire, Issues Related to Growth Problem and Height Questionnaire (IRGPH) and the Growth Hormone Injection Questionnaire (GHIQ), however self-report measurements of short stature youth are noted to be lacking (The European QoLISSY Group, 2013b). Table 2 below provides a summary of the HrQoL concepts measured by the respective instruments.

Table 2: Examples of health-related quality of life (HrQoL) instruments used for short stature youth.

HrQoL Instrument used for Short Stature Youth	**QoL concepts measured**	**Developer(s)**
PedsQL Generic Core Scales	Physical, emotional, social and school-related domains	James W. Varni (1998)
KIDSCREEN	Physical well-being, psychological well-being, mood and emotions, self-perception, autonomy, parent relations and home life, financial resources, peers and social support, school environment, bullying	Ulrike Ravens-Sieberer et al. (2001-2004)
DISABKIDS	Mental (independence, emotion), social (inclusion, exclusion), and physical (limitations, medications) domains	DISABKIDS group (2002)
KINDL	Physical, emotional, self-esteem, family, friends, school, disease domains	Ravens-Sieberer & Monika Bullinger (1998)
Self-Perception Profile	Global self-worth, athletic competence, and romantic appeal	Susan Harter (1985)
Youth Self Report	Anxiety, depression, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior	Achenbach System of Empirically Based Assessment (1991)
Idiopathic Short Stature QoL (ISSQOL)	Vitality (energy level)	Bucciardini et al. (2006)
Child Quality of Life (TACQOL-S)	Physical abilities, vitality, contact with peers, contact with adults, and body	Vogels et al. (2003)
Issues Related to Growth Problem and Height Questionnaire (IRGPH)	Concerns & satisfaction about appearance, advantages/disadvantages of height, stigmatization & juvenilization, competence in physical activities, and school reputation.	Sandberg & Mazur (1990)
Growth Hormone Injection Questionnaire (GHIQ),	Feelings related to GH treatment and “injection issues”	Cramer et al. (2003)

1.5 QoLISSY Instrument: Original Study, Development, and Original Results

Because of the lack of instruments that measure the HrQoL of short stature youth, the Quality of Life in Short Stature Youth (QoLISSY) instrument was developed. The QoLISSY tool assesses several aspects of quality of life (physical, emotional, social, beliefs about height, and treatment-related factors), targets short-stature youth (The European QoLISSY Group, 2013b), and has been translated and validated into 5 languages: German, English, Swedish, French, and Spanish. The original aim of the QoLISSY project was to create a psychometrically sound and cross-culturally valid tool to assess the impacts of short stature on the QoL in children/adolescents from their own perspective with the added perspective of their parents.

The QoLISSY instrument includes a patient report (for children ages 8-12 and 13-18 years old) and a parent report instrument (for parents of children ages 4-7, 8-12, and 13-18 years old.). These QoLISSY instruments all cover three main QoL domains (physical, emotional, and social), as well as three additional domains of QoL (coping, beliefs, and treatment), which include items that pertain to the coping strategies used by the children, beliefs of short stature, and growth treatment. The QoLISSY instrument also provides a summary HrQoL score (Total Score), which is the sum score of the Physical, Emotional, and Social scale. The parent report instrument also includes two extra domains specifically for parents: Future and Effects on Parents. These scales assess what the parents think their children worry about regarding their future (ex. “My child worries that when he/she is older, his/her height will make getting certain jobs harder” or “My child worries that when he/she is older, his/her height will make getting a girlfriend/boyfriend harder.”) and also how the short stature of their children affect the parents (ex. “My child’s growth problems make me feel anxious” or “I am worried that I am overprotective of him/her.”). Possible answer options were presented on a 5-point Likert scale (Not at all/never, Slightly/seldom, Moderately/quite often, Very/very often, Extremely/always).

The development of this instrument involved international cooperation between 5 countries (Sweden, Germany, France, UK, and Spain), in which children and adolescents diagnosed with GHD and ISS (and their parents) were recruited to participate in the item elicitation process of the questionnaire (via focus group discussions and cognitive debriefing). After creating items for the questionnaire, the QoLISSY’s psychometric performance was evaluated in a field test, as well as

a re-test test to evaluate the questionnaire's test-retest reliability. A sample of 60 patients and their parents were recruited from each country.

The field test revealed that in the self-report, children had an elevated level of quality of life, and the scales had good reliability values (Cronbach's alpha, split-half reliability, test-test reliability, and intraclass correlation coefficient), as shown in Table 3. The psychometric performance of the QoLISSY's parent scales was also satisfactory, as shown in Table 4.

Table 3: Descriptive statistics and reliability results for children subscales and total quality of life score of the QoLISSY (The European QoLISSY Group, 2013).

QoLISSY Self Report									
Domain	Descriptives						Reliability		
	n	Mean	SD	Skewness	% Floor	% Ceiling	**α***	Split-half*	ICC*
Physical	268	73.69	22.08	-0.95	0.4	12.3	0.84	0.83	0.80
Social	268	72.94	22.93	-0.78	0.4	10.4	0.87	0.83	0.80
Emotional	268	72.69	23.87	-0.94	1.1	10.1	0.88	0.88	0.85
Coping	257	55.60	22.38	-0.19	1.2	1.2	0.82	0.65	0.56
Beliefs	266	69.13	28.59	-0.78	3.8	21.1	0.85	0.85	0.83
Treatment	152	55.12	21.06	-0.13	0.7	07.	0.87	0.74	0.73
Total Score	268	73.10	21.39	-0.80	0.4	4.9	0.95	0.92	0.88

*Values at 0.7 and over is considered good for group comparisons

Table 4: Descriptive statistics and reliability results for parent subscales and total quality of life score of the QoLISSY (The European QoLISSY Group, 2013).

QoLISSY Parent Report									
Domain	Descriptives						Reliability		
	n	Mean	SD	Skewness	% Floor	% Ceiling	**α***	Split-half*	ICC*
Physical	317	73.80	23.20	-0.76	1.3	13.2	0.86	0.86	0.84
Social	313	69.41	25.24	-0.60	0.3	12.5	0.90	0.88	0.84
Emotional	315	68.50	22.90	-0.66	0.3	5.4	0.88	0.90	0.70
Coping	286	45.07	20.96	0.03	3.1	0.7	0.83	0.65	0.64
Beliefs	310	67.62	28.95	-0.62	2.6	22.2	0.90	0.89	0.72
Treatment	163	55.18	20.60	-0.23	0.6	1.2	0.88	0.78	0.74
Future	303	74.85	26.45	-1.11	1.3	20.5	0.90	0.86	0.74
Effect on Parents	313	65.68	24.48	-0.44	1.6	5.1	0.90	0.82	0.88
Total Score	313	69.97	22.03	-0.65	0.3	2.9	0.95	0.91	0.86

*Values at 0.7 and over is considered good for group comparisons

1.6 QoLISSY Instrument: First Use in a Randomized Open Label Comparator Trial and Initial Results

Until a randomized open label comparator trial conducted by Mauras et al. (Mauras et al., 2016), the QoLISSY instrument had never been used in a randomized clinical trial of a growth intervention. It was not only found that combination therapy (AI+GH) increases height potential, but also that the treatment positively impacts HrQoL in short stature youth (Mauras et al., in preparation).

Although these results are promising and provide support that the combination treatment AI + GH is a viable treatment option with beneficial outcomes to drug regulatory bodies, a natural follow up question to the study results is whether QoLISSY has indeed accurately and consistently measured the targeted construct (HrQoL) for these short stature adolescents. In other words, is the QoLISSY instrument a valid tool to use for this target population? The answer to this question can be discovered by performing appropriate tests that assess the psychometric properties of the QoLISSY instrument, which will be conducted and interpreted in the present thesis.

1.7 Research Goals of the Thesis

Because this is the QoLISSY's first usage in a randomized open label comparator trial, this thesis will aim to describe and assess the psychometric performance of the English and Spanish version of the QoLISSY that was used in a randomized open label comparator trial that explored the effects of AI growth therapy in adolescent boys diagnosed with ISS and their parents in the US and in Chile.

The following research questions will also be investigated:

1. Is the QoLISSY instrument a psychometrically sound tool for use in randomized open label comparator trials?
2. Can QoLISSY detect differences of parent and/or child judgments of HrQoL, thus lending support for the idea of (dis)agreement between proxy and self-reported PROs, and what are the implications of these results?
3. Can the QoLISSY instrument detect changes of HrQoL over time?

In addition to these aims, the present thesis will give a summarized explanation about psychometric assessment and explore whether the results of this study contradict or support previous results in similar studies.

2 Methods

In this section of the thesis, a brief description of the original design of the randomized open label comparator trial will be provided, as well as explanations of various psychometric properties that may be explored in the QoLISSY instrument. Lastly, the data analysis process of the thesis will be introduced.

2.1 Design of the Randomized Open Label Comparator Trial (Mauras et al., 2016)

With the cooperation of different endocrine clinics located in the US and Chile, 76 boys diagnosed with idiopathic short stature, ages 12 to 18 years old and their parents were asked to participate in a randomized open label comparator trial that aimed to compare the impacts of AI treatment, GH treatment and a combination of AI and GH (denoted henceforth as AI/GH). All patients were treatment naïve and had normal birth weight. Participants were randomized into one of the three treatment types (AI, GH, or AI/GH). Clinical variables such as bone age and anthropometry were obtained in addition to HrQoL assessments using the QoLISSY and the generic KIDSCREEN questionnaires before and after 12, 24, and 36 months of treatment.

2.2 Quality Criteria for Measurement Properties of Health Status Questionnaires

A criteria list was developed to be used to detect shortcoming and gaps in knowledge of measurement properties and to help design validation studies (Terwee et al., 2007). The present thesis will use this criteria list as a basis for the evaluation of the psychometric performance of the QoLISSY in its use in the abovementioned randomized open label comparator trial. In the following paragraphs, each quality criterion listed is explained.

Content validity is defined to be the extent in which the items of an instrument represent the concept the instrument is intended to measure (Fields, 2009c; Guyatt, Feeny, & Patrick, 1993). Terwee et al. provide a list of characteristics that a health assessment should include in order to prove content validity (2007):

- Clear statement of what concept the instrument intends to measure
- Clearly states which target population the instrument is intended to be used in
- Definition of the concepts the instrument is meant to measure
- Justification for item selection and reduction
- Interpretability of items (the questionnaire itself should be short and simple and should be understandable for each appropriate age group the instrument is made for)

Internal consistency is defined as the extent to which items of a scale are correlated with other scales, indicating whether all these scales measure the same concept that the questionnaire is supposed to measure (Terwee et al., 2007). To evaluate internal consistency, the recommended method is to conduct a factor analysis or principal component analysis on a sample population of at least n= 100 to evaluate whether the items form one overall scale or should be divided into more than one scale. It is also recommended that, if a questionnaire has subscales, Cronbach's alpha values for each subscale be calculated separately (Fields, 2009b; Terwee et al., 2007).

Criterion validity is defined as the extent to which a health instrument's score relates to a gold standard instrument that measures the same intended construct (Terwee et al., 2007). Another definition of criterion validity is whether an instrument measures what it is supposed to measure (Fields, 2009c) by comparing its score with another instrument's scores that measures the same construct and see if these scores correspond (Fields, 2009a). In the quality criteria list for measurement properties of health status instruments by Terwee and colleagues, an instrument has criterion validity if it can fulfill two requirements. First, the instrument developers must present a convincing argument the gold standard instrument they used is a gold standard instrument. And second, the scores of the instrument under study and the gold standard instrument must have correlations of at least 0.70 (2007).

Construct validity is defined as the extent in which scores of an instrument relate to other instruments (Terwee et al., 2007; Westen & Rosenthal, 2003). Arguably, construct validity is the most important property of an instrument because if it lacks this, then whatever results that are

obtained using this instrument become difficult to interpret. Despite the importance of this concept, there is no simple way to evaluate whether an instrument is indeed construct valid.

Typically, researchers who develop instruments argue for the construct validity of their instrument by presenting correlations between their instrument and other similar instruments that should theoretically measure the same construct and by presenting negative correlations between two different instruments that do not measure the same construct (Westen & Rosenthal, 2003). In simpler terms, researchers try to provide evidence that their instrument measures what it should (convergent validity) and doesn't measure what it shouldn't (divergent validity).

Reproducibility is defined as the extent to which an instrument yields similar answers even after repeated measurement (Terwee et al., 2007). Terwee et al. deem an instrument to have reproducibility if it can show that it has agreement and is reliable and responsive. These concepts are defined below.

Agreement is the extent to which scores of a scale after repeated measurement are similar to each other, which is important when an instrument is being evaluated for its ability to distinguish clinically important changes from measurement error (Terwee et al., 2007). Terwee and colleagues deem an instrument to show agreement if (a) the defined minimal important change is less than the smallest detectable change of score, (b) minimal important change of score is outside the limits of agreement, or (c) there is a convincing argument that agreement of scores is acceptable.

Reliability is defined as the degree to which patients can be distinguished between each other regardless of measurement error (Terwee et al., 2007) or as the degree to which a questionnaire measures the same way even when subjects answer the questionnaire at different time points (Fields, 2009c; The European QoLISSY Group, 2013b). High reliability is favored for discriminative purposes (ex. the health instrument can distinguish between patients with less or more severe forms of a disease). Terwee et al. recommend calculating intra-class correlations (ICC) or weighted Kappa values to evaluate the reliability of an instrument, in which values over 0.70 show acceptable reliability. However, others suggest calculating the Cronbach's alpha for each scale, which appears to be the most common way to evaluate the reliability of instruments (Fields, 2009b).

Responsiveness is defined as an instrument's ability to detect minimal clinically important differences (MCID) over time, even if these differences are small. MCID is defined to be the smallest difference of a score in the domain of interest that patients believe is beneficial (Cook,

2008). Defining MCIDs has proven problematic because of patients' variability in evaluating improvement. When asked to report health changes from baseline health, some may report their health status in comparison to expectations of what their health should be or to healthy peers (Cook, 2008). In addition to this dilemma, retrospective evaluation of improvement may be subject to recall bias because patients may not remember the nature of their health condition at baseline measurement. In addition to this dilemma of defining MCID, there is also no gold standard of how to evaluate responsiveness, and depending on which method used, the calculated MCID can vary widely (Kosinski, Zhao, Dedhiya, Osterhaus, & Ware, 2000). Because there is no standard way to calculate MCID, this has resulted in problems in its methodology and interpretation.

However, Terwee and colleagues have made suggestions about how responsiveness may be evaluated in a health instrument. They deemed an instrument to be responsive if the ratio of the change seen in patients who have undergone an intervention of known efficacy and matched patients who have not undergone the intervention is more than 1.96. Another method they suggested was to evaluate the area under the receiver operating characteristics curve (AUC), which distinguishes patients who have or have not changed based on an external criterion. They deemed an AUC of at least 0.70 to reflect acceptable responsiveness of a health instrument.

Floor and ceiling effects are detected through the number or percentage of respondents who have achieved the lowest or highest possible score. An instrument has a floor or ceiling effect if more than 15% of respondents have achieved the lowest or highest possible score in a sample size of at least 50 patients (Lim et al., 2015; McHorney, 1999; Terwee et al., 2007). If a floor or ceiling effect is present, this may be indicative of limited content validity (lack of "extreme" items at the lower or upper end of the scale), reliability (those who have achieved either extreme score can't be distinguished from each other), and responsiveness (changes cannot be measured in patients who have scored the highest or lowest possible score) (Terwee et al., 2007).

Interpretability is defined as the extent to which qualitative meaning can be assigned to a quantitative score. Terwee and colleagues recommend that instrument developers should define what change in score is clinically meaningful and provide means and SD of scores of a reference population and relevant subgroup of patients that (a) are expected to differ in scores, (b) have undergone a treatment of known efficacy (baseline and after treatment data measurement), and (c) have reported global ratings of change. They deem an instrument to be interpretable if developers

of the instrument have presented the means and SDs of scores of the abovementioned groups and have provided a definition of what score is clinically meaningful.

2.3 Data Analyses

All statistical analyses were conducted using SPSS 23. The following subsections detail the data preparation and data analyses done in this thesis. The significance level for all analyses was set at $p<0.05$.

2.3.1 Data Preparation

Because most of the items in the QoLISSY are formulated positively and it is in concordance to the scoring that a higher score reflects higher HrQoL, negatively formulated items were recoded to reflect higher HrQoL (ex. "Have you felt sad?" and respective answer options 1="Not at all," 2="Sometimes," 3="Often," 4="Always" are recoded in which "Not at all" reflects best HrQoL). Mean scale scores were computed and missing data that were random and less than 20% of the values were replaced with the individual mean score for each variable. All scores were transformed from raw scores to 0 to 100 scores, with higher values representing higher HrQoL.

2.3.2 Sample Statistics and Descriptive Analyses of Scales

Sample characteristics (means of age, height, SDS height, and respective SD values) were calculated for both Chilean and American samples. These variables were also analyzed using an independent sample analysis of variance (ANOVA) for all intervention groups at baseline to detect any significant differences. Because the sizes of the US and Chile samples are small and because no significant differences were found between these two samples, data from both samples were used together in all the analyses included in this thesis. After determining sample characteristics, QoLISSY scale distribution characteristics such as mean, SD, skewness, kurtosis, and floor and ceiling effects were examined. Floor and ceiling effects were calculated by determining the worst and best possible scores (0 and 100, respectively) and determining the percentage of respondents who achieved these scores per scale. If these percentages exceeded 15%, a floor or ceiling effect was indicated (Lim et al., 2015; McHorney, 1999; Terwee et al., 2007).

2.3.3 Content Validity & Internal Consistency

Content validity was evaluated based on the criteria stated by Terwee et al. Because the sample size in this study was less than 100, the Cronbach's alpha values for each scale were assessed to evaluate the internal consistency of the QoLISSY, in which acceptable values were defined to be between 0.70 and 0.95 (Fields, 2009b; Nunnally & Bernstein, 1994; Terwee et al., 2007).

2.3.4 Construct Validity

Construct validity was assessed by evaluating the convergent and discriminant validity of the QoLISSY instrument. Convergent validity was evaluated by calculating Pearson correlation coefficients between the domains of the QoLISSY instrument and the KIDSCREEN-52 instrument domains, which is a questionnaire that is intended to measure HrQoL of adolescents between 8 and 18 years, albeit it is not condition-specific.

There are three versions of the KIDSCREEN questionnaire— a long version (KIDSCREEN-52), which covers ten HrQoL dimensions, as well as a short version (KIDSCREEN-27) and a global HrQoL assessment (KIDSCREEN-10 Index). All versions of the KIDSCREEN questionnaire have a self-report and proxy report available. The KIDSCREEN-52 was administered in the randomized open label comparator trial. The KIDSCREEN-52 includes the following domains in both the self- and proxy-report: Physical Wellbeing, Psychological Wellbeing, Moods & Emotions, Self-Perception, Autonomy, Parents, Financial, Peers, School, and Bullying (KIDSCREEN Group Europe, 2006).

Before conducting analyses, the KIDSCREEN-52 data was prepared as follows: KIDSCREEN items that are negatively formulated were recoded so that higher scores reflect higher HrQoL (KIDSCREEN Group Europe, 2006). Items of each scales were summed up (raw scale score). Study participants with missing data were not included in the calculation of raw scale scores. These raw scale scores were exchanged for Rasch person parameters estimates, which are provided in the manual, and then transformed into z-values and t-values. These values are based on data from an international survey. Preparing the data this way allows t-values to have scale means of 50 and an SD of 10. Higher t-values equate to higher HrQoL.

Regarding convergent validity, the Physical, Social, and Emotional domains of the QoLISSY were expected to have the highest correlation values with the Physical Wellbeing, Psychological

Wellbeing and Mood, and Social domains of the KIDSCREEN instrument due to similar questions and content. Pearson correlation coefficients of more than 0.5 were regarded as large correlation values, 0.31 to 0.5 as moderate, and 0.1 to 0.3 as low (Mukaka, 2012; The European QoLISSY Group, 2013b).

In addition to this, convergent validity was also assessed by evaluating the inter-correlation of the QoLISSY scales with each other to determine if the individual scales are indeed correlated with the same construct (HrQoL). This assessment was done using Pearson correlations, with large, moderate, and low correlation coefficients being defined by the same criteria stated above.

As a first step to evaluate discriminant validity, Pearson correlation analyses using scale data (at both data collection points, as well as the self- and parent-report) and possible relevant demographic data (ex. age, previous GH treatment, height) were conducted to determine possible covariates that can be included in MANCOVA and ANCOVA analyses, if appropriate. Box's M Test of Equality of Covariance Matrices was evaluated to determine if homogeneity of the coefficients for the covariates across the levels of factor (AI, GH, or combination treatment) can be assumed and if the model's results are valid.

Discriminant validity (also known as known-groups validity) was evaluated by conducting multivariate analyses of variances (MANOVA) and, if appropriate, multivariate analyses of covariances (MANCOVA) on all QoLISSY scales except for the Total Score scale. For the Total Score, Analyses of Variance (ANOVA) and, if appropriate, Analyses of Covariance (ANCOVA) were used. These analyses were conducted to determine if there are differences between scale scores in the study's participants who received GH, AI, or the combination therapy.

2.3.5 Responsiveness

Responsiveness was evaluated by conducting repeated measurement ANOVA analyses (for the Total Score scale) and repeated measurement MANOVA analyses (for all other QoLISSY scales) for the child and parent report for both measurement points. These analyses were conducted to determine if the QoLISSY can detect changes in HrQoL, and to determine if changes of HrQoL are due to the interaction between the treatment type received and time. If no interaction was found, follow up analyses where conducted to determine if changes of HrQoL where simply due to time or treatment type. In this model, time (baseline and 24-months) was appointed as the within-subjects factor and the treatment type received (AI, GH, or combo therapy) was the between-

subjects factor. Profile plots were also created for preliminary analyses of possible interactions between treatment and time.

2.3.6 Parent-Child (Dis-)Agreement

Correlations between parent and child scores were also evaluated using Pearson correlation coefficients for both measurement points.

2.3.7 Not Evaluated: Interpretability, Agreement, Reliability, and Criterion Validity

Interpretability was not evaluated because this construct falls outside of the realm of the topic of this thesis and is an issue of instrument development.

Agreement and reliability (functions of reproducibility) were not evaluated because they do not fit into the design of this intervention study. Typically, these properties are assessed during the development of an HrQoL instrument in a test-retest phase, when a questionnaire is administered 1-2 weeks after the original administration and scores are not expected to differ. This process was done during the development of the QoLISSY instrument, which yielded satisfactory ICC values (The European QoLISSY Group, 2013a)

Because no gold standard instrument yet exists to measure the HrQoL in short stature youth, criterion validity was also not statistically evaluated.

3 Results

3.1 Sample Description and Statistics

A total of n=76 boys participated at baseline and n=57 at the 24-month measurement point, of which n=50 boys were from the US and n=26 boys were from Chile. Table 5 presents descriptive statistics of the sample at both measurement points and countries.

Table 5: Descriptive Statistics of the sample at baseline and 24 months after treatment.

	Baseline				24 Months			
Group	**N**	**Mean age (SD)***	**Mean height in cm (SD)***	**Mean height SDS (SD)***	**N**	**Mean age (SD)***	**Mean height in cm (SD)***	**Mean height SDS (SD)***
AI	**25** US= 16 Chile=9	14,16 (1.0)	145.7 (5.5)	-2.2 (0.3)	**16** US= 12; Chile= 4	16.16 (1.0)	159.9 (5.0)	-1.7 (0.5)
GH	**25** US=17, Chile=8	14.09 (1.2)	144.2 (7.0)	-2.4 (0.4)	**20** US=13; Chile= 7	16,09 (1.2)	161.1 (7.7)	-1.4 (0.7)
AI/GH	**26** US= 17, Chile=9	14.04 (1.1)	144.5 (6.8)	-2.3 (0.4)	**21** US= 13; Chile=8	16.03 (1.1)	163.1 (5.3)	-1.3 (0.6)
Total	**76** US= 50; Chile=26	14.09 (1.2)	144.8 (6.4)	-2.3 (0.4)	**57** US=38; Chile=19	16.09 (1.1)	161.5 (6.3)	-1.5 (0.6)

* Means are calculated after combining both US and Chile samples together

ANOVA tests for both baseline and 24-month data also revealed that there is no significant difference in mean age between American and Chilean boys—this holds true for mean height and SD height. The significant level was placed at $p<0.05$

Because the US and Chilean samples showed no significant differences, the data analyses in this thesis combine these two samples together. Otherwise the sample size would have been too small and would have further limited the interpretability of the results.

After combining the US and Chilean samples together, ANOVA tests for both measurement points also show that there are no significant differences between intervention groups in terms of mean age, height, and SD height. The significant level was placed at $p<0.05$.

3.2 Descriptive Analyses and Reliability Analyses of Scales (All Treatment Types)

The score of the QoLISSY subscales and Total Score represent the HrQoL of short-statured children and adolescents from their point of view (child report) and their parents' point of view (parent report). To interpret these scores, all scores were transformed from raw scores to 0 to 100 scores with higher values representing higher HrQoL.

With the exception of the Coping scale and the Beliefs scale (parent-report only) at baseline and the Coping scale (child-report only) at 24 months, scale characteristics (mean standard

deviation) of both the child and parent reports at baseline and 24 months show a negative skewness, indicating a favoring of a higher HrQoL within the range of 0 to 100. A ceiling effect was detected in the Beliefs scale in the child report at 24 months. All QoLISSY scales of the child report had respectable Cronbach alpha values, ranging between α=0.77 (Physical and Coping) and α=0.94 (Total Score) at baseline and between α=0.82 (Physical) and α=0.94 (Total Score) at 24 months (Table 6).

With the Coping scale as an exception, all QoLISSY scales of the parent report also had respectable Cronbach alpha values, ranging between α=0.87 (Social and Emotional) and α=0.95 (Total Score) at baseline and between α=0.89 (Physical and Emotional) and α=0.95 (Total Score) at 24 months (Table 6).

Table 6: Descriptive Statistics and Reliability for Total Score and Subscales at Baseline and 24 Months.

Descriptive Statistics and Reliability for Total Score and Subscales																			
		Baseline Domains (# of items)								24 months Domains (# of items)								
Descriptive Analyses	**Report Type**	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Future (5)**	Effect on Parents (11)**	Total Score (22)	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Treatment (14)	Future (5)**	Effect on Parents (11)**	Total Score (22)
n	*Child report*	75	75	75	75	75			75	58	58	58	57	57	51			58
	Parent report	75	74	75	73	74	74	74	74	58	58	58	58	58	50	74	74	74
Mean	*Child report*	62.53	61.80	63.18	45.11	49.58			62.51	77.61	74.92	75.62	45.28	61.26	55.26	--	--	76.05
	Parent report	57.82	50.69	49.94	44.99	44.34	51.89	54.43	52.84	75.75	71.00	65.48	48.20	55.50	64.87	51.89	54.43	70.74
SD	*Child report*	18.63	19.23	20.81	17.51	27.17			17.42	16.11	17.66	18.31	21.62	29.64	21.01			15.71
	Parent report	18.87	20.64	21.48	14.46	26.05	26.54	21.72	18.59	19.07	20.74	20.31	16.84	25.49	19.28	26.54	21.72	19.08
Skewness	*Child report*	-0.57	-0.41	-0.83	0.33	-0.13			-0.66	-1.16	-0.86	-0.68	0.24	-0.39	-0.28			-0.70
	Parent report	-0.20	-0.31	-0.41	0.61	0.14	-0.11	-0.05	-0.27	-1.14	-1.26	-0.91	-0.01	-0.12	-0.87	-0.11	-0.05	-1.22
Kurtosis	*Child report*	0.33	-0.53	0.55	0.66	-0.81			0.65	2.21	0.46	-0.32	-0.28	-0.77	-1.04			-0.13
	Parent report	-0.57	0.55	0.55	0.56	0.55	0.55	0.55	-0.60	1.34	1.75	0.93	0.15	-0.60	1.22	-0.71	-0.92	1.71
% Floor	*Child report*	0.0	0.0	1.3	1.3	6.6			0.0	0.0	0.0	0.0	1.8	5.3	0.0			0.0
	Parent Report	0.0	0.0	1.3	0.0	5.4	5.3	0.0	0.0	0.0	0.0	0.0	1.7	3.4	0.0	3.5	0.0	0.0
% Ceiling	*Child report*	0.0	0.0	1.3	1.3	4.0			0.0	8.6	3.4	5.2	1.8	**15.8**	0.0			1.7
	Parent report	0.0	0.0	0.0	0.0	1.3	2.6	0.0	0.0	5.2	3.40	1.7	0.0	6.9	2.0	12.3	1.7	1.7
Reliability Analyses*																		
α	*Child report*	**0.77**	**0.84**	**0.85**	**0.77**	**0.87**			**0.94**	**0.82**	**0.86**	**0.89**	**0.83**	**0.92**	**0.91**			**0.94**
	Parent report	**0.89**	**0.87**	**0.87**	0.65	**0.91**	**0.90**	**0.89**	**0.95**	**0.89**	**0.90**	**0.89**	**0.79**	**0.90**	**0.91**	**0.93**	**0.91**	**0.95**

*Only included in the parent report

3.3 Descriptive Analyses and Reliability Analyses of Scales by Treatment Type

3.3.1 Aromatase Inhibitors (AI)

With the exception of the Physical scale (parent-report only), Coping scale, Beliefs scale, Future and Total Scales (parent-report only) at baseline and the Physical, Coping, and Total Score scales (parent-report only) and the Beliefs scale at 24 months, scale characteristics (mean, standard deviation) of both the child and parent reports at baseline and 24 months show a negative skewness, indicating a favoring of a higher HrQoL within the range of 0 to 100. No floor or ceiling effects were found at both measurement points and reports.

All QoLISSY scales of the child report had respectable Cronbach alpha values, ranging between α=0.70 (Emotional) and α=0.89 (Total Score) at baseline and between α=0.78 (Physical and Social) and α=0.93 (Total Score) at 24 months (Table 7). All QoLISSY scales of the parent report had respectable Cronbach alpha values, ranging between α=0.71 (Coping) and α=0.93 (Total Score) at baseline and between α=0.72 (Physical) and α=0.93 (Future and Total Score) at 24 months (Table 7).

3.3.2 Growth Hormone (GH)

With the exception of the Coping, Future, and Beliefs scales (parent-report only) and the Social scale (child-report only) at baseline and the Treatment scale (child-report only) and the Coping scale, scale characteristics (mean standard deviation) of both the child and parent reports both measurement points show a negative skewness, indicating a favoring of a higher HrQoL within the range of 0 to 100. A floor effect was found in the Beliefs scale of the child report at baseline measurement.

All QoLISSY scales of the child report had respectable Cronbach alpha values at both measurement points, ranging between α=0.70 (Coping) and α=0.94 (Total Score) at baseline and between α=0.72 (Coping) and α=0.95 (Treatment) at 24 months (Table 8). All QoLISSY scales of the parent report except the Coping scale at baseline measurement had respectable Cronbach alpha values, ranging between α=0.84 (Physical) and α=0.96 (Beliefs) at baseline and between α=0.79 (Social) and α=0.93 (Treatment) at 24 months (Table 8).

3.3.3 Combination Treatment (AI/GH)

With the exception of the Beliefs scales and the Effect on Parents scale (parent-report only) at baseline and the Coping scale (child-report only) and the Beliefs scale (parent-report only) at 24 months, scale characteristics (mean standard deviation) of both the child and parent reports at baseline and 24 months show a negative skewness, indicating a favoring of a higher HrQoL within the range of 0 to 100. Ceiling effects were found in the Beliefs and Physical scales (child report) at 24 months, as well as a ceiling effect in the Future scale (parent-report only).

With the exception of the Coping scale, all QoLISSY scales of the child report had respectable Cronbach alpha values, ranging between α=0.71 (Physical) and α=0.91 (Total Score) at baseline and between α=0.85 (Coping) and α=0.96 (Total Score) at 24 months (Table 9). All QoLISSY scales of the parent report except the Coping scale at baseline had respectable Cronbach alpha values, ranging between α=0.77 (Physical) and α=0.94 (Total Score) at baseline and α=0.79 (Coping) and α=0.97 (Total Score) at 24 months (Table 9).

Table 7: Descriptive Statistics and Reliability for Total Score and Subscales of AI Group at Baseline and 24 Months.

Descriptive Statistics and Reliability for Total Score and Subscales of AI Treatment type																			
		Baseline Domains (# of items)								24 months Domains (# of items)									
Descriptive Analyses	**Report Type**	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Future (5)**	Effect on Parents (11)**	Total Score (22)	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Treatment (14)	Future (5)**	Effect on Parents (11)**	Total Score (22)
n	*Child report*	24	24	24	24	24			24	16	16	16	16	16	10			16
	Parent report	24	23	24	22	23	23	23	23	16	16	16	16	16	10	16	16	16
Mean	*Child report*	67.71	66.80	70.05	45.46	52.60			68.19	73.54	69.34	71.76	41.09	58.98	48.86			71.55
	Parent report	56.91	51.67	50.74	48.47	42.12	50.00	58.51	53.17	74.48	70.03	63.09	45.97	53.52	56.63	60.94	66.34	69.20
SD	*Child report*	17.08	16.27	14.24	20.55	23.67			14.11	14.39	16.39	18.43	22.97	27.48	18.13			15.04
	Parent report	17.21	19.91	21.16	17.35	22.23	27.51	21.42	17.57	13.85	14.32	18.76	14.25	27.19	15.80	28.12	20.38	13.90
Skewness	*Child report*	-0.23	-0.64	-0.84	0.76	0.00			-0.55	-0.30	-0.28	-0.31	-0.29	0.07	-0.19			-0.42
	Parent report	0.42	-0.56	-0.07	1.09	0.59	0.34	-0.16	0.03	0.16	-0.20	-0.62	0.29	0.15	-0.05	-0.93	-0.40	0.22
Kurtosis	*Child report*	-1.07	-0.04	-0.01	1.06	-0.09			-0.52	0.31	-0.32	-0.67	-1.21	-1.17	-0.68			-0.05
	Parent report	-0.48	-0.08	-0.70	0.81	0.77	-0.41	-0.99	0.16	-0.43	-0.40	0.70	-0.13	0.01	-0.15	0.80	-0.30	-0.89
% Floor	*Child report*	0.0	0.0	0.0	0.0	4.2			0.0	0.0	0.0	0.0	6.3	0.0	0.0			0.0
	Parent Report	0.0	0.0	0.0	0.0	4.3	4.3	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	6.3	0.0	0.0
% Ceiling	*Child report*	0.0	0.0	0.0	4.2	4.2			0.0	6.3	0.0	0.0	0.0	12.5	0.0			0.0
	Parent report	0.0	0.0	0.0	0.0	0.0	4.3	0.0	0.0	6.3	0.0	0.0	0.0	12.5	0.0	6.3	0.0	0.0
Reliability Analyses*																		
α	*Child report*	**0.72**	**0.76**	**0.70**	**0.84**	**0.79**			**0.89**	**0.78**	**0.78**	**0.88**	**0.87**	**0.92**	**0.86**			**0.93**
	Parent report	**0.77**	**0.86**	**0.86**	**0.71**	**0.85**	**0.89**	**0.87**	**0.93**	**0.72**	**0.83**	**0.92**	**0.75**	**0.92**	**0.81**	**0.93**	**0.89**	**0.93**

*Only included in the parent report

Table 8: Descriptive Statistics and Reliability for Total Score and Subscales of GH Group at Baseline and 24 Months.

Descriptive Statistics and Reliability for Total Score and Subscales of GH Treatment type																			
		Baseline Domains (# of items)								24 months Domains (# of items)									
Descriptive Analyses	**Report Type**	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Future (5)**	Effect on Parents (11)**	Total Score (22)	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Treatment (14)	Future (5)**	Effect on Parents (11)**	Total Score (22)	
n	*Child report*	25	25	25	25	25			25	21	21	21	20	20	20			21	
	Parent report	25	25	25	25	25	25	25	25	21	21	21	21	21	20	21	21	21	
Mean	*Child report*	57.60	55.88	57.63	39.70	43.25			57.03	75.99	73.96	72.62	43.68	56.77	50.71			74.19	
	Parent report	53.00	46.27	45.00	41.07	42.50	47.40	49.36	48.09	75.87	72.34	66.71	48.82	58.04	64.04	71.19	68.04	71.64	
SD	*Child report*	21.26	20.75	24.64	15.72	31.19			20.47	16.61	14.93	16.30	16.04	29.59	23.55			13.41	
	Parent report	22.29	19.91	24.26	14.40	30.51	29.93	23.62	20.19	20.17	20.58	20.12	17.48	23.65	20.36	23.34	21.02	19.58	
Skewness	*Child report*	-0.49	0.35	-0.43	-0.03	-0.15			-0.22	-1.78	-0.84	-0.52	0.60	-0.52	0.04			-0.58	
	Parent report	-0.10	-0.11	-0.25	0.16	0.00	0.05	-0.08	-0.15	-1.74	-1.83	-1.29	0.25	-0.64	-1.21	-1.88	-1.38	-1.83	
Kurtosis	*Child report*	0.53	-0.43	0.11	1.20	-1.56			0.65	5.29	0.64	-0.59	-1.17	-0.38	-1.23			-0.44	
	Parent report	-0.86	-0.91	-0.78	-0.84	-1.64	-1.25	-1.01	-1.12	3.80	4.08	2.82	-1.33	0.39	2.47	4.30	1.49	4.37	
% Floor	*Child report*	0.0	0.0	4.0	4.0	**16.0**			0.0	0.0	0.0	0.0	0.0	10.0	0.0			0.0	
	Parent Report	0.0	0.0	4.0	0.0	12.0	8.0	0.0	0.0	0.0	0.0	0.0	0.0	4.8	0.0	4.8	0.0	0.0	
% Ceiling	*Child report*	0.0	0.0	4.0	0.0	0.0			0.0	0.0	0.0	0.0	0.0	5.0	0.0			0.0	
	Parent report	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.5	0.0	0.0	
Reliability Analyses*																			
α	*Child report*	**0.83**	**0.84**	**0.89**	**0.70**	**0.91**			**0.94**	**0.79**	**0.79**	**0.85**	**0.72**	**0.90**	**0.95**			**0.90**	
	Parent report	**0.84**	**0.86**	**0.90**	0.67	**0.96**	**0.93**	**0.92**	**0.95**	**0.92**	**0.79**	**0.88**	**0.83**	**0.89**	**0.93**	**0.92**	**0.92**	**0.92**	

*Only included in the parent report

Table 9: Descriptive Statistics and Reliability for Total Score and Subscales of AI/GH Group at Baseline and 24 Months.

Descriptive Statistics and Reliability for Total Score and Subscales of AI/GH Treatment type																			
		Baseline Domains (# of items)								24 months Domains (# of items)								
Descriptive Analyses	**Report Type**	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Belief (4)	Future (5)**	Effect on Parents (11)**	Total Score (22)	Physical (6)	Social (8)	Emotional (8)	Coping (10)	Beliefs (4)	Treatment (14)	Future (5)**	Effect on Parents (11)**	Total Score (22)
n	*Child report*	26	26	26	26	26			26	21	21	21	21	21	21			21
	Parent report	26	26	26	26	26	26	26	26	21	21	21	21	21	20	20	21	21
Mean	*Child report*	62.50	62.89	62.19	50.00	52.88			62.53	82.34	80.14	81.55	50.00	67.26	62.64			81.34
	Parent report	63.30	54.09	53.97	45.82	48.08	57.88	55.68	57.12	76.59	70.39	66.07	49.29	54.46	69.82	69.44	73.51	71.02
SD	*Child report*	16.58	19.42	20.86	15.08	25.99			15.87	16.40	20.25	19.41	25.09	31.62	18.18			17.46
	Parent report	15.77	21.97	18.62	11.11	25.11	21.69	19.89	17.43	21.99	25.40	22.37	18.58	26.97	19.08	28.38	23.54	22.51
Skewness	*Child report*	-0.78	-0.94	-0.81	-0.08	0.21			-0.96	-1.50	-1.61	-1.48	0.27	-0.68	-0.68			-1.43
	Parent report	-0.33	-0.43	-0.78	-0.15	0.18	-0.61	0.34	-0.51	-1.03	-1.09	-0.91	-0.42	0.08	-1.19	-0.69	-0.78	-1.16
Kurtosis	*Child report*	0.31	0.18	0.29	-0.66	-0.77			1.00	2.96	2.49	1.83	-0.44	-0.54	-0.55			1.57
	Parent report	-0.81	-0.34	0.16	-1.20	-0.86	0.84	-1.17	-0.25	0.01	0.65	0.68	1.75	-1.17	2.53	-0.51	-0.76	0.73
% Floor	*Child report*	0.0	0.0	0.0	0.0	0.0			0.0	0.0	0.0	0.0	0.0	4.8	0.0			0.0
	Parent Report	0.0	0.0	0.0	0.0	0.0	3.8	0.0	0.0	0.0	0.0	0.0	4.8	0.0	0.0	0.0	0.0	0.0
% Ceiling	*Child report*	0.0	0.0	0.0	0.0	7.7			0.0	**19.0**	9.5	14.3	4.8	**28.6**	0.0			4.8
	Parent report	0.0	0.0	0.0	0.0	3.8	3.8	0.0	0.0	9.5	9.5	4.8	0.0	9.5	5.0	**20.0**	4.8	4.8
Reliability Analyses*																		
α	*Child report*	**0.71**	**0.89**	**0.83**	0.68	**0.88**			**0.91**	**0.89**	**0.93**	**0.92**	**0.85**	**0.93**	**0.87**			**0.96**
	Parent report	**0.77**	**0.89**	**0.85**	0.57	**0.88**	**0.85**	**0.88**	**0.94**	**0.93**	**0.95**	**0.88**	**0.79**	**0.90**	**0.91**	**0.94**	**0.92**	**0.97**

*Only included in the parent report

3.4 Content Validity

The content validity of the QoLISSY instrument was evaluated based on the following criteria stated by Terwee et al. found in Table 10. In summary, the QoLISSY instrument fulfilled all criteria for good content validity—evidence for each criterion is listed in Table 10.

Table 10: Fulfillment of Criteria for Content Validity

Criteria for Content Validity	Fulfillment of Criterion (**Yes/No**); Evidence from QoLISSY Manual (The European QoLISSY Group, 2013b)
1. A clear statement of the measurement aim of the instrument.	**Yes**; "The QoLISSY Instrument can be used to assess the health-related quality of life in short stature youth regardless of its cause" (Pg. 10)
2. Clearly states which target population the instrument is intended to be used in	**Yes**; "It [the QoLISSY instrument] is usable for children and adolescents between the ages of 8 and 18 years and for parents of children with short stature between the ages of 4 and 18 years." (Pg. 10)
3. Definition of concepts that the instrument is meant to measure	**Yes**; "The QoLISSY instrument is based on the definition of health-related quality of life involving mental, social, and physical components of quality of life." (Pg. 31)"
4. Justification for item selection and reduction	**Yes**; Locus group methodology was used for item generation that is indeed relevant for children and adolescents with short stature (pg. 34). Pilot testing and cognitive debriefing of the pilot QoLISSY instrument was conducted to assess the psychometric performance of items, as well their relevance and understandability (pg. 48-68) in order for item selection and reduction for use in field testing. A field test and re-test with psychometric testing were conducted to assess the feasibility of the instrument, score distribution, validity of the instrument, and whether any changes to the instrument needed to be made (pg. 69).
5. Interpretability of items (the questionnaire itself should be short and simple and should not require reading skills beyond the age of 12 years old.)	**Yes**; The QoLISSY instrument is only 6 pages long and uses age-appropriate language Time to finish the questionnaire is between 10-50 minutes. (pg. 130-142).

3.5 Construct Validity

Construct validity was assessed by evaluating the convergent and discriminant validity of the QoLISSY instrument via Pearson Correlation analyses. The following subsections summarize the results of these analyses.

3.5.1 Convergent Validity: Pearson Correlations

3.5.1.1 KIDSCREEN-52 & QoLISSY Pearson Correlations

In the child report of the QoLISSY at baseline measurement, the Physical scale of the QoLISSY showed moderate to high correlations with only the Moods & Emotions (r= 0.50) and Autonomy (r=0.30) scale of the KIDSCREEN. The Social scale moderately correlated with the Moods & Emotions, Self Perception, and Autonomy scales of the KIDSCREEN instrument, and it correlated highly with the Bullying scale (r= 0.60). The Social scale also showed a significantly low correlation with the Parent scale of the KIDSCREEN (r= 0.24). The Emotional scale had a low to high correlation (ranging from r= 0.24 for Physical Wellbeing to r= 0.57 for Moods & Emotions) with all scales of the KIDSCREEN instrument, except for the Financial, Peers, and School scales. The Beliefs scale of the QoLISSY slightly or moderately correlated with the Moods & Emotions, Self Perception, and Autonomy scales. The Total Score of the QoLISSY correlated highly with the Moods & Emotions (r= 0.59) and Bullying (r= 0.50) scale, moderately with the Self Perception (r= 0.38) and Autonomy scales (r= 0.44), and lowly with the Parents (r= 0.27) scale of the KIDSCREEN instrument. The Social scale and the Bullying scale of the KIDSCREEN showed the highest correlation (r=0.60). The lowest significant correlations were found between the Social & Parents scale, as well as the Beliefs & Autonomy scale. Table 11 shows a complete summary of the Pearson correlations between the QoLISSY Children scales and the KIDSCREEN scales at baseline measurement.

In the child report of the QoLISSY at 24 months, the Physical scale of the QoLISSY showed no significant correlations with any KIDSCREEN scales. The Social scale moderately correlated with Moods and Emotions (r= 0.43) KIDSCREEN scale, and slightly correlated with the Self Perception scale (r= 0.27). The Emotional scale showed moderate to high correlations with the Moods and Emotions (r= 0.50), Self Perception scale (r= 0.33), and Bullying (r= 0.38) KIDSCREEN scales. The Coping scale moderately correlated with the Parents (r= 41), Peers (r= 0.33), and School (r= 0.31) scales, while the Beliefs scale was only moderately correlated with the

Moods and Emotions scale (r= 0.37). For the Treatment Effects scale, only the School scale of the KIDSCREEN showed a significant correlation (r= 0.30) with this QoLLISY scale. The Total Score scale of the QoLISSY showed moderate correlations with the Mood and Emotions (r= 0.41) and Bullying scales (r= 0.37), and a slight correlation with the Self Perception scale (r= 0.29) of the KIDSCREEN instrument. The highest significant correlation was found between the QoLISSY Emotional scale and KIDSCREEN Mood & Emotions scale (r=0.50), and the lowest being between the Social and the Self-Perception scale (r=0.27). Table 12 shows a complete summary of the Pearson correlations between the QoLISSY Children scales and the KIDSCREEN scales at 24 months.

Table 11: Scale Correlations of QoLISSY and KIDSCREEN at Baseline (Children Report).

QoLISSY Domains (# of items)	Pearson Correlation Coefficients of QoLISSY and KIDSCREEN Scales at Baseline (Child Report)									
	KIDSCREEN Domains (# of items)									
	Physical (5)	Psychological Wellbeing (6)	Moods & Emotions (7)	Self Perception (5)	Autonomy (5)	Parents (6)	Financial (3)	Peers (6)	School (6)	Bullying (3)
Physical (6)	-0.18	0.11	**0.50****	0.22	**0.30***	0.14	0.08	0.18	0.13	0.24
Social (8)	0.06	0.10	**0.49****	**0.34****	**0.39****	**0.24***	0.16	0.21	0.18	**0.60****
Emotional (8)	**0.24***	**0.30***	**0.57****	**0.44****	**0.47****	**0.34****	0.08	0.24	0.21	**0.48****
Coping (10)	0.23	0.11	0.09	-0.13	**0.28***	0.24*	0.03	0.24	0.03	0.10
Beliefs (4)	0.13	0.09	**0.36****	**0.28***	**0.24***	0.14	0.10	0.10	0.10	0.22
Total Score (22)	0.05	0.19	**0.59****	**0.38****	**0.44****	**0.27***	0.13	0.24	0.20	**0.50****

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

Table 12: Scale Correlations of QoLISSY and KIDSCREEN at 24 Months (Children Report).

QoLISSY Domains (# of items)	Pearson Correlation Coefficients of QoLISSY and KIDSCREEN Scales at 24-Months (Child Report)									
	KIDSCREEN Domains (# of items)									
	Physical (5)	Psychological Wellbeing (6)	Moods & Emotions (7)	Self Perception (5)	Autonomy (5)	Parents (6)	Financial (3)	Peers (6)	School (6)	Bullying (3)
Physical (6)	-0.14	-0.09	0.19	0.17	-0.01	-0.09	-0.18	-0.02	0.04	**0.32***
Social (8)	-0.04	-0.13	**0.43****	**0.27***	-0.04	-0.06	-0.10	-0.03	0.15	0.29
Emotional (8)	0.05	0.00	**0.50****	**0.33***	0.12	0.04	-0.17	0.07	0.20	**0.38***
Coping (10)	0.16	0.02	0.08	0.12	0.18	**0.41****	0.14	**0.33***	**0.31***	-0.04
Beliefs (4)	-0.04	-0.11	**0.37****	0.18	-0.08	-0.17	-0.18	-0.01	0.22	0.25
Treatment Effects (14)	-0.22	-0.19	0.11	-0.09	-0.07	-0.13	-0.14	0.13	**0.30***	-0.05
Total Score (22)	-0.05	-0.08	**0.41****	**0.29***	0.03	-0.04	-0.17	0.01	0.15	**0.37***

* Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

In the parent report of the QoLISSY at baseline measurement, the Physical scale slightly correlated with the Parents scale (r= 0.28), moderately correlated with the Psychological Well Being (r= 0.33), Mood & Emotions (r= 0.34), and Bullying (r= 0.40) scales, and highly correlated with the Self Perception scale (r= 0.55) of the KIDSCREEN instrument. The Social scale showed a moderate correlation with the Bullying scale (r= 0.42) and a high correlation with the Self Perception scale (r= 0.59). The Emotional scale of the QoLISSY instrument moderately correlated with the Psychological Well Being, Parents, School, and Bullying scales (correlations ranging from r= 0.33 to r= 0.43), and highly correlated with the Self Perception scale (r= 0.53) of the KIDSCREEN instrument. The Coping scale only correlated slightly with the Self Perception scale (r= 0.29), while the Beliefs scale had moderate correlations with the Mood & Emotions (r= 0.37) and Bullying (r= 0.45) scales and a high correlation with the Self Perception scale (r= 0.55). The Future scale moderately correlated with the Psychological Well Being, Moods & Emotions, and Bullying scales (correlations ranging from r= 0.35 to r= 0.39) and highly correlated with the Self Perception scale (r= 0.60) of the KIDSCREEN instrument. The Effect on Parents scale showed a moderate correlation with the Autonomy, Parents, and School scales (correlation ranging from r= 0.33 to r= 0.42), as well as high correlations with the Psychological Well Being, Moods & Emotions, Self Perception, and Bullying scales (ranging from r= 0.56 to r= 0.61). The Total Score scale slightly correlated with the School scale (r= 0.29), moderately correlated with the Psychological Well Being, Moods & Emotions, Parents, and Bullying scales (ranging from r= 0.31 to r= 0.45), and highly correlated with the Self Perception scale (r= 0.60) of the KIDSCREEN questionnaire. The highest significant correlation was found between the QoLISSY Future and Total Score scales and the KIDSCREEN Self-Perception scale (r=0.60), the lowest being between the Physical and the Parents scales (r=0.28). Table 13 shows a complete summary of the Pearson correlations between the QoLISSY Parent scales and the KIDSCREEN scales at baseline.

In the parent report of the QoLISSY instrument at 24 months, the Physical scale did not significantly correlate with any KIDSCREEN scale, with the exception of a moderate correlation with the Bullying scale (r= 0.39). The Social scale also only correlated moderately with the Self Perception (r= 0.37) and Bullying (r= 0.38) scales of the KIDSCREEN. The Emotional scale showed a moderate correlation with the Moods & Emotions scale (r= 0.33) and a high correlation with the Self Perception scale (r= 0.51). The Beliefs scale only showed a moderate correlation

with the Self Perception scale (r= 0.33), while the Effect on Parents scale had moderate correlations with the Psychological Well Being, Moods & Emotions, Self Perception, Parents, and Peers scales (ranging from r =0.39 and r= 0.49) of the KIDSCREEN instrument. The Total Score scale of the QoLISSY only showed moderate correlations with the Self Perception (r= 0.41) and Bullying (r= 0.38) scales of the KIDSCREEN. The Coping, Treatment Effects, and Future scales of the QoLISSY did not correlate significantly with any KIDSCREEN scales. The highest significant correlation was found between the QoLISSY Emotional scales and the KIDSCREEN Self-Perception scale (r=0.51), while the lowest was found between the Emotional and the Moods & Emotions scales (r=0.33).

Table 14 shows a complete summary of the Pearson correlations between the QoLISSY Parent scales and the KIDSCREEN scales at 24 months.

Table 13: Scale Correlations of QoLISSY and KIDSCREEN at baseline (Parent Report).

	Pearson Correlation Coefficients of QoLISSY and KIDSCREEN Scales at Baseline (Parent Report)									
QoLISSY Domains (# of items)	KIDSCREEN Domains (# of items)									
	Physical (5)	Psychological Wellbeing (6)	Moods & Emotions (7)	Self Perception (5)	Autonomy (5)	Parents (6)	Financial (3)	Peers (6)	School (6)	Bullying (3)
Physical (6)	0.06	**0.33***	**0.34***	**0.55****	0.12	**0.28***	0.11	0.00	0.21	**0.40****
Social (8)	0.00	0.22	0.20	**0.59****	0.05	0.23	0.10	-0.05	0.26	**0.42****
Emotional (8)	0.19	**0.40****	**0.33***	**0.53****	0.20	**0.36***	0.11	0.17	**0.33***	**0.43****
Coping (10)	0.14	0.15	0.13	**0.29***	0.24	0.17	-0.03	0.25	0.17	0.26
Beliefs (4)	0.05	0.27	**0.37****	**0.55****	0.12	0.15	-0.05	-0.02	0.24	**0.45****
Future (5)	0.00	**0.38****	**0.35***	**0.60****	0.06	0.27	0.04	0.10	0.27	**0.39****
Effect on Parents (11)	0.06	**0.57****	**0.56****	**0.61****	**0.33***	**0.42****	0.09	0.14	**0.34***	**0.57****
Total Score (22)	0.09	**0.34***	**0.31***	**0.60****	0.13	**0.32***	0.12	0.05	**0.29***	**0.45****

* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

Table 14: Scale Correlations of QoLISSY and KIDSCREEN at 24 months (Parent Report).

	Pearson Correlation Coefficients of QoLISSY and KIDSCREEN Scales at 24-Months (Parent Report)									
QoLISSY Domains (# of items)	KIDSCREEN Domains (# of items)									
	Physical (5)	Psychological Wellbeing (6)	Moods & Emotions (7)	Self Perception (5)	Autonomy (5)	Parents (6)	Financial (3)	Peers (6)	School (6)	Bullying (3)
Physical (6)	-0.12	0.19	0.20	0.29	-0.12	0.25	-0.05	0.22	0.20	0.**39***
Social (8)	-0.21	0.20	0.19	**0.37***	-0.04	0.31	0.01	0.27	0.24	**0.38***
Emotional (8)	-0.05	0.28	**0.33***	**0.51****	0.03	0.31	-0.03	0.29	0.23	0.31
Coping (10)	0.20	0.24	0.12	0.32	-0.15	0.14	-0.13	0.18	0.05	0.06
Beliefs (4)	-0.05	0.20	0.26	**0.33***	0.01	0.06	-0.09	0.24	0.14	0.22
Treatment Effects (14)	0.00	-0.01	0.14	0.19	-0.08	-0.04	0.00	-0.20	-0.15	0.01
Future (5)	-0.06	0.15	0.26	0.31	0.06	0.05	0.08	0.30	0.22	0.32
Effect on Parents (11)	0.15	**0.41***	**0.42****	**0.39***	0.26	**0.39***	0.02	**0.49****	0.24	0.28
Total Score (22)	-0.13	0.24	0.25	**0.41***	-0.05	0.30	-0.02	0.28	0.24	**0.38***

* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

3.5.1.2 Inter-Correlation of QoLISSY Scales (Child & Parent Report)

As an additional measure for convergent validity, the inter-correlation coefficients of the QoLISSY scales were analyzed.

In the children report at baseline, core scales of the QoLISSY instrument (Physical, Social, Emotional) show large-scale intercorrelations. Beliefs also show large intercorrelations with the Social, Emotional, and Total Score scales and a moderate intercorrelation with the Physical scale. The Total Scale score shows large-scale intercorrelations with Physical, Social, Emotional, and Beliefs. The Coping scale did not significantly intercorrelate with any QoLISSY scale (Table 15).

Table 15: InterScale Correlations of QoLISSY at Baseline (Children Report)

InterScale Correlations of QoLISSY at Baseline (Children Report)						
Child QoLISSY Domains (# of items)	Correlation Statistics	Child QoLISSY Domains (# of items)				
		Child Physical (6)	Child Social (8)	Child Emotional (8)	Child Coping (10)	Child Beliefs (4)
Child Social (8)	Pearson Correlation	**0.67****				
	Sig. (2-tailed)	0.00				
	N	75				
Child Emotional (8)	Pearson Correlation	**0.61****	**0.79****			
	Sig. (2-tailed)	0.00	0.00			
	N	75	75			
Child Coping (10)	Pearson Correlation	-0.03	-0.06	-0.04		
	Sig. (2-tailed)	0.83	0.58	0.73		
	N	75	75	75		
Child Beliefs (4)	Pearson Correlation	**0.44****	**0.58****	**0.67****	0.05	
	Sig. (2-tailed)	0.00	0.00	0.00	0.68	
	N	75	75	75	75	
Child Total Score (22)	Pearson Correlation	**0.84****	**0.92****	**0.91****	-0.05	**0.64****
	Sig. (2-tailed)	0.00	0.00	0.00	0.68	0.00
	N	75	75	75	75	75

** Correlation is significant at the 0.01 level (2-tailed)

In the parent report, core scales of the QoLISSY instrument (Physical, Social, Emotional) show large-scale intercorrelations. The Coping scale showed low scale correlations with the Social and Emotional scales. The Beliefs scale had large intercorrelation with the core scales. The Total Score shows large intercorrelation with all QoLISSY scales, with the exception of the Coping scale (r=0.31, moderate correlation) (Table 16).

Table 16: Inter Scale Correlations of QoLISSY at Baseline (Parent Report)

Inter Scale Correlations of QoLISSY at Baseline (Parent Report)							
Parent QoLISSY Scales (# of items)	Correlation Statistics	Parent QoLISSY Scales (# of items)					
		Parent Physical (6)	Parent Social (8)	Parent Emotional (8)	Parent Coping (10)	Parent Beliefs (4)	Parent Total Score (22)
Parent Physical (6)	Pearson Correlation						
	Sig. (2-tailed)						
	N						
Parent Social (8)	Pearson Correlation	**0.81****					
	Sig. (2-tailed)	0.00					
	N	74					
Parent Emotional (8)	Pearson Correlation	**0.66****	**0.76****				
	Sig. (2-tailed)	0.00	0.00				
	N	75	74				
Parent Coping (10)	Pearson Correlation	0.22	**0.27***	**0.36****			
	Sig. (2-tailed)	0.07	0.02	0.00			
	N	73	73	73			
Parent Beliefs (4)	Pearson Correlation	**0.52****	**0.59****	**0.66****	0.21		
	Sig. (2-tailed)	0.00	0.00	0.00	0.08		
	N	74	73	74	73		
Parent Total Score (22)	Pearson Correlation	**0.90****	**0.94****	**0.90****	**0.31****	**0.66****	
	Sig. (2-tailed)	0.00	0.00	0.00	0.01	0.00	
	N	74	74	74	73	73	

*Correlation is significant at the 0.01 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed)

In the children report at 24 months, core scales of the QoLISSY instrument (Physical, Social, Emotional) show large-scale intercorrelations. The Beliefs also show moderate intercorrelation with the core scales of the QoLISSY instrument. The Total Scale score shows large-scale intercorrelations with Physical, Social, Emotional, and Beliefs scales. The Coping and Treatment scales did not significantly intercorrelate with any QoLISSY scale (Table 17).

Table 17: Inter Scale Correlations of QoLISSY at 24 Months (Child Report)

InterScale Correlations of QoLISSY at 24 Months (Child Report)							
Child QoLISSY Domains (# of items)	Correlation Statistics	Child QoLISSY Domains (# of items)					
		Child Physical (6)	Child Social (8)	Child Emotional (8)	Child Coping (10)	Child Beliefs (4)	Child Treatment (14)
Child Social (8)	Pearson Correlation	**0.67****					
	Sig. (2-tailed)	0.00					
	N	58					
Child Emotional (8)	Pearson Correlation	**0.66****					
	Sig. (2-tailed)	0.00					
	N	58					
Child Coping (10)	Pearson Correlation	0.04	0.07	0.11			
	Sig. (2-tailed)	0.74	0.60	0.44			
	N	57	57	57			
Child Beliefs (4)	Pearson Correlation	**0.51****	**0.66****	**0.71****	-0.04		
	Sig. (2-tailed)	0.00	0.00	0.00	0.79		
	N	57	57	57	57		
Child Treatment (14)	Pearson Correlation	0.07	0.10	0.16	0.27	0.08	
	Sig. (2-tailed)	0.62	0.47	0.25	0.06	0.58	
	N	51	51	51	50	50	
Child Total Score (22)	Pearson Correlation	**0.85****	**0.93****	**0.93****	0.08	**0.70****	0.126
	Sig. (2-tailed)	0.00	0.00	0.00	0.54	0.00	0.38
	N	58	58	58	57	57	51

*Correlation is significant at the 0.01 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

In the parent report, core scales of the QoLISSY instrument (Physical, Social, Emotional) show large-scale intercorrelations. The Coping scale showed a low scale correlation with the Emotional scale. The Beliefs scale had moderate intercorrelation with the core scales. The Total Score shows large intercorrelation with all QoLISSY scales, except for the Coping and Treatment scales. The Coping and Treatment scales did not significantly intercorrelate with any QoLISSY scale (Table 18).

Table 18: Inter Scale Correlations of QoLISSY at 24 Months (Parent Report)

Inter Scale Correlations of QoLISSY at 24 Months (Parent Report)							
Parent QoLISSY Scales (# of items)	Correlation Statistics	Parent QoLISSY Scales (# of items)					
		Parent Physical (6)	Parent Social (8)	Parent Emotional (8)	Parent Coping (10)	Parent Beliefs (4)	Parent Treatment (14)
Parent Social (8)	Pearson Correlation	0.88**					
	Sig. (2-tailed)	0.00					
	N	58					
Parent Emotional (8)	Pearson Correlation	**0.83****	**0.87****				
	Sig. (2-tailed)	0.00	0.00				
	N	58	58				
Parent Coping (10)	Pearson Correlation	0.14	0.09	**0.26***			
	Sig. (2-tailed)	0.30	0.53	0.05			
	N	58	58	58			
Parent Beliefs (4)	Pearson Correlation	**0.63****	**0.68****	**0.68****	-0.06		
	Sig. (2-tailed)	0.00	0.00	0.00	0.65		
	N	58	58	58	58		
Parent Treatment (14)	Pearson Correlation	0.21	0.22	0.36*	0.18	0.11	
	Sig. (2-tailed)	0.14	0.13	0.01	0.21	0.46	
	N	50	50	50	50	50	
Parent Total Score (22)	Pearson Correlation	**0.95****	**0.96****	**0.95****	0.17	**0.70****	0.28
	Sig. (2-tailed)	0.00	0.00	0.00	0.20	0.00	0.05
	N	58	58	58	58	58	50

*Correlation is significant at the 0.01 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

3.5.2 Discriminant Validity (Known-Groups Validity): Multivariate Analyses

Pearson correlation analyses were conducted between the QoLISSY scales (for both reports and measurement points) and various demographic variables (eg. previous GH treatment, history of long-term disability or medical condition, height, and age) to identify possible covariates for use in multivariate analyses to evaluate the discriminant validity of the QoLISSY instrument.

Demographic variables were selected as covariates if they correlated significantly with 3 or more scales of the QoLISSY instrument.

Based on these Pearson correlation analyses, *previous GH treatment* was entered as a covariate into the model for data from the parent report at baseline. In the child report and parent report at 24 months, *height* was entered as a covariate into the model. After entering the appropriate covariates into the model, ANCOVA (for the Total Score scale) and MANCOVA (all other QoLISSY scales) analyses were conducted. Because no possible covariates were identified for baseline data from the child report, an ANOVA (for the Total Score scale) and MANOVA were used for this data.

Box M's Test for MANCOVAs and Levene's Test of Variances for ANCOVAS were used to evaluate the homogeneity of covariates (*height* and *previous GH treatment*) in the multivariate analyses and determine if results of the model are valid:

Previous GH treatment: After conducting the Box's M Test of homogeneity on parent-reported data at baseline, the significance level resulted in a value less than 0.05 for the MANCOVA analyses, which suggest that the assumptions of the model are not met and the results of this model may not be valid.

For the ANCOVA that used *previous GH treatment* as a covariate (Total Score of the parent report at baseline measurement), the Levene's Test of Variances resulted in a significant level more than 0.05 (p= 0.461), therefore homogeneity of the coefficient for the covariate can be assumed.

Height: Results of the Box's M Test and Levene's Test of Variances that has *height* entered as a covariable showed that the homogeneity of the coefficient for the covariates can be assumed.

Child Report: Using Pillai's trace, no significant multivariate effects of treatment type on QoLISSY scales in the children report were found at baseline (Pillai's trace = 0.165, $F_{(10, 138)}$ = 1.242, p = 0.270, η_p^2 = 0.083) and at 24 months (Pillai's trace = 0.206, $F_{(12, 86)}$ = 0.824, p = 0.625, η_p^2 = 0.103), which shows that HrQoL was not significantly different between treatment type from baseline to the 24-month measurement. Univariate analyses with all individual children QoLISSY scales also show no significant effects between the types of treatment on HrQoL at baseline (Table 19) and at 24 months (Table 20). In the ANOVA analyses for the Total Score of the children report, no significant effects of treatment type on the Total Score at both baseline ($F_{(2,761)}$=2.619, p=0.080; η_p^2=0.068) and at 24 months ($F_{(2, 258)}$=1.310, p=0.278; η_p^2=0.047) were found.

Parent report: Using Pillai's trace, no significant multivariate effect of treatment type on the HrQoL of the children was found in parent report at baseline (Pillai Trace=0.137; $F_{(14, 128)}$=0.674, p=0.796; η_p^2=0.069 and at 24 months (Pillai Trace=0.235; $F_{(16, 76)}$=0.632, p=0.848; η_p^2=0.117). Univariate analyses with all individual parent QoLISSY scales also show no significant effects between the types of treatment on HrQoL at baseline (Table 19) and at 24 months (Table 20). In the ANOVA analyses of the Total Score of the parent report, no significant effects of treatment type on the Total Score at both baseline ($F_{(2, 293)}$=0.930, p=0.399; η_p^2=0.026) and at 24 months ($F_{(2, 125)}$=0.538, p=0.587; η_p^2=0.020) were found.

Table 19: Score differences of the QoLISSY scales between the AI, GH, and AI/GH group (Child and Parent report) at baseline measurement.

Baseline		AI			GH			AI/GH			Statistics			
	Report	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**F**	**df**	**p**	**η_p^2**
Physical	*Child*	24	67.71	17.08	25	57.60	21.26	26	62.50	18.63	1.844	2	0.166	0.049
	Parent[a]	22	57.58	17.80	25	53.00	22.29	26	63.30	15.77	1.271	2	0.287	0.036
Social	*Child*	24	66.80	16.27	25	55.88	20.80	26	62.89	19.42	2.100	2	0.130	0.055
	Parent[a]	22	51.75	20.37	25	46.27	19.91	26	54.09	21.97	0.536	2	0.587	0.015
Emotional	*Child*	24	70.05	14.24	25	57.63	24.64	26	62.19	20.86	2.307	2	0.107	0.060
	Parent[a]	22	50.43	22.08	25	45.00	24.26	26	53.97	18.62	0.634	2	0.533	0.018
Coping	*Child*	24	45.46	20.55	25	39.70	15.71	26	50.00	15.08	2.289	2	0.109	0.060
	Parent[a]	22	48.47	17.35	25	41.10	14.40	26	45.82	11.11	1.476	2	0.236	0.041
Beliefs	*Child*	24	52.60	23.67	25	43.25	31.19	26	52.88	25.99	1.020	2	0.366	0.028
	Parent[a]	22	43.75	21.30	25	42.50	30.51	26	48.08	25.84	0.117	2	0.890	0.003
Future*	*Parent*[a]	22	52.27	25.85	25	47.40	29.93	26	57.88	21.69	0.595	2	0.554	0.017
Effect on Parents*	*Parent*[a]	22	59.00	21.79	25	49.36	23.62	26	55.68	19.89	0.974	2	0.383	0.027
Total Score	*Child*	24	68.19	14.11	25	57.03	20.47	26	62.51	17.42	2.619	2	0.080	0.068
	Parent	23	53.17	17.57	25	48.10	20.19	26	57.12	17.43	0.930	2	0.399	0.026

*Only included in the parent report of the QoLISSY instrument.

[a] Box M Test: p= 0.034; Variance homogeneity cannot be assumed.

Note: The following covariate was entered into the model: *previous GH (growth hormone) treatment* (parent scales).

Table 20: Score differences of the QoLISSY scales between the AI, GH, and AI/GH group (Child and Parent report) at 24 months.

24 Months		AI			GH			AI/GH			Statistics			
	Report	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**F**	**df**	**p**	**η_p^2**
Physical	*Child*	10	73.75	15.84	19	75.66	16.57	20	82.08	16.78	0.664	2	0.520	0.029
	Parent	10	76.25	12.89	20	76.42	20.54	18	78.24	22.17	0.316	2	0.731	0.014
Social	*Child*	10	71.88	15.66	19	73.85	15.60	20	80.87	20.50	0.779	2	0.465	0.033
	Parent	10	73.30	11.63	20	73.15	20.77	18	70.83	26.80	0.827	2	0.444	0.036
Emotional	*Child*	10	74.82	19.84	19	72.09	17.04	20	81.56	19.92	0.859	2	0.430	0.037
	Parent	10	65.63	19.71	20	66.92	20.62	18	67.01	23.12	0.475	2	0.625	0.021
Coping	*Child*	10	35.50	26.37	19	42.16	14.93	20	51.13	25.19	1.916	2	0.159	0.078
	Parent	10	47.50	13.84	20	49.88	17.23	18	49.31	19.92	0.099	2	0.906	0.004
Beliefs	*Child*	10	62.50	27.48	19	59.76	27.13	20	67.19	32.44	0.172	2	0.843	0.008
	Parent	10	58.13	27.17	20	57.81	24.24	18	55.90	28.23	0.376	2	0.689	0.017
Treatment	*Child*	10	48.86	18.13	19	51.03	24.15	20	61.85	18.27	1.464	2	0.242	0.061
	Parent	10	56.63	15.80	20	64.04	20.36	18	67.86	18.64	1.301	2	0.282	0.056
Future*	*Parent*	10	63.50	27.59	20	70.75	23.86	18	67.99	29.62	1.118	2	0.336	0.048
Effect on Parents*	*Parent*	10	65.23	22.24	20	68.07	21.57	18	75.66	22.12	0.757	2	0.475	0.033
Total Score	*Child*	16	71.55	15.04	21	74.19	13.41	20	81.51	17.90	1.310	2	0.278	0.047
	Parent	16	69.20	13.90	21	71.64	19.58	20	70.92	23.09	0.538	2	0.587	0.020

*Only included in the parent report of the QoLISSY instrument.
Note: The following covariates were entered into the model: *height* (parent and child scales).

3.6 Responsiveness (HrQoL Change Over Time)

To test the responsiveness of the QoLISSY instrument within the course of growth hormone treatment, repeated measures MANOVAs/ANOVAs were performed. Because the study design only included two measurement points, sphericity can be assumed.

Child Report: No significant interaction effect was found between treatment type and time in the MANOVA (Pillai-Trace=0.139; $F_{(10, 102)}$=0.764, *p*=0.662; η_p^2= 0.070) and ANOVA (Pillai-Trace=0.071; $F_{(2, 55)}$=2.106, *p*=0.131; η_p^2= 0.071). When visually interpreting the plot graphs for possible interaction between time and treatment type for each QoLISSY scale, the Physical, Belief, and Total Score scales initially appear to have interaction. However, these interactions were found to be non-significant in the univariate effect analyses (Table 21). However, significant multivariate main effects of time were found in the child report in the MANOVA (Pillai-Trace=0.431; $F_{(5, 50)}$=7.586; *p*=0.00; η_p^2=0.431) and ANOVA (Pillai-Trace=0.512; $F_{(1, 55)}$=28.14; *p*=0.00; η_p^2=0.338).

Parent Report: No significant interaction effect was found between treatment type and time in the MANOVA (Pillai-Trace=0.187; $F_{(14, 96)}$=0.187, *p*=0.763; η_p^2= 0.093) and ANOVA (Pillai-Trace=0.053; $F_{(2, 55)}$=1.527, *p*=0.226; η_p^2= 0.053). When visually interpreting the plot graphs for possible interaction between time and treatment type for each QoLISSY scale, Physical, Social, Emotional, Believe, Coping, Future, Effect on Parents, and Total Score scales initial show interactions, but these interactions were later found to be non-significant in the univariate effect analyses (Table 22). However, significant multivariate main effects of time were found in the parent report in the MANOVA (Pillai-Trace=0.613; $F_{(7, 47)}$=10.627; *p*=0.00; η_p^2=0.613) and ANOVA (Pillai-Trace=;0.504 $F_{(1, 55)}$=55.90, *p*=0.00; η_p^2= 0.504).

Table 21: Changes in health-related quality of life (HrQoL) from baseline to 24-month measurement (child report).

Responsiveness of QoLISSY: Changes of HrQoL from baseline to 24-month measurement (Child Report)*														
		AI			**GH**			**AI/GH**			**Statistics**			
	Time	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**F**	**df**	**p**	**η_p^2**
Physical	*Baseline*	16	66.41	14.63	20	57.42	23.38	21	65.08	15.51	2.414	2	0.099	0.082
	24 Month	16	73.54	14.39	20	76.67	16.74	21	82.34	16.40				
Social	*Baseline*	16	64.06	12.45	20	57.50	22.31	21	65.18	19.92	1.403	2	0.255	0.049
	24 Month	16	69.34	16.39	20	74.22	15.27	21	80.14	20.25				
Emotional	*Baseline*	16	67.77	13.39	20	57.34	27.17	21	64.20	21.58	1.556	2	0.220	0.054
	24 Month	16	71.76	18.43	20	72.50	16.71	21	81.55	19.41				
Coping	*Baseline*	16	44.29	15.26	20	38.88	17.20	21	49.29	16.45	0.564	2	0.572	0.020
	24 Month	16	41.09	22.97	20	43.68	16.04	21	50.00	25.09				
Belief	*Baseline*	16	56.64	18.47	20	41.88	32.89	21	57.74	26.51	0.826	2	0.443	0.030
	24 Month	16	58.98	27.48	20	56.77	29.59	21	67.26	31.62				
Total Score**	*Baseline*	16	66.08	11.32	20	57.42	22.65	21	64.82	15.71	1.412	2	0.252	0.049
	24 Month	16	71.55	15.04	21	57.84	22.16	21	64.82	15.71				

*The Treatment scale was not included because it was only assessed after 24 months of receiving treatment.
** Only the Total Score scale was analyzed using a repeated measurement ANOVA, while a repeated measurement MANOVA was used for the other scales.

Table 22: Changes in health-related quality of life (HrQoL) from baseline to 24-month measurement (parent report).

Responsiveness of QoLISSY: Changes of HrQoL from baseline to 24-month measurement (Parent Report)*														
		AI			**GH**			**AI/GH**			**Statistics**			
	Time	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**F**	**df**	**p**	**η_p^2**
Physical	*Baseline*	15	55.56	16.57	21	53.57	23.69	20	63.54	15.41	0.976	2	0.387	0.035
	24 Month	15	74.44	14.34	21	75.87	20.17	20	78.33	21.02				
Social	*Baseline*	15	50.48	19.23	21	47.05	19.58	20	53.13	23.23	0.594	2	0.594	0.019
	24 Month	15	71.99	12.41	21	72.34	20.58	20	71.56	25.47				
Emotional	*Baseline*	15	48.33	20.07	21	46.13	24.68	20	54.84	19.44	0.837	2	0.438	0.031
	24 Month	15	64.17	18.89	21	66.71	20.12	20	67.50	21.94				
Coping	*Baseline*	15	45.67	14.65	21	40.56	15.35	20	44.56	10.48	0.721	2	0.491	0.026
	24 Month	15	45.87	14.74	21	48.82	17.48	20	48.88	18.96				
Belief	*Baseline*	15	39.58	17.78	21	44.35	29.71	20	46.56	20.37	0.351	2	0.706	0.013
	24 Month	15	55.42	27.02	21	58.04	23.65	20	54.69	27.65				
Future	*Baseline*	15	50.00	24.13	21	48.57	29.84	20	56.00	22.80	0.900	2	0.413	0.033
	24 Month	15	62.00	28.77	21	71.19	23.34	20	69.44	28.38				
Effect on Parents	*Baseline*	15	59.56	19.47	21	49.13	24.60	20	56.14	20.20	1.700	2	0.192	0.060
	24 Month	15	67.12	20.85	21	68.04	21.02	20	74.68	23.51				
Total Score**	*Baseline*	16	51.46	16.17	21	48.91	20.53	21	58.00	18.10	0.426	2	0.655	0.015
	24 Month	15	70.20	13.77	21	71.64	19.58	20	72.47	22.07				

*The Treatment scale was not included because it was only assessed after 24 months of receiving treatment.
** Only the Total Score scale was analyzed using a repeated measurement ANOVA, while a repeated measurement MANOVA was used for the other scales.

3.7 Parent-Child (Dis)Agreement: Correlation Analyses

All parent and children HrQoL scores at baseline measurement had significant moderate or large correlation coefficients, with Physical and Social showing the lowest correlation (r=0.42) and Emotional showing the highest correlation (r=0.59) (Table 23). With the exception of the Coping scale (which was not significant) and the Beliefs scale (which had a significant low correlation coefficient), all parent and children HrQoL scores at 24 months had significant moderate or large correlation coefficients, with Social showing the lowest correlation (r=0.43) and Emotion and Total Score showing the largest correlation (r=0.58) (Table 24).

Table 23: Pearson Correlations of Parent and Child Scales of the QoLISSY at Baseline.

Parent-Child Scale Correlations of QoLISSY at Baseline						
Child QoLISSY Domains (# of items)	Parent QoLISSY Domains (# of items)					
	Parent Physical (6)	Parent Social (8)	Parent Emotional (8)	Parent Coping (10)	Parent Beliefs (4)	Parent Total Score (22)
Child Physical (6)	**0.42****	**0.44****	**0.46****	0.22	,291*	**0.50****
Child Social (8)	**0.30***	**0.42****	**0.50****	0.22	**0.27***	**0.45****
Child Emotional (8)	**0.35****	**0.42****	**0.59****	**0.27***	**0.43****	**0.50****
Child Coping (10)	0.20	0.12	0.16	**0.34****	**0.28***	0.17
Child Beliefs (4)	**0.37****	**0.41****	**0.56****	0.20	**0.47****	**0.50****
Total Score (22)	**0.40****	**0.48****	**0.58****	**0.27***	**0.37****	**0.54****

** Correlation is significant at the 0.01 level (2-tailed)

Table 24: Pearson Correlations of Parent and Child Scales of the QoLISSY at 24 Months.

Parent-Child Scale Correlations of QoLISSY at 24 Months (Children Report)							
Child QoLISSY Domains (# of items)	Parent QoLISSY Domains (# of items)						
	Parent Physical (6)	Parent Social (8)	Parent Emotional (8)	Parent Coping (10)	Parent Beliefs (4)	Parent Treatment Effects (14)	Parent Total Score (22)
Child Physical (6)	**0.54****	**0.54****	**0.62****	0.13	**0.50****	**0.47****	**0.60****
Child Social (8)	**0.36****	**0.43****	**0.55****	0.19	**0.42****	**0.20**	**0.47****
Child Emotional (8)	**0.41****	**0.46****	**0.58****	0.20	**0.46****	**0.28***	**0.51****
Child Coping (10)	0.18	0.10	0.11	0.21	-0.04	0.13	0.14
Child Beliefs (4)	0.20	0.24	**0.39****	0.18	**0.28***	**0.28***	**0.29***
Child Treatment Effects (14)	0.21	0.18	0.13	-0.10	0.10	**0.45****	0.18
Child Total Score (22)	**0.48****	**0.52****	**0.64****	0.20	**0.51****	**0.35***	**0.58****

** Correlation is significant at the 0.01 level (2-tailed)

4 Discussion

Regarding the psychometric aspects analyzed in this thesis, the QoLISSY instrument performed satisfactorily in terms of having good internal consistency, convergent validity (including inter-scale correlations), and content validity. This is also true for most scales in terms of skewness, floor and ceiling effects, and divergent validity.

4.1 Internal Consistency

In terms of internal consistency, most scales in each treatment type had good Cronbach's alpha values (defined as values over 0.70, taken from the suggestions of Terwee (2007) and Nunnally and Bernstein (1994), which implies that the items of the QoLISSY questionnaire of each subscale are correlated with the other subscales, and thus all measure the desired concept — HrQoL. Other studies exploring the psychometric performance of the QoLISSY instrument also showed good internal consistency for most, if not all, QoLISSY scales for both the parent and child report (Bullinger et al., 2013, 2014, 2015; J. Quitmann et al., 2017; Rohenkohl, Stalman, Kamp, Bullinger, & Quitmann, 2016; The European QoLISSY Group, 2013a).

However, this result was not observed in the Coping scale in the parent report at baseline measurement for all treatment types. This result may suggest that this scale does not fit well with the other scales of the parent version of the QoLISSY and may not be as relevant a concept for the measurement of HrQoL as the other scales. This is a novel result because the Coping scale in other validations studies of the QoLISSY instrument showed good Cronbach alpha values (generally over 0.80) for both the parent and child report (Bullinger et al., 2013, 2014, 2015; J. Quitmann et al., 2017; Rohenkohl et al., 2016; The European QoLISSY Group, 2013a).

4.2 Convergent Validity

After conducting Pearson correlations between the QoLISSY and KIDSCREEN instrument as one way to assess convergent validity, the scales of child report of the QoLISSY showed mostly moderate to high correlations with many of the KIDSCREEN scales. This result suggests that the QoLISSY instrument does indeed measure the concept, HrQoL.

An interesting anomaly to note in this convergent validity analysis to is that the Physical scale of both the child and parent report of the QoLISSY did not have significantly high correlations with the Physical scale of the KIDSCREEN at either measurement points. These results may be viewed as surprising because the Physical scale of both of these questionnaires should be quite similar (both should measure the physical aspects of HrQoL). The Physical scale of the QoLISSY instrument also showed low correlations to the Physical scale of the KIDSCREEN in the original QoLISSY study for both the child and parent report (Bullinger et al., 2013), as well as in other validation studies of the QoLISSY instrument (Bullinger et al., 2015; J. Quitmann et al., 2017; Rohenkohl et al., 2016).

Although these discrepancies may point toward limited construct validity of the QoLISSY, it may also simply be an indicator that the Physical scale of the QoLISSY captures specific problems related to short stature which the generic KIDSCREEN instrument does not.

Table 25 shows a side by side comparison of the items included in the Physical domain of the QoLISSY and KIDSCREEN questionnaires. When comparing the items between these two instruments, it is clear that even though both scales do deal with physical aspects of HrQoL, the item content of the QoLISSY instrument addresses specific problems that short stature children and adolescents may face, while the items in the KIDSCREEN Physical scale cover only general questions about the physical aspects of HrQoL. These differences in item content and the lack of significant correlations of the Physical scales of the QoLISSY and KIDSCREEN instruments further support the importance of disease-specific instruments because they can catch the unique challenges that affected people experience that a generic instrument cannot.

Table 25: Comparison of the Physical scale items of the QoLISSY and KIDSCREEN instruments

Comparison of Physical Scale Items	
QoLISSY	KIDSCREEN
My height prevents me from doing things that other children my age do	In general, how would you say your health is?
Because of my height I have problems everyday	Have you felt fit and well?
Because of my height I have more trouble reaching things than others my age	Have you been physically active?
Because of my height I depend others	Have you been able to run well?
Because of my height I depend others	Have you felt full of energy?
In sports I can keep up with others my age	
It bothers me that others my age can go on fairground rides and I can't	

The core scales of the QoLISSY also show large scale intercorrelations in both the child and parent report, which further supports the idea that they measure the same concept—that being the main domains of HrQoL. The Coping and Treatment scales were either not significantly intercorrelated with any other scale or only had low intercorrelations, which further suggests that these scales do not fit well with the other scales of the QoLISSY and may not be as relevant a concept for the measurement of HrQoL as the other scales. These results were also observed in the original QoLISSY study (Bullinger et al., 2013).

4.3 Content Validity

According to analyses of the development process of the QoLISSY instrument (information taken from the QoLISSY manual), the criteria for good content validity by Terwee et al. (2006) where fulfilled with adequate and persuasive evidence. This implies that the items in the QoLISSY instrument comprehensively represent the concept it intends to measure (HrQoL) and (according to the criteria list) that the development and selection processes of items in the QoLISSY are transparent with adequate arguments for why each item was included. The content validity analyses also confirm that the items of the QoLISSY are understandable, age-appropriate, and that they present a relatively little response burden with a possible completion time of 10 minutes.

4.4 Floor & Ceiling Effects

When analyzing for floor and ceiling effects at scale level, it appears that content validity may be limited in this study due to the presence of a floor effect in the Beliefs scale of the child report (GH Group at baseline measurement), ceiling effects in the Physical and Belief scales in the child report (AI/GH group at 24 months), a ceiling effect in the Future scale (AI/GH group at 24 months), and a ceiling effect in the Physical scale of the parent report (AI/GH group at 24 months).

These ceiling and floor effects may have occurred because the concepts that these scales are intended to capture may require more variance in items and answer option to comprehensively measure them. This may be especially true for the Beliefs scale — for example, the original European dataset also found ceiling effects (over 20%) in the Beliefs scale in both the child and parent report (The European QoLISSY Group, 2013a). Several other validation studies also reported a ceiling effect in the Beliefs scale for both the child and parent report of the QoLISSY instrument (Bullinger et al., 2013, 2015; Rohenkohl et al., 2016). Similar to the results of this study, the original European study of the QoLISY project, as well as other validation studies for the American-English and Danish versions of the QoLISSY also found ceiling effects in the Future scale (Bullinger et al., 2015; Rohenkohl et al., 2016; The European QoLISSY Group, 2013a).

However, because no floor or ceiling effects were found in the Total Score in this study, these floor and ceiling effects in the subscales could simply be indicators of the relative high opinion of both the children and their parents on the treatment the children received after experiencing it for a period (Lim et al., 2015). On the other hand, the presence of floor and ceiling effects in these subscales may also imply that when left as is, the items in these scales may not be appropriate enough to detect significant observation of change, thus limiting the responsiveness of the QoLISSY instrument (Terwee et al., 2007).

4.5 Skewness & Divergent Validity Analyses

When analyzing the descriptive statistics (skewness) of the QoLLISY instrument (for all treatment types), with a few exceptions, all treatment types report good HrQoL even at baseline measurement. These results are also supported by the divergent validity (known-groups validity) analyses of the QoLISSY instrument, which showed that the QoLISSY scale scores did not significantly vary by treatment type even after controlling for possible influencing factors like

height and *previous GH treatment* at both measurement points. Although the QoLISSY instrument in this study did not show good divergent validity, these results are unsurprising, when considering the fact that the sample itself did not vary significantly in terms of height even before starting treatment.

On the other hand, the Coping scale often revealed a positive skewness in the descriptive statistical analyses (see Table 6, 7, 8, and 9). When disregarding the Coping scale's poor performance in the internal consistency and convergent validity analyses, the positive skewness found in this scale suggests lower HrQoL. If this is true, then it may be an indicator that this patient group has problems coping with their short stature and may benefit from psychosocial group interventions to promote usage of coping strategies (Rohenkohl et al., 2016).

However, because this scale did not perform well in the internal consistency and convergent validity analyses, this suggests that the Coping scale may not be as relevant a concept for the measurement of HrQoL as the other scales. Therefore, the positive skewness found in this scale may not be an indicator of poor HrQoL. Instead, this result may suggest that the study participants simply do not bother the participants as much as it has been thought to (Attanasio et al., 2005; Blum et al., 2003). Another explanation for this diverging result may be because of the overall age of the sample — boys in their teenage years — who may have developed sufficient coping mechanisms and resources that mitigate the negative effects of short stature on their HrQoL.

Table 26: Items of the Coping Scale

Sometimes things might not be easy for you. Please tell what you think of or do to feel better.	I tell myself it is okay to be short.
	I try to get used to my height.
	If others tease me I stand up for myself.
	If others tease me my friends stand up for me.
	If others tease me I try to talk to them.
	If I feel bad about my height I spend time with my friends.
	If I feel bad about my height I try to think about things I am good at.
	If I feel bad about my height I talk about it to family and or friends.
	If I feel bad about my height I try to forget about it.

4.6 Responsiveness

In the repeated measurement MANOVA/ANOVA analyses used to measure the responsiveness of the QoLISSY instrument, no significant interaction effects were found between treatment type and time for both the child and parent report of the QoLISSY instrument (Table 20) despite initial plot graphs showing an interaction between these variables. However significant multivariate main effects of time were found in both the child and parent reports. These results suggest the QoLISSY can detect significant changes of HrQoL between baseline and 24-month, even though these changes only appear to be due to the passage of time and not necessarily as a result of receiving GH treatment.

These results may be due to the inherent limitation of the study's design—namely, the lack of a control group that did not receive growth treatment for comparison. Although a control group would have been a valuable component to the study, a control group was not included in the study design because it would be unethical to have purposefully deprived some boys the chance to grow taller, especially since there is already evidence that GH therapy and AI therapy do induce growth

in adolescents (Cohen et al., 2008; Dahlgren, 2011; Geffner, 2009; Hero, 2016; McGrath & O'Grady, 2015; Ranke et al., 2007; Richmond & Rogol, 2016).

However, it should be noted that the method used to analyze the responsiveness of the QoLISSY instrument is not one of the more "recommended" methods that other studies use in PRO research — one of which is calculating the minimally important difference (MID), which is also sometimes called minimally clinically important difference (MCID). There is a variety of methods to calculate MIDs/MCIDs that fall into two categories: distribution-based and anchor-based approaches (Dawson, Doll, Coffey, & Jenkinson, 2007). Anchor-based approaches are the preferred method in PRO research (Fayers & Hays, 2014), which involves using an external criterion as a measure stick for what a MCID should be (Dawson et al., 2007; Fayers & Hays, 2014).

Examples of external criterion include respondent's definition of what constitutes as a meaningful change, pre-determined cut-off points for pertinent clinical endpoints, etc. (Dawson et al., 2007). However, in PRO research, the method currently used most often in studies and is advocated for is asking respondents to take an initial assessment and a later assessment of whether a global change has occurred and if so, to what extent (Fayers & Hays, 2014; McLeod, Coon, Martin, Fehnel, & Hays, 2011; Nixon, Doll, Kerr, Burge, & Naegeli, 2016).

Nixon et al. further differentiate between two types of global ratings of change that can be assessed: global ratings of change (according to Fayers et al., also called "transition rating") and global ratings of concept. The main difference between the two is that the latter requires respondents to provide a global rating of their current state at each assessment point (ex. baseline and 24-months after intervention), whereas global ratings of change require respondents to recall how their state was at one measurement point to their state in the present (Nixon et al., 2016). For both the global rating of concept and change, the change of these ratings between measurement points are then calculated and analyzed.

Taking the study by Nixon et al., they developed examples for global ratings of concept and change items to assess mobility (as a core subcategory under the overall concept studied: physical functioning):

Global rating of concept item: "Overall, how much difficulty have you had with physical functioning (ex. walking, eating, walking stairs) *in the last week* as a result of x condition?" where

possible answer options would be “no difficulty,” “little difficulty,” some difficulty,” moderate difficulty,” and “severe difficulty.”

Global rating of change item: “Overall, how has your physical functioning (ex. walking, eating, walking stairs) as a results of x condition changed *since your last visit/last completion of this questionnaire*?” with the following possible answer options “much worse,” a little worse,” “no change,” “little better” and “much better.”

Because of the higher chance of recall bias when using global ratings of change and its unsuitableness for use of retrospective health assessments on health conditions that have high symptom variability over short periods of time, use of global ratings of concept items is recommended in PRO research (McLeod et al., 2011; Nixon et al., 2016; Wyrwich, Norquist, Lenderking, & Acaster, 2013). However, this more widely used and recommended anchor-method is currently not possible with the QoLISSY instrument. This is because there is no global rating of change or concept item that requires respondents to provide a global rating of their current state.

Although there is still an ongoing debate on how to measure responsiveness and no gold standard process exists, it would be perhaps beneficial for further item development (specifically for global ratings of concept items) to be conducted for the QoLISSY instrument so that responsiveness can be calculated and analyzed in a more “standardized” way that allows easy comparison of results with other similar studies.

4.7 Parent-Child (Dis)Agreement

The results of this study show that parent-reported HrQoL has significant moderate to large correlations to children-reported HrQoL at baseline measurement. Most importantly, this trend is true for the core domains of HrQoL (Physical, Beliefs, and Social). Apart from the Coping and Beliefs scale, this trend was also true at the 24-month measurement. These results may suggest that although parent-reported HrQoL data of their children may reflect their children’s own judgement of their HrQoL, the parent-report still shows disagreement after their children have undergone treatment. This could imply that self-reported data may describe a more accurate picture of post-treatment HrQoL changes than a parent-report does, especially in acquiring information about coping strategies and changes of beliefs of short stature.

When comparing these results with the descriptive statistics of the parent report (see Tables 7, 8, 9, and 10), worse HrQoL scores in the Beliefs scale and better scores for the Coping scale were

reported by parents than the scores that the children reported themselves for both measurement points. Also, the mean scores of the parent report (for most scales for all treatment types at both measurement points) were lower than the mean scores from the child report. Although the Pearson correlation analyses show that many of the children and parent scales do show good correlation values, these results may show support to the idea that parents/proxy reports tend to report worse health outcomes than what their children themselves would report (Balen, Sinnema, & Geenen, 2006; Bullinger et al., 2009; N. Silva, Crespo, Carona, Bullinger, & Canavarro, 2015).

The results of this study also provide support for the idea that agreement between proxy and self-reports are highest in areas of HrQoL that are easily observable (ex. physical aspects and less so for social aspects) (Eiser & Morse, 2001; Rajmil, Lopez, Lopez-Aguila, & Alonso, 2013). Although self-reported data about HrQoL has been regarded to be a superior and more accurate source of information (Cremeens, Eiser, & Blades, 2006; Julia Quitmann et al., 2016), many studies investigating the agreement between these reports have suggested analyzing both reports because it may bring a better understanding of the disease experience and its impact on quality of life on the patient. (Julia Quitmann et al., 2016; David E. Sandberg et al., 2004; Vetter, Bridgewater, & McGwin, 2012).

5 Conclusions

The results presented in this thesis support the ideas that (1) PRO research in pediatric populations require both the self and a proxy report to provide a comprehensive view of HrQoL (2) disease-specific HrQoL instrument do indeed capture unique problems that affected persons face, which general HrQoL instrument cannot measure, and (3) the QoLISSY is a psychometrically-sound instrument.

In terms of the psychometric performance, the QoLISSY instrument, with the exception of the Coping scale, was similar to what has been observed in the original QoLISSY study and in other validation studies. The QoLISSY in this study showed good, internal consistency, convergent validity (including inter-scale correlations), and content validity. This is also true for most scales of the QoLISSY scale in terms of skewness, floor and ceiling effects, and divergent validity.

The results of this thesis suggest the QoLISSY instrument can detect significant changes of HrQoL between baseline and 24-months. However, the overall verdict of the responsiveness of the

QoLISSY questionnaire is inconclusive because it was not analyzed using a more "standardized" method, namely calculating the MID (minimally important difference), due to the QoLISSY instrument's lack of global rating of change or concept item that require respondents to provide a global rating of their current state.

Although significant HrQoL differences were found between baseline and 24-months, these differences appear to be the result of the passage of time and not necessarily because the treatment received. This result may have been influenced by the lack of a control group in the study's design. However, a prospective observational study that did have a control group (ISS diagnosed children who received no GH treatment) and used the QoLISSY questionnaire to measure HrQoL before and after receiving growth treatment found significant interaction effects between time and treatment type (received GH treatment or untreated) reported increased scores in the Physical, Social, and Emotional scales between the baseline and 12-month measurement point in the child-report (Julia Quitmann et al., 2018). Therefore, the QoLISSY instrument is indeed responsive enough to detect HrQoL changes in study settings that include a control group in its design.

Regardless of the whether the improvement of HrQoL was due to time or treatment, this thesis found that the group that received the combination therapy (GH and AI) reported higher HrQoL in all scales than the other two treatment types for both the child and parent report. This result may suggest that the use of combination therapy may indeed be a favorable treatment option for short stature that not only increases height (Mauras et al. 2016), but also increase affected patients' HrQoL—or at the very least, it does not decrease their HrQoL.

Despite these promising results, the limitations of this study should also be considered. One such limitation is the lack of a control group that would have not received treatment. However, this aspect in the study design was inevitable due to the ethnical problems that would have arose if some study participants did not receive treatment, and thus be denied the opportunity to grow taller. The lack of a "global rating of HrQoL" item in the QoLISSY instrument may also be seen as a limitation because it prevents the calculation of more "standardized" methods of assessing responsiveness, which may affect the comparability of the results with other studies.

Another possible limitation to the study may be the selected age group (between 12 and 18 years old) because it is uncommon for older adolescents (>13 years old) to be treated with GH due to it being more effective when children have not yet gone through puberty (Richmond & Rogol,

2016). Therefore, the chance of a simple developmental delay in growth cannot be completely be ruled out for the children who received GH.

Despite these possible study limitations, the QoLISSY questionnaire appears to be a psychometrically sound instrument that could be used as a treatment outcome indicator. However, further improvement of the questionnaire should be explored, such as developing a global rating of HrQoL change or concept item to allow ease of comparison with other studies that strive to explore the responsiveness of the QoLISSY instrument in intervention studies. Although the results of this thesis show that the core scales of the QoLISSY questionnaire (Physical, Social, and Emotional) and the Total Score scale are psychometrically robust scales, the psychometric performance of other non-core scales (ex. the Coping, Beliefs, and Treatment) should be further scrutinized in future studies. However, despite not performing psychometrically as well as the core-scales of the QoLISSY instrument, the Coping, Beliefs, and Treatment scales provide valuable, complementary information about the experience of short stature, perhaps allowing healthcare givers to become aware if extra support (ex. psychological intervention) should be provided.

In conclusion, the QoLISSY instrument can be used to track HrQoL changes over time, explore the experiences associated with short stature (and its treatment) through the perspectives of both the patient and his or her parents, and to highlight areas in life (ex. well-being and functioning) of short-statured children and adolescents that can be improved through intervention.

6 References

Abe, S., Okumura, A., Mukae, T., Nakazawa, T., Niijima, S.-I., Yamashiro, Y., & Shimizu, T. (2009). Depressive tendency in children with growth hormone deficiency. *Journal of Paediatrics and Child Health*, *45*(11), 636–640. https://doi.org/10.1111/j.1440-1754.2009.01586.x

Alonso, J., Bartlett, S. J., Rose, M., Aaronson, N. K., Chaplin, J. E., Efficace, F., … Forrest, C. B. (2013). The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. *Health and Quality of Life Outcomes*, *11*(1), 210. https://doi.org/10.1186/1477-7525-11-210

Attanasio, A. F., Shavrikova, E. P., Blum, W. F., & Shalet, S. M. (2005). Quality of life in childhood onset growth hormone-deficient patients in the transition phase from childhood to adulthood. *The Journal of Clinical Endocrinology and Metabolism*, *90*(8), 4525–4529. https://doi.org/10.1210/jc.2005-0439

August, G. P., Lippe, B. M., Blethen, S. L., Rosenfeld, R. G., Seelig, S. A., Johanson, A. J., … Sherman, B. M. (1990). Growth hormone treatment in the United States: demographic and diagnostic features of 2331 children. *The Journal of Pediatrics*, *116*(6), 899–903.

Balen, H. V., Sinnema, G., & Geenen, R. (2006). Growing up with idiopathic short stature: psychosocial development and hormone treatment; a critical review. *Archives of Disease in Childhood*, *91*(5), 433–439. https://doi.org/10.1136/adc.2005.086942

Blum, W. F., Shavrikova, E. P., Edwards, D. J., Rosilio, M., Hartman, M. L., Marin, F., … Herschbach, P. (2003). Decreased quality of life in adult patients with growth hormone deficiency compared with general populations using the new, validated, self-weighted questionnaire, questions on life satisfaction hypopituitarism module. *The Journal of*

Clinical Endocrinology and Metabolism, *88*(9), 4158–4167. https://doi.org/10.1210/jc.2002-021792

Borjesson, A. E., Lagerquist, M. K., Windahl, S. H., & Ohlsson, C. (2013). The role of estrogen receptor alpha in the regulation of bone and growth plate cartilage. *Cellular and Molecular Life Sciences : CMLS*, *70*(21), 4023–4037. https://doi.org/10.1007/s00018-013-1317-1

Brutt, A. L., Sandberg, D. E., Chaplin, J., Wollmann, H., Noeker, M., Koltowska-Haggstrom, M., & Bullinger, M. (2009). Assessment of health-related quality of life and patient satisfaction in children and adolescents with growth hormone deficiency or idiopathic short stature - part 1: a critical evaluation of available tools. *Hormone Research*, *72*(2), 65–73. https://doi.org/10.1159/000232158

Bullinger, M. (2002). Assessing health quality life in medicine. An overview over concepts, methods, and applications in international research. *Restorative Neurology and Nueroscience*, *20*, 93–101.

Bullinger, M., Koltowska-Haggstrom, M., Sandberg, D., Chaplin, J., Wollmann, H., Noeker, M., & Brutt, A. L. (2009). Health-related quality of life of children and adolescents with growth hormone deficiency or idiopathic short stature - part 2: available results and future directions. *Hormone Research*, *72*(2), 74–81. https://doi.org/10.1159/000232159

Bullinger, M., Quitmann, J., Power, M., Herdman, M., Mimoun, E., DeBusk, K., … Chaplin, J. E. (2013). Assessing the quality of life of health-referred children and adolescents with short stature: development and psychometric testing of the QoLISSY instrument. *Health and Quality of Life Outcomes*, *11*, 76. https://doi.org/10.1186/1477-7525-11-76

Bullinger, M., Quitmann, J., Silva, N., Rohenkohl, A., Chaplin, J. E., DeBusk, K., … Power, M. (2014). Cross-cultural equivalence of the patient- and parent-reported quality of life in

short stature youth (QoLISSY) questionnaire. *Hormone Research in Paediatrics*, *82*(1), 18–30. https://doi.org/10.1159/000358832

Bullinger, M., Sommer, R., Pleil, A., Mauras, N., Ross, J., Newfield, R., … Quitmann, J. (2015). Evaluation of the American-English Quality of Life in Short Stature Youth (QoLISSY) questionnaire in the United States. *Health and Quality of Life Outcomes*, *13*, 43. https://doi.org/10.1186/s12955-015-0236-2

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., … PROMIS Cooperative Group. (2007). The Patient-Reported Outcomes Measurement Information System: (PROMIS) Progress of an NIH Roadmap Cooperative Group During its First Two Years. *Medical Care*, *45*(5), S3–S11. https://doi.org/10.1097/01.mlr.0000258615.42478.55

Cohen, P., Rogol, A. D., Deal, C. L., Saenger, P., Reiter, E. O., Ross, J. L., … Wit, J. M. (2008). Consensus statement on the diagnosis and treatment of children with idiopathic short stature: a summary of the Growth Hormone Research Society, the Lawson Wilkins Pediatric Endocrine Society, and the European Society for Paediatric Endocrinology Workshop. *The Journal of Clinical Endocrinology and Metabolism*, *93*(11), 4210–4217. https://doi.org/10.1210/jc.2008-0509

Cook, C. E. (2008). Clinimetrics Corner: The Minimal Clinically Important Change Score (MCID): A Necessary Pretense. *The Journal of Manual & Manipulative Therapy*, *16*(4), E82-83. https://doi.org/10.1179/jmt.2008.16.4.82E

Coons, S. J., Rao, S., Keininger, D. L., & Hays, R. D. (2000). A comparative review of generic quality-of-life instruments. *PharmacoEconomics*, *17*(1), 13–35.

Cremeens, J., Eiser, C., & Blades, M. (2006). Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Quality of Life Research :*

An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation, *15*(4), 739–754. https://doi.org/10.1007/s11136-005-4184-x

Dahlgren, J. (2011). Growth outcomes in individuals with idiopathic short stature treated with growth hormone therapy. *Hormone Research in Paediatrics*, *76 Suppl 3*, 42–45. https://doi.org/10.1159/000330158

Dawson, J., Doll, H., Coffey, J., & Jenkinson, C. (2007). Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis and Cartilage*, *15*(8), 918–931. https://doi.org/10.1016/j.joca.2007.02.003

DeSalvo, K., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality Prediction with a Single General Self-Rated Health Question: A Meta-Analysis. *J Gen Intern Med*, *21*(3), 267–275.

Deshpande, P. R., Rajan, S., Sudeepthi, B. L., & Abdul Nazir, C. P. (2011). Patient-reported outcomes: A new era in clinical research. *Perspectives in Clinical Research*, *2*(4), 137–144. https://doi.org/10.4103/2229-3485.86879

Eiser, C., & Morse, R. (2001). Can parents rate their child's health-related quality of life? Results of a systematic review. *Quality of Life Research*, *10*(4), 347–457.

Emons, J., Chagin, A. S., Savendahl, L., Karperien, M., & Wit, J. M. (2011). Mechanisms of growth plate maturation and epiphyseal fusion. *Hormone Research in Paediatrics*, *75*(6), 383–391. https://doi.org/10.1159/000327788

Eshet, R., Maor, G., Ben Ari, T., Ben Eliezer, M., Gat-Yablonski, G., & Phillip, M. (2004). The aromatase inhibitor letrozole increases epiphyseal growth plate height and tibial length in peripubertal male mice. *The Journal of Endocrinology*, *182*(1), 165–172.

Fayers, P. M., & Hays, R. D. (2014). Don't middle your MIDs: regression to the mean shrinks estimates of minimally important differences. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *23*(1), 1–4. https://doi.org/10.1007/s11136-013-0443-4

Ferris, J. A., & Geffner, M. E. (2017). Are aromatase inhibitors in boys with predicted short stature and/or rapidly advancing bone age effective and safe? *Journal of Pediatric Endocrinology & Metabolism : JPEM*, *30*(3), 311–317. https://doi.org/10.1515/jpem-2016-0219

Fields, A. (2009a). Criterion validity. In *Discovering Statistics Using SPSS* (3rd ed., p. 784). London: SAGE Publications Ltd.

Fields, A. (2009b). Interpreting Cronbach's α (some cautionary tales …). In *Discovering Statistics Using SPSS* (3rd ed., pp. 675–676). London: SAGE Publications Ltd.

Fields, A. (2009c). Validity and reliability. In *Discovering Statistics Using SPSS* (3rd ed., pp. 11–12). London: SAGE Publications Ltd.

Geffner, M. E. (2009). Aromatase inhibitors to augment height: continued caution and study required. *Journal of Clinical Research in Pediatric Endocrinology*, *1*(6), 256–261. https://doi.org/10.4274/jcrpe.v1i6.256

Geisler, A., Lass, N., Reinsch, N., Uysal, Y., Singer, V., Ravens-Sieberer, U., & Reinehr, T. (2012). Quality of life in children and adolescents with growth hormone deficiency: association with growth hormone treatment. *Hormone Research in Paediatrics*, *78*(2), 94–99. https://doi.org/10.1159/000341151

GH Research Society. (2000). Consensus guidelines for the diagnosis and treatment of growth hormone (GH) deficiency in childhood and adolescence: summary statement of the GH

Research Society. GH Research Society. *The Journal of Clinical Endocrinology and Metabolism*, *85*(11), 3990–3993. https://doi.org/10.1210/jcem.85.11.6984

Grimberg, A., DiVall, S. A., Polychronakos, C., Allen, D. B., Cohen, L. E., Quintos, J. B., … Murad, M. H. (2016). Guidelines for Growth Hormone and Insulin-Like Growth Factor-I Treatment in Children and Adolescents: Growth Hormone Deficiency, Idiopathic Short Stature, and Primary Insulin-Like Growth Factor-I Deficiency. *Hormone Research in Paediatrics*, *86*(6), 361–397. https://doi.org/10.1159/000452150

Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, *118*(8), 622–629.

Hero, M. (2016). Aromatase Inhibitors in the Treatment of Short Stature. *Endocrine Development*, *30*, 130–140. https://doi.org/10.1159/000439338

Hess, R. A. (2003). Estrogen in the adult male reproductive tract: A review. *Reproductive Biology and Endocrinology : RB&E*, *1*, 52. https://doi.org/10.1186/1477-7827-1-52

Higginson, I., & Carr, A. (2001). Using quality of life measures in the clinical setting. *BMJ*, *322*, 1297–1300.

Kaplan, R., & Ries, A. (2007). Quality of life: concept and definition. *Journal of Chronic Obstructive Pulmonary Disease*, *4*, 263–271.

KIDSCREEN Group Europe. (2006). *The KIDSCREEN questionnaires: Quality of life questionnaires for children and adolescents, a handbook*. Lengerich, Germany: Pabst Science Publishers.

Kosinski, M., Zhao, S. Z., Dedhiya, S., Osterhaus, J. T., & Ware, J. E. J. (2000). Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis and Rheumatism*, *43*(7),

1478–1487. https://doi.org/10.1002/1529-0131(200007)43:7<1478::AID-ANR10>3.0.CO;2-M

Lim, C. R., Harris, K., Dawson, J., Beard, D. J., Fitzpatrick, R., & Price, A. J. (2015). Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open*, *5*(7), e007765. https://doi.org/10.1136/bmjopen-2015-007765

Mackie, E. J., Ahmed, Y. A., Tatarczuch, L., Chen, K.-S., & Mirams, M. (2008). Endochondral ossification: how cartilage is converted into bone in the developing skeleton. *The International Journal of Biochemistry & Cell Biology*, *40*(1), 46–62. https://doi.org/10.1016/j.biocel.2007.06.009

Matza, L. S., Swensen, A. R., Flood, E. M., Secnik, K., & Leidy, N. K. (2004). Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *7*(1), 79–92. https://doi.org/10.1111/j.1524-4733.2004.71273.x

Mauras, N., Ross, J. L., Gagliardi, P., Yu, Y. M., Hossain, J., Permuy, J., … Bullinger, M. (in preparation). *Growth Hormone & Aromatase Inhibitors Positively Impact Quality of Life in Short Stature Youth*. Manuscript in preparation.

Mauras, N., Ross, J. L., Gagliardi, P., Yu, Y. M., Hossain, J., Permuy, J., … Mericq, V. (2016). Randomized Trial of Aromatase Inhibitors, Growth Hormone, or Combination in Pubertal Boys with Idiopathic, Short Stature. *The Journal of Clinical Endocrinology and Metabolism*, *101*(12), 4984–4993. https://doi.org/10.1210/jc.2016-2891

McGrath, N., & O'Grady, M. J. (2015). Aromatase inhibitors for short stature in male children and adolescents. *The Cochrane Database of Systematic Reviews*, (10), CD010888. https://doi.org/10.1002/14651858.CD010888.pub2

McHorney, C. A. (1999). Health status assessment methods for adults: past accomplishments and future challenges. *Annu Rev Public Health*, *20*. https://doi.org/10.1146/annurev.publhealth.20.1.309

McLeod, L. D., Coon, C. D., Martin, S. A., Fehnel, S. E., & Hays, R. D. (2011). Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Review of Pharmacoeconomics & Outcomes Research*, *11*(2), 163–169. https://doi.org/10.1586/erp.11.12

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi*, *24*(3), 69–71.

Nilsson, O., Weise, M., Landman, E. B. M., Meyers, J. L., Barnes, K. M., & Baron, J. (2014). Evidence that estrogen hastens epiphyseal fusion and cessation of longitudinal bone growth by irreversibly depleting the number of resting zone progenitor cells in female rabbits. *Endocrinology*, *155*(8), 2892–2899. https://doi.org/10.1210/en.2013-2175

Nixon, A., Doll, H., Kerr, C., Burge, R., & Naegeli, A. N. (2016). Interpreting change from patient reported outcome (PRO) endpoints: patient global ratings of concept versus patient global ratings of change, a case study among osteoporosis patients. *Health and Quality of Life Outcomes*, *14*, 25. https://doi.org/10.1186/s12955-016-0427-5

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

Price, V., Klaassen, R., Bolton-Maggs, P., Grainger, J., Curtis, C., Wakefield, Ci., … Young, N. (2009). Measuring disease-specific quality of life in rare populations: a practical appraoch to cross-cultural translation. *Health and Quality of Life Outcomes*, *7*, 92.

Quitmann, J., Giammarco, A., Maghnie, M., Napoli, F., Di Giovanni, I., Carducci, C., … Sommer, R. (2017). Validation of the Italian Quality of Life in Short Stature Youth (QoLISSY) questionnaire. *Journal of Endocrinological Investigation*, *40*(10), 1077–1084. https://doi.org/10.1007/s40618-017-0667-1

Quitmann, Julia, Blömeke, J., Silva, N., Witt, S., Akkurt, I., Christian Dunstheimer, … Dörr, H.-G. (2018). *Quality of life of short-statured children born small for gestaional age and idiopathic growth hormone deficiency treated with growth hormone*. Manuscript in preparation.

Quitmann, Julia, Rohenkohl, A., Sommer, R., Bullinger, M., & Silva, N. (2016). Explaining parent-child (dis)agreement in generic and short stature-specific health-related quality of life reports: do family and social relationships matter? *Health and Quality of Life Outcomes*, *14*, 150.

Rajmil, L., Lopez, A. R., Lopez-Aguila, S., & Alonso, J. (2013). Parent-child agreement on health-related quality of life (HRQOL): a longitudinal study. *Health and Quality of Life Outcomes*, *11*, 101. https://doi.org/10.1186/1477-7525-11-101

Ranke, M. B., Lindberg, A., Price, D. A., Darendeliler, F., Albertsson-Wikland, K., Wilton, P., & Reiter, E. O. (2007). Age at growth hormone therapy start and first-year responsiveness to growth hormone are major determinants of height outcome in idiopathic short stature. *Hormone Research*, *68*(2), 53–62. https://doi.org/10.1159/000098707

Ravens-Sieberer, U., Ellert, U., & Erhart, M. (2007). [Health-related quality of life of children and adolescents in Germany. Norm data from the German Health Interview and Examination Survey (KiGGS)]. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz*, *50*(5–6), 810–818. https://doi.org/10.1007/s00103-007-0244-4

Ravens-Sieberer, U., Erhart, M., Wille, N., & Bullinger, M. (2008). Health-related quality of life in children and adolescents in Germany: results of the BELLA study. *European Child & Adolescent Psychiatry*, *17 Suppl 1*, 148–156. https://doi.org/10.1007/s00787-008-1016-x

Richmond, E., & Rogol, A. D. (2016). Treatment of growth hormone deficiency in children, adolescents and at the transitional age. *Best Practice & Research. Clinical Endocrinology & Metabolism*, *30*(6), 749–755. https://doi.org/10.1016/j.beem.2016.11.005

Rohenkohl, A., Stalman, S., Kamp, G., Bullinger, M., & Quitmann, J. (2016). Psychometric performance of the Quality of Life in Short Stature Youth (QoLISSY) questionnaire in the Netherlands. *European Journal of Pediatrics*, *175*(3), 347–354. https://doi.org/10.1007/s00431-015-2656-8

Sandberg, D. E., Brook, A. E., & Campos, S. P. (1994). Short stature: a psychosocial burden requiring growth hormone therapy? *Pediatrics*, *94*(6 Pt 1), 832–840.

Sandberg, David E., Bukowski, W. M., Fung, C. M., & Noll, R. B. (2004). Height and social adjustment: are extremes a cause for concern and action? *Pediatrics*, *114*(3), 744–750. https://doi.org/10.1542/peds.2003-1169-L

Schepers, S. A., Haverman, L., Zadeh, S., Grootenhuis, M. A., & Wiener, L. (2016). Healthcare Professionals' Preferences and Perceived Barriers for Routine Assessment of Patient-Reported Outcomes in Pediatric Oncology Practice: Moving Toward International

Processes of Change. *Pediatric Blood & Cancer*, *63*(12), 2181–2188. https://doi.org/10.1002/pbc.26135

Silva, N., Crespo, C., Carona, C., Bullinger, M., & Canavarro, M. C. (2015). Why the (dis)agreement? Family context and child-parent perspectives on health-related quality of life and psychological problems in paediatric asthma. *Child: Care, Health and Development*, *41*(1), 112–121. https://doi.org/10.1111/cch.12147

Silva, Neuza, Bullinger, M., Quitmann, J., Ravens-Sieberer, U., & Rohenkohl, A. (2013). HRQoL of European children and adolescents with short stature as assessed with generic (KIDSCREEN) and chronic-generic (DISABKIDS) instruments. *Expert Review of Pharmacoeconomics & Outcomes Research*, *13*(6), 817–827. https://doi.org/10.1586/14737167.2013.847366

Solans, M., Pane, S., Estrada, M.-D., Serra-Sutton, V., Berra, S., Herdman, M., … Rajmil, L. (2008). Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *11*(4), 742–764. https://doi.org/10.1111/j.1524-4733.2007.00293.x

Stephen, M. D., Varni, J. W., Limbers, C. A., Yafi, M., Heptulla, R. A., Renukuntla, V. S., … Brosnan, P. G. (2011). Health-related quality of life and cognitive functioning in pediatric short stature: comparison of growth-hormone-naive, growth-hormone-treated, and healthy samples. *European Journal of Pediatrics*, *170*(3), 351–358. https://doi.org/10.1007/s00431-010-1299-z

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., … de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties

of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

The European QoLISSY Group. (2013a). Field Test and Retest with Psychometric Testing. In *Quality of Life in Short Stature Youth* (pp. 79–89). Lengerich, Germany: Pabst Science Publishers.

The European QoLISSY Group. (2013b). *Quality of Life in Short Stature Youth: The QoLISSY Questionnaire User's Manual*. Lengerich: Pabst Science Publishers.

US CDCP. (2016, May 31). HrQoL Concepts [Government]. Retrieved from https://www.cdc.gov/hrqol/concept.htm

Versteegh, M., Leunis, A., Uyl-de Groot, C., & Stolk, E. (2012). Condition-specific preference-based measures: benefit or burden? *Value in Health*, *15*, 504–513.

Vetter, T. R., Bridgewater, C. L., & McGwin, G. J. (2012). An observational study of patient versus parental perceptions of health-related quality of life in children and adolescents with a chronic pain condition: who should the clinician believe? *Health and Quality of Life Outcomes*, *10*, 85. https://doi.org/10.1186/1477-7525-10-85

Voss, L. D., & Mulligan, J. (2000). Bullying in school: are short pupils at risk? Questionnaire study in a cohort. *BMJ (Clinical Research Ed.)*, *320*(7235), 612–613.

Waqar Rabbani, M., Imran Khan, W., Bilal Afzal, A., & Rabbani, W. (2013). Causes of short stature identified in children presenting at a tertiary care hospital in Multan Pakistan. *Pakistan Journal of Medical Sciences*, *29*(1), 53–57. https://doi.org/10.12669/pjms.291.2688

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: two simple measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618.

Wit, J. M., Reiter, E. O., Ross, J. L., Saenger, P. H., Savage, M. O., Rogol, A. D., & Cohen, P. (2008). Idiopathic short stature: management and growth hormone treatment. *Growth Hormone & IGF Research : Official Journal of the Growth Hormone Research Society and the International IGF Research Society*, *18*(2), 111–135. https://doi.org/10.1016/j.ghir.2007.11.003

World Health Organization. (1997). Programme on Mental Health: WHOQOL Measuring Quality of Life. World Health Organization.

Wu, H., Li, H., & Gao, Q. (2013). Psychometric properties of the Chinese version of the pediatric quality of life inventory 4.0 Generic core scales among children with short stature. *Health and Quality of Life Outcomes*, *11*, 87. https://doi.org/10.1186/1477-7525-11-87

Wyrwich, K. W., Norquist, J. M., Lenderking, W. R., & Acaster, S. (2013). Methods for interpreting change over time in patient-reported outcome measures. *Quality of Life Research : An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *22*(3), 475–483. https://doi.org/10.1007/s11136-012-0175-x

7 Statutory Declaration

I declare that I have written this thesis independently and did not use any other source other than those that have been declared in the reference section of this thesis. All information from these sources have been marked explicitly in the manuscript via quotations or by citations.

_____	_____
Date	Signature