FUNCTIONAL OUTCOME MEASURES IN HAEMOPHILIA –A SYSTEMATIC REVIEW

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Master of Philosophy (MPhil)

By

Maheshi Wijesekera

August 2016

Abstract: Functional Outcome Measures in Haemophilia- A Systematic Review

Maheshi Wijesekera

Background: Haemophilia is an inherited bleeding disorder that results in haemarthrosis leading to chronic arthropathy in those with severe forms of the disease. It causes significant disability and affects a patient's quality of life. Functional outcome measures enable the healthcare professionals to assess the patients' ability to carry out activities of daily living providing an important input to the assessment of joint disease.

Objectives: This study aims to carry out a systematic review to identify the existing functional outcome measures used in the adult English speaking, haemophiliac population and evaluate these instruments based on their development methodology, measurement properties and other properties.

Methods: Both PubMed and Scopus databases were searched to identify suitable outcome measures. Once the search identified the instruments, each instrument was searched to identify the relevant pilot and validation studies. Development methodology of each instrument was summarised. The measurement properties were evaluated using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) 4-point checklist. These measurement properties included internal consistency, reliability, measurement error, construct validity, criterion validity, content validity and responsiveness. The other properties that were assessed include interpretability, generalisability, precision, conceptual model, measurement model, acceptability, feasibility and burden.

Results: There were three main outcome measures used to assess function in the adult haemophiliac population. These were the Haemophilia Activities List (HAL), the Functional Independence Score in Haemophilia (FISH) and the Haemophilia Exercise Project–Test-Questionnaire (HEP-Test-Q). Information on the development of instruments was only well provided in the HAL. However the COSMIN checklist proved that the HAL had not assessed all measurement properties. The FISH and the HEP-Test-Q, did not possess very good methodological quality of its measurement properties. With regards to the other properties, all three instruments were acceptable but interpretability was poor. The HAL and the HEP-Test-Q were precise. The conceptual model instruments assessed function in different forms, whereas the measurement model was treated as a reflective model in all three instruments. The HEP-Test-Q had the most amount of burden in comparison to the other instruments. The main limitation of this study was that the FISH, a performance based instrument was evaluated using the COSMIN checklist that was developed to assess patient reported outcomes.

Conclusion: This systematic review suggests that the existing instruments produced to assess function in adult persons with haemophilia have not been adequately validated and that the methodology undertaken for this process consists of certain drawbacks. This suggests that there is scope for a new instrument to assess function in the English speaking adult haemophiliac population.

Acknowledgements

First and foremost I would like to thank my primary supervisor, Professor Simon Frostick for giving me the opportunity to undertake this MPhil degree and letting me to further pursue my interest in musculoskeletal sciences. The support and guidance he has provided me throughout has made this year a period of both intellectual and professional development.

Mr Joseph Alsousou, my secondary supervisor and everyone in the musculoskeletal research group who have constantly helped me and supported me over the past year.

Special mention should go to the haemophilia teams at the Royal Liverpool University Hospital and Guy's and St Thomas' Hospital for giving me further insight into this rare disease and helping me understand the disease from a clinical perspective.

I would also like to thank Dr Richard Jackson and Miss Silvia Cicconi for their assistance on such short notice and helping me with the statistical evaluation of studies in carrying out this systematic review.

I am grateful to my family for supporting me this year. If it weren't for my parents financially supporting me, this degree would not have been possible. I am especially thankful to my mother for helping me with the constant proof reading of manuscripts of this thesis.

Finally, I would like to thank Mu-Chun, Mina, Kavitha and Henal for providing me with moral support during this very challenging year.

Table of Contents

List of Figures VI
List of tables VII
List of abbreviations VIII
Glossary of terms X
1 CHAPTER 1- Introduction 1
1.1 Haemophilia 1
1.1.1 Disease overview 1
1.1.2 Historical perspective 2
1.1.3 Epidemiology 2
1.1.4 Classification 3
1.1.5 Clinical course 3
1.1.6 Complications 4
1.1.7 Management of haemophilia 10
1.1.8 Management of haemarthrosis 15
1.2 Disability 18
1.2.1 Overview and historical perspectives 18
1.2.2 Definitions and concepts 19
1.2.3 Conceptual frameworks 19
1.2.4 Major disability models 21
1.2.5 Measurement of disability/ function 23
1.3 Definitions of concepts 27
1.3.1 Theories of methodology 27
1.3.2 Measurement models 28
1.3.3 COSMIN 31
1.3.4 Measurement properties 31
1.3.5 Other properties 34
1.4 Outcome measures in haemophilia 36
1.4.1 Physical examination tools 37
1.4.2 Imaging scores 38
1.4.3 Quality of Life measures 39
1.5 Hypothesis 41
1.6 Aim 42
2 CHAPTER 2- Methods 43

2.1 Literature search 43
2.1.1 Inclusion criteria 44
2.1.2 Exclusion criteria 44
2.2 Evaluation 46
2.2.1 Development of instruments 47
2.2.2 Measurement Properties 47
2.2.3 Other properties 48
3 CHAPTER 3- Results 51
3.1 Literature search 51
3.2 Evaluation 54
3.2.1 Haemophilia Activities List 54
3.2.2 Functional Independence Score in Haemophilia 59
3.2.3 HEP-test-Q score 63
4 CHAPTER 4- Discussion 70
4.1 Literature search 70
4.2 Critical appraisal tool 70
4.3 Evaluation 72
4.3.1 Development of instruments 72
4.3.2 Measurement properties of instruments 73
4.3.3 Other properties 78
4.4 Assessment of joint disease 83
4.5 Limitations 85
5 CHAPTER 5- Further Study and Conclusions 86
5.1 Further study 86
5.2 Conclusion 86
6 REFERENCES 88
7 APPENDICES 121
7.1 Appendix 1- COSMIN table of definitions 121
7.2 Appendix 2- Haemophilia Activities List 123
7.3 Appendix 3- Functional Independence Score in Haemophilia 134
7.4 Appendix 4- HEP-Test-Q 135
7.5 Appendix 5- ISOQOL minimum standard definitions 137

List of Figures

Figure 1 Pedigree diagram of inheritance pattern of haemophilia 1
Figure 2 Pathophysiology of haemophilic arthropathy 8
Figure 3 Components of the ICF classification system[100] 22
Figure 4 Reflective model versus Formative model 29
Figure 5 COSMIN taxonomy[129] 34
Figure 6 Flow diagram of literature search 52

List of tables

Table 1 Reflective versus Formative measurement models[126] 30
Table 2 MeSH terms used 44
Table 3 Inclusion and exclusion criteria 45
Table 4 Data extraction for interpretability 49
Table 5 Interpretation of readability scores 50
Table 6 Instrument distribution in literature search 53
Table 7 Individual searches per instrument 53
Table 8 Generalisability results of studies 68
Table 9 Evaluation of measurement properties using the COSMIN 4-point checklist 69

List of abbreviations

ADL	Activities of Daily Living
AIMS	Arthritis Impact Measurement Scale
CHO-KLAT	Canadian Haemophilia Outcomes -Kids Life Assessment Tool
COPM	Canadian Occupational Performance Measure
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
DNA	Deoxyribonucleic acid
FISH	Functional Independence Score in Haemophilia
HAART	Highly Active Anti-Retroviral Therapy
HAL	Haemophilia Activities List
HAQ	Health Assessment Questionnaire
HBV	Hepatitis B Virus
HCV	Hepatitis C Virus
HEP-Test-Q	Haemophilia Exercise Project- Test-Questionnaire
HIV	Human Immunodeficiency Virus
HJHS	Haemophilia Joint Health Score
HRA	Health Research Authority
HR-QoL	Health Related- Quality of Life
ICC	Intra-class Correlation Coefficient
ICF	International Classification of Functioning, disability and health
ICIDH	International Classification of Impairments, Disabilities and Handicaps
IPA	Impact on Participation and Autonomy questionnaire

IPSG	International Prophylaxis Study Group
ISOQOL	International Society for Quality Of Life
MACTAR	McMaster-Toronto Arthritis Patient preference Disability Questionnaire
MDM2	Mouse Double Minute 2 homolog
MRI	Magnetic Resonant Imaging
NHS	National Health Service
OJS	Orthopaedic Joint Score
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM	Patient Reported Outcome Measure
QoL	Quality of Life
ROM	Range of Motion
SF-12/36	Short Form-12/36
UK	United Kingdom
UKHCDO	United Kingdom Haemophilia Centre Doctors' Organisation
WFH	World Federation of Haemophilia
WHO	World Health Organisation
WOMAC	Western Ontario and McMaster Universities Arthritis Index

Glossary of terms

Acceptability	The willingness or ability of patients from the target population to complete the questionnaire.
Clinimetrics	The measurement of clinical phenomenon by an instrument.
Concept	The specific goal of the instrument or the global event that is being measured.
Conceptual framework/ model	A schematic representation of the theory, explaining the underlying phenomena between items, domains, subconcepts, concepts and the scores produced by the instrument.
Construct	Well defined and precise subject of measurements that are usually unobservable characteristics
Construct validity	The extent to which an instrument validly measures the construct meant to be measured.
Content validity	The ability of an instrument to reflect the construct being measured in its content.
Criterion validity	Extent to which a score of an instrument reflects the score of its 'gold standard'.
Cross-cultural validity	Ability of items to mirror the performance of the translated or culturally adapted instrument to the original instrument

Face validity	The extent to which the items of an instrument represents the construct being measured.
Feasibility	This assesses the burden of an instrument on the healthcare professional.
Instrument	A method to capture data which may take the form of a questionnaire as well as the information and documentation that support its use.
Internal consistency	The extent to which the items are interrelated.
Interpretability	The extent to which a person can accredit qualitative meaning to the score of an instrument.
Item	An individual question, statement or task that is evaluated to address a particular concept.
Measurement model	Measurement models depict the causal relationships between the construct measured and its items.
Precision	The ability of a score of an instrument to make distinctions.
Psychometrics	The field of study concerned with the theory and technique of psychological measurement.
Responsiveness	The ability of an instrument to detect change over time in the construct to be measured.

Structural validity	The extent to which the dimensionality of the construct being measured is reflected by the scores of the instrument.

1 CHAPTER 1- Introduction

1.1 Haemophilia

1.1.1 Disease overview

Haemophilia is an inherited bleeding disorder and is the most common of all the severe bleeding disorders. It is caused by a deficiency of a single clotting factor; factor VIII in haemophilia A (Classic haemophilia) and factor IX in haemophilia B (also known as Christmas disease) comprising 85% and 15% of cases respectively. It follows an X-linked recessive mutation thus predominantly affects males.[1] In the absence of a family history the rate of spontaneous mutations is 30%.[2] The carriers of this condition (females) lack about 50% of the respective clotting factor but this suffices for normal clotting to take place. About 10% of carriers may have less than 50% of normal amount of factor making them more susceptible to bleed especially after surgery or trauma.[3, 4]

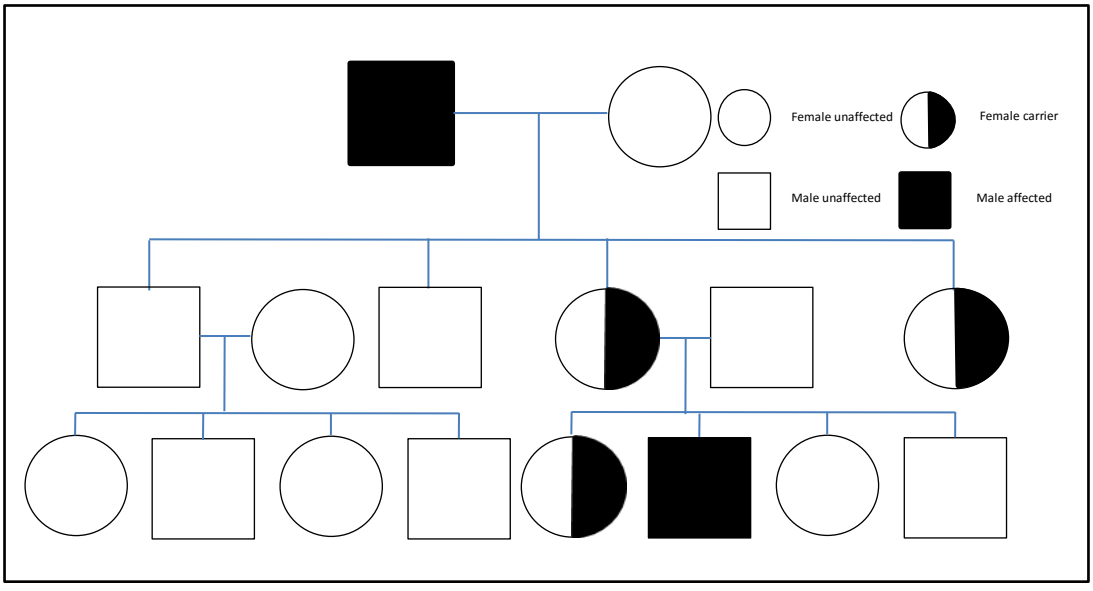

Figure 1 **Pedigree diagram of inheritance pattern of haemophilia**

A rare form of haemophilia that results from a lack of factor XI is haemophilia C (also known as Rosenthal syndrome). This is thought not to be a 'true' haemophilia as it is an autosomal recessive condition and affects both males and females equally.[5] This condition is not usually associated with spontaneous haemarthrosis.[6] Another form of the disease that does not carry a genetic predisposition is, 'acquired haemophilia.' It is caused by autoantibody production against Factor VIII and presents with bleeding but in later life. Acquired haemophilia predominantly involves skin bruising and soft tissue bleeding whereas haemarthrosis is unusual.[7] In about 50% of acquired haemophilia cases the cause is unknown although the known causes include drug interactions, cancer, immune disorders and pregnancy or post-partum.[8]

The deficiency of coagulation factor in haemophilia results in the patient being more likely to bleed, often without an apparent cause and this can occur in one site or multiple sites. Prior to treatment availability congenital haemophilia most commonly affects the musculoskeletal (80%) and central nervous systems (20%) and results in various complications which will be explained further below.[9]

Further discussion is limited to haemophilia A and B.

1.1.2 Historical perspective

Historical writing on haemophilia dates back to the Rabbinic writings in Judaism in the 2nd Century AD. This disease was later considered a 'royal disease' as Queen Victoria was a carrier of haemophilia B, which resulted in some of her royal descendants across the continent such as Russia, Germany and Spain, inheriting the disorder.[10, 11]

1.1.3 Epidemiology

Haemophilia has an average prevalence of one per 10,000 of the population.[9] More specifically, incidence among males is 1 in 5,000 for haemophilia A and 1 in 30,000

for haemophilia B.[2] According to the United Kingdom (UK) survey statistics in 2014, the number of persons diagnosed with haemophilia was 6811, out of which 5646 had haemophilia A and 1165 persons had haemophilia B.[12] Haemophilia has no particular affinity towards a certain race or ethnic group.[13, 14]

1.1.4 Classification

Classification of persons with haemophilia is based on the level of coagulation factors present in blood. The normal reference range for factor levels in blood is 50-150 $IUmL^{-1}$ (50%-150%). Mild haemophilia is associated with a factor activity of 0.06-0.40 $IUmL^{-1}$ (6%-40%), whereas moderate and severe haemophilia are associated with factor activity levels of 0.01-0.05 $IUmL^{-1}$ (1%-5%) and of <0.01 $IUmL^{-1}$ (<1%) respectively.[15] These levels of factor generally remain unchanged throughout life. Mild forms of the disease may go undiagnosed if they do not manifest as symptoms although the level of factor present in the blood is low.[16 17]

1.1.5 Clinical course

Families with a known background of the disease will be offered appropriate testing on every male child due to the inheritance pattern of this disease. When facilities are available the investigations are arranged antenatally including genetics that partly involves screening of male relatives. For patients who present with certain signs of the disease but in the absence of a family history, a new mutation would be tested for. Features in the undiagnosed neonate can range from intracranial haemorrhage, extra-cranial haemorrhage, bleeding from puncture sites, internal bleeding, haemarthrosis, bleeding from the umbilical stump or bleeding following circumcision. In such neonates, it is imperative to carry out investigations in order to seek a diagnosis.[18] Assays of factors in clot formation usually form the basis of initial diagnosis.

As the severity of haemophilia increases, more intensive follow-up, where a comprehensive care program with frequent clinical and laboratory review is carried out.[19] In the UK, patients with severe or moderate disease should be offered a clinical or multidisciplinary review twice a year as a minimum whilst those with mild disease are assessed annually.[19] With sufficient treatment and care as seen in most western nations, persons with haemophilia live near to normal lives indicating a good prognosis. However, without adequate treatment as observed in low-income nations, most children with severe disease will die young.[20]

1.1.6 Complications

1.1.6.1 Overview

As patients with haemophilia are prone to bleeding, the clinical features are associated with bleeding from the respective body system. Patients with severe haemophilia suffer with spontaneous bleeding of joints and other tissues while those with moderate disease suffer with less spontaneous bleeding but still remain at a high risk of bleeding after minor trauma. Whereas those with mild disease are only susceptible to bleeding after major trauma.[17] Post traumatic bleeding after dental or other surgical procedures are seen.[21] The central nervous system may be affected as these patients have an increased risk of intracranial bleeds after traumatic injury.[22]

The complications of haemophilia are closely linked with the musculoskeletal system and consist mainly of haemarthrosis and less commonly pseudotumours resulting in fractures.[23, 24] Pseudotumours are a clinical entity caused by chronic encapsulated haematomas. This rare phenomenon may be of osseous or soft tissue origin and result in the destruction of adjacent structures depending on the anatomic site.[25] Gilbert et al[26] described two distinct groups of pseudotumours. Proximal pseudotumors are seen in the adult population and originates in axial skeleton around the pelvis and proximal femur. Whereas distal pseudotumors predominantly affect the skeletally immature and

is caused by intraosseous bleeding in the cancellous bone.[26]The incidence of pseudotumors have decreased with better control of the disease.[27] Persons with haemophilia who are immunocompromised due to associated comorbidities may be more susceptible to developing septic arthritis, soft tissue infections and low bone mineral density.[28, 29]

The severity of haemophilia is often reflected in the clinical features exhibited such as spontaneous haemarthroses accompanied by functional impairment seen in the severe form of the disease. In mild haemophilia, as bleeding is only likely to follow trauma or surgery, these patients are treated as and when they present. These patients have very minimal functional impairment caused by their haemophilia.[30, 31]

Patients with haemophilia not only suffer from the obvious physical symptoms but also have psychosocial issues associated with the genetics, severity and chronicity of the disease. The psychological manifestations can include poor self-esteem, anxiety and stress as a result of various factors such as social restrictions due to financial difficulties.[32]

1.1.6.2 Haemarthrosis

Haemarthrosis is important as it is the most common complication of severe haemophilia and if treated inappropriately, can result in highly disabling morbidity such as chronic arthropathy and flexion deformities.[24, 33] Generally joints have a higher tendency to bleed because there are low levels of tissue factor, a crucial component for haemostasis within the joints. This, combined with a lack of either factor VIII or factor IX results in a higher tendency to bleed into the joints.[34, 35] Prior to the availability of current treatments for the disease, haemarthrosis could occur up to twenty times a year in those with severe disease.[36]

1.1.6.2.1 Target joint

Most patients with haemophilia who experience recurrent haemarthrosis are said to develop a target joint, where initial destruction of a joint will take place. Such a target joint is more likely to develop in patients with severe disease than those with moderate or mild haemophilia. Those with severe disease may also have more than one target joint.[37, 38] There are multiple definitions used for a 'target joint' as defined by various organisations.[39-41] The International Society of Thrombosis and Haemostasis defines a target joint as "three or more spontaneous bleeds into a single joint within a consecutive 6 month period".[42] This group further goes on to say that if there are less than 2 bleeds into a joint over a consecutive 12 month period it is no longer known as a target joint.[42]

These target joints are more susceptible to develop chronic synovitis and cause damage to both cartilage and bone. Haemarthrosis most commonly affects the knees, ankles and elbows. Other joints such as hips and shoulders are rarely affected, showing a preference towards the weight bearing joints.[23] The appearance of deformity caused by chronic contractures in haemophilic arthropathy may take the form of an equinus deformity in the ankle, or a flexion contracture involving the knee or elbow joints.[24]

1.1.6.2.2 Pathophysiology

Haemophilic arthropathy occurs in three stages, it is initially set off by acute haemarthrosis which results in chronic synovitis and finally progresses to degenerative arthritis.[43] During the acute insult of haemarthrosis, the pathological process manifests itself as joint capsular distension and vasodilatation of the capsular and epiphyseal bone vessels, resulting in destruction of these structures. The disintegration of the cartilage occurs with recurrent haemarthroses.[44] The pathogenesis that follows this is incompletely understood. It is thought that the iron present in red blood cells plays a

pivotal role in the inflammatory process.[43, 45] The breakdown product of haem, haemosiderin initiates the production of pro-inflammatory cytokines; particularly interleukins 1, 6 and tumour necrosis factor-alpha.[46-49] Furthermore, it induces genes that are present in the synovial tissue such as Mouse Double Minute 2 homolog (MDM2) that results in cellular proliferation as seen in synovitis.[50]

Following the development of haemarthrosis, the synovium attempts to reabsorb excess blood, thereby resulting in hypertrophic and fragile synovial lining with an increased tendency to rebleed. The target joint contains abundant vascular and hypertrophic tissue which have a tendency to become impinged between articular surfaces and has a higher affinity to bleed resulting in a vicious cycle of haemarthrosis-synovitis-haemarthrosis[43, 51] The blood and the inflammatory mediators that are released into the joint interferes with the homeostasis of articular cartilage especially if not aspirated early.[51-57] The synthesis of proteoglycans by chondrocytes are affected by the presence of intra-articular blood, leading to the apoptosis of chondrocytes.[24] Haemophilic arthropathy bears resemblance to osteoarthritis and rheumatoid arthritis, with regards to the progressive degeneration, inflammation of hyaline cartilage and synovial hypertrophy. The mentioned phenomenon progresses until the joint is completely destroyed.[51, 58, 59]

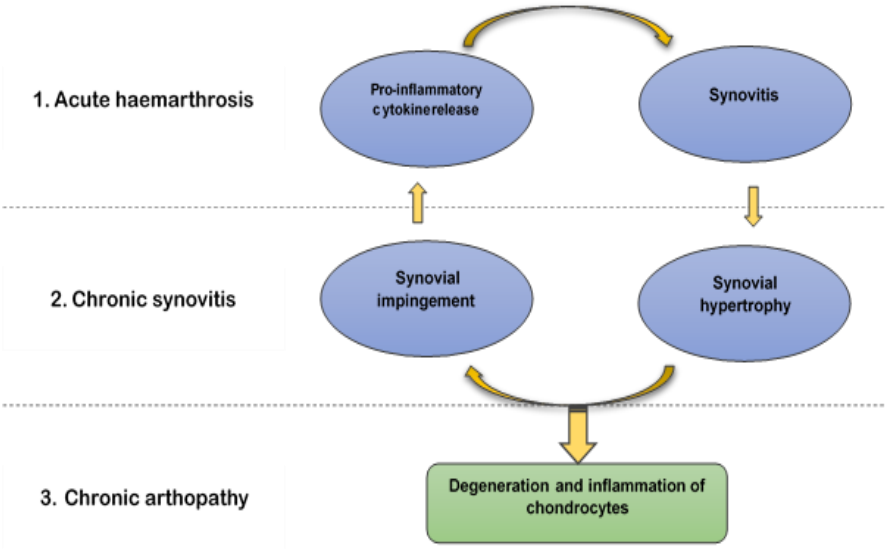

Figure 2 Pathophysiology of haemophilic arthropathy

1.1.6.2.3 Epidemiology

The chronic degenerative changes explained above affect 90% of persons with severe haemophilia whereby 1-6 major joints are affected by the age of 20-30 years.[24] The Range of Motion (ROM) in joints is affected in almost 85% of the patients with the severe disease.[43, 60] In the United States, 36% of those with severe haemophilia reported needing assistance with mobilising and 21% with haemophilia missed either school or work (mean of 10 days per year) due to their joint problems.[61] These high levels of functional impedance suggest that preventative intervention is of paramount importance.

1.1.6.2.4 Clinical manifestations

Symptoms associated with acute haemarthrosis are tingling, limited ROM and pain at the affected joint that is generally followed by rapid joint swelling and bruising. In weight bearing joints, weight bearing may be impossible. On examination there may be ecchymosis, with a local increase in temperature and tenderness with the joint usually held in a flexed position.[62, 63]

Arthropathy presents with a myriad of symptoms that greatly affects the patient's quality of life.[23, 38, 60, 64, 65] Recurrent haemarthrosis may lead to muscle atrophy, ankylosis, reduced ROM and eventually arthropathy.[66, 67] Arthropathy, once established is progressive and irreversible.[68]

1.1.6.3 Psychosocial aspects

The psychosocial challenges in congenital haemophilia are observed at three levels. Those that come with the diagnosis and treatment of haemophilia, the impact on the individual with haemophilia, and the impact on family life and social challenges encountered.[69]

The challenges surrounding the diagnosis and treatment of haemophilia have reduced over the years with the availability of prophylaxis and home treatment in developed nations.[69] However, challenges remain such as the use of blood products due to concerns over the transmission of infections and prion disease as portrayed in the media.[70, 71] In families that have not encountered haemophilia before, the initial diagnosis not only affects the emotional wellbeing of the family but also financially.[69] Parents of those initially diagnosed with haemophilia, often experience a prolonged stage of denial or withdrawal thereby affecting their wellbeing.

Persons with haemophilia experience anxiety surrounding the increased risk of bleeding, chronic joint pain and reduced function.[69] In the cohort who have contracted Human Immunodeficiency Virus (HIV) or Hepatitis C (HCV) patients often experience depression due to associated comorbidities as well as the stigma attached to these conditions.[72] During their childhood persons with haemophilia may experience low self-esteem and may feel isolated due to the fear of inducing a bleed.[73] Mothers of persons with haemophilia frequently experience guilt for passing on the disease and play a significant role in the disease management of their children.[74] This

guilt may result in frustration or depression and as a consequence may result in rejection or overprotection of the child.[75] With the main focus of the family being the person with haemophilia, other siblings may feel neglected and this may have an impact on the family dynamics.[74, 76, 77] A new diagnosis of haemophilia in some families may negatively affect relationships between parents or even lead to abandonment by the father .[69]

Reproductive technology may offer the opportunity to screen for an affected male with haemophilia among carriers and consequently aid in decision making amongst pregnant mothers. In cultures where this may not be executed and where cultural shaming for having a child with a disability take place, affected families will conceal the diagnosis in order ensure that their daughter is able to marry.[78]

The psychosocial issues in persons with haemophilia are similar to other chronic disease, but certain characteristics are unique to haemophilia. The unpredictable nature of bleeding, its associated complications and the psychological burden on the carriers.[73] The severity of the psychosocial impact in these three categories varies across the severity of the disease and in the presence of inhibitors.[79] Therefore it is crucial to provide psychosocial support early on not only to the patient but also their families.

1.1.7 Management of haemophilia

1.1.7.1 Overview

The conventional method of treatment consists of intravenous replacement of the deficient factor using commercially prepared factor concentrates. The two mainstays of treatment are ‘on-demand’ therapy and prophylaxis. ‘On-demand’ therapy is when factor concentrate administration takes place after a bleeding event. Prophylaxis is the

administration of treatment in the absence of bleeding with an aim of preventing any episodes of bleeding and can be given in the short-term or long-term.

Regardless of the definition used for a target joint, it results in delayed diagnosis of the affected joint. Thus long-term prophylaxis remains the gold standard treatment in those with severe disease in the western nations.[16, 80]

1.1.7.2 Historical Perspective

A therapeutic breakthrough for haemophilia arose with the introduction of cryoprecipitate in its treatment in the 1960s. The introduction of factor concentrates followed a decade later. These were associated with complications as blood products were not screened for infectious diseases such as HIV and HCV. Consequently those who received unscreened factor were predisposed to developing these infections.[24] This was followed by the introduction of screened blood products to minimize this effect.[81, 82] There remains continuing concerns about the theoretical transmission of prion disease such as variant Creutzfeldt's Jakob Disease (vCJD).[83]

Prior to the introduction of blood screening, there were about 7250 persons with haemophilia. All those with severe haemophilia (n= 2262) were said to be diagnosed with HCV as a consequence of contaminated blood products.[84] There were 1246 persons with severe haemophilia diagnosed with HIV.[84] Most of these patients were reported to have progressed to chronic liver disease and some patients were co-infected with HBV, HCV and/ or HIV. Since the introduction of HIV screening for blood products in 1985 the cases of HIV among persons with haemophilia started reducing. In 2006 there were about 360 patients in the UK with a diagnosis of HIV as a consequence of the treatment received for their bleeding disorder.[19] (These numbers are not haemophilia specific.) Persons with haemophilia that are immunocompromised, often carry a higher risk of liver disease and carcinoma of the

liver that is increased by coinfection with HCV.[85, 86] In addition persons with haemophilia that are immunocompromised have a significant risk of secondary malignancies such as lymphoma.[87]

Since the introduction of highly active anti-retroviral therapy (HAART) the morbidity of HIV positive persons with haemophilia has decreased.[84]However the use of HAART come with side effects such as unusual bleeding in this population with protease inhibitors.[88]

Current practice in the UK routinely offers vaccinations to persons with haemophilia against hepatitis A and HBV in order to prevent contraction of such viruses.[89]

1.1.7.3 Treatment options

Two main types of clotting factor concentrates are available including plasma derived clotting factor concentrates and recombinant factor concentrates. The plasma derived clotting factor concentrate requires screening for infectious agents while the recombinant factor does not, as the factor concentrates are produced using genetic and deoxyribonucleic acid (DNA) technology. Therefore recombinant factor remains the favourable treatment of choice as recommended by the United Kingdom Haemophilia Centre Doctors' Organisation (UKHCDO).[19] This was introduced in the UK in 1999.

Desmopressin a synthetic vasopressin analogue that can be administered intravenously, subcutaneously or as an intranasal spray to increase the release of factor VIII is used in the mild forms of the disease.[16] In an attempt to halt blood loss, antifibrinolytics such as tranexamic acid or epsilon aminocaproic acid may be administered. This acts by preventing the formation of the fibrinolytic enzyme. This is often used for oral bleeds or after dental surgery.[90]

1.1.7.4 Treatment Regimens

Initially, the management option used for haemarthrosis was, 'on-demand' therapy. This treatment regimen resulted in patients developing severe destruction of their joints by the age of 20 or 30 years. These trends observed by specialists led to the current treatment options of regular prophylaxis to reduce the incidence of haemarthrosis.

Newly diagnosed haemophilia patients should receive primary prophylaxis by the second joint bleed or significant soft tissue bleed according to recommendations by the UKHCDO.[91] Secondary prophylaxis, is when factor administration follows several articular haemorrhages.[68]

The goal of prophylaxis is to prevent the development of associated complications such as chronic arthropathy. The advancement of therapeutic options in developed nations has seen self-administration at home as a popular option for those with haemophilia. Advantages associated with this treatment method are convenience and rapid access to the factor. This lowers the level of side effects and complications such as the frequency of haemarthosis and progression of arthropathy.[91] Due to the financial burden and the mode of coagulation factors administration, primary prophylaxis is only recommended in developed countries. It is noteworthy that self-administration therapy is associated with a risk to patients who receive chronic intravenous therapy and includes the complications of an indwelling venous catheter. The majority of patients with haemophilia do not have access to self-prepared factor or hospital treatment as they come from low-income countries, thereby leaving these patients often treated 'on-demand' or remain untreated.[24, 68]

When haemophilic patients participate in strenuous activities such as sport, episodic prophylaxis where factor replacement is given for short period intervals, may confer benefit by avoiding unanticipated bleeds.[63]

1.1.7.5 Epidemiology

In the World Federation of Haemophilia (WFH) annual global survey in 2014, the percentage estimate of children who were on prophylaxis in the UK was 95% and the percentage estimate of those above the age of 18 who were on prophylaxis was 70%.[12]

1.1.7.6 Inhibitor development

Factor replacement is associated with the development of inhibitors, which are antibodies that render factor concentrates ineffective and therefore, may allow uncontrolled bleeding. Inhibitors may develop in about 30% of those with haemophilia A after administration of its factor concentrate although this may be transient in about half of those.[19, 92] Such patients often have worsened complications of joint disease that result in poor Quality of Life (QoL).

In patients who have developed inhibitors, the treatment of acute bleeding episodes is extremely difficult. In such situations, more factor replacement or a different type of clotting factor (such as recombinant factor VIIa) will be required. This increases the financial burden for the provision of adequate care for such patients.[93]

1.1.7.7 Evolution of joint function with current treatment regimens

As treatment regimens for haemophilia advanced in the developed world, there was a noticeable change in the observed joint function. In the past, patients were restricted from high intensity tasks such as sport, as an increased level of uncontrollable bleeding in the musculoskeletal system was observed.[94] This was followed by a temporary period of immobilisation caused by the acute insult. As haemarthrosis usually begins

at a young age, immobilisation caused inadequate muscle growth.[95] The functional capacity proved to be poor in these patients as there was minimal muscle bulk to compensate for the destructed joint.[95-97]

The perception held regarding the limitation of physical activity has reformed with the availability of novel treatment regimens with patients being encouraged to participate in sport.[98] This has been regarded as a good method of decreasing the incidence of haemarthroses. The concept of increased activity has led to the establishment of physiotherapy as a means for improving joint function.[95]

1.1.8 Management of haemarthrosis

1.1.8.1 Overview

In spite of the regular factor replacement, haemarthrosis and arthropathy can still develop as some articular bleeds and subclinical hemorrhages are unavoidable.[24] There are two main stays in the management of haemarthrosis, both with aims to impede the development of arthropathy: 1) management of acute haemarthrosis 2) management of the recurrent haemarthrosis that is results in the development of chronic arthropathy.

1.1.8.2 Management of acute haemarthrosis

In an acute haemarthrosis, in persons with haemophilia, administration of the relevant factor concentrates is the first line treatment. This can be done at home in those who are able to self-administer. Adjuvant treatment with analgesia aims to relieve pain and restore movement.[33] Arthrocentesis may be carried out following extreme episodes of haemarthrosis and should be performed once factor concentrate has been administered.[99] Other recommended therapeutic measures involve physiotherapy that include rest, ice, compression and elevation, but their consequences and practicality should be taken into account. Physiotherapy is available in different forms, provides

well rounded rehabilitation following haemarthosis as it has an effect on decreasing the bleeding frequency, joint instability and improves ROM and muscle bulk.[63, 100]

1.1.8.3 Management of chronic arthropathy

The knee joint used to be the most commonly affected joint in persons with haemophilia but since the introduction of prophylaxis, the ankle joint has the highest incidence of haemarthrosis.[101] This is thought to be due to the higher activity rate as patients have acquired the skill to administer prophylactic and treatment regimens from home, placing the ankle at a more vulnerable position.[101]

Despite conservative management options being freely available, surgical intervention aims to correct deformity pertained to the joint. This highlights the importance of a multidisciplinary approach especially involving the haematologists and the orthopaedic surgeons.

1.1.8.3.1 Non-operative Management

The initial management of joint disease comprises the use of physiotherapy, orthotics and use of corrective devices.

Physiotherapy can vary from hydrotherapy, strength training and balance strength, balance training, and isometric to isotonic exercises, ultrasound, pulsed short waved diathermy and sports therapy.[102] In some cases where pain is severe, transcutaneous electrical nerve stimulation may be attempted as it may also help with reducing inflammation and improve mobility. Recent advancements in therapies such as the programmed sports therapy[95] has also been implemented in some countries. Physiotherapy is vital not only in the prevention of haemarthrosis but also is responsible for the rehabilitation in severe arthropathy as well as in the post-surgical care. Advances in physiotherapy, such as programmed sports therapy have more recently been established in developed countries. This enables patients to carry out

regulated therapy independently at home.[95] In low income countries where clotting factors are not widely available, physiotherapy plays a vital role as it is more economically feasible.

The intra-articular injection of hyaluronic acid- a natural substance of the synovial fluid and cartilage has been recommended and has been used for the treatment of haemophilic arthropathy of the knee. This is a promising option in whom non-operative therapy has failed and where surgical intervention is impracticable.[103]

1.1.8.3.2 Surgical management

Surgical intervention is often performed on those who present with arthropathy as a consequence of haemarthrosis. This may take the form of soft tissue procedures (i.e. synovectomy, release of muscle contractures), osteotomies and arthrodesis or total joint replacements in those with severe pain and disability.

Synovectomy is a surgical procedure that is frequently performed on persons with haemophilia with chronic haemarthrosis and synovitis; the aim is to delay total joint replacements. This procedure involves excision of the inflamed and hypertrophic synovium in order to decrease the propensity of recurrent haemarthrosis in a particular target joint. Having an estimated 70-100% reduction in the frequency of haemarthrosis, synovectomy maximises the effects of medical treatment and averts the progression to symptomatic arthritis.[104-108] Synovectomy may be performed as either surgical (open or arthroscopic), chemical or as radiosynovectomy. Chemical synovectomy involves the injection of a chemically active agent. Whereas radiosynovectomy (radionuclide synovectomy) involves intra-articular injection of a radionuclide to induce fibrosis. Each method has its advantages as well as disadvantages but the success rates remain similar for each procedure.[63] It should be

noted that even though success with synovectomies have been documented, it does not halt disease progression to arthropathy.

Total joint replacements have been performed in persons with haemophilia where the joints focused on are the knee, hip, ankle, elbow and shoulder joints. Previous studies on knee arthroplasty indicate high rates of complications due to the mode of factor replacement (i.e. pulse infusion).[109, 110] Recent studies using continuous infusion of factor concentrates, have shown a lower reported frequency of bleeding and pain related complications.[28]

During the end stage disease **arthrodesis** may be performed to reduce pain especially in the ankles, although arthrodesis of other joints have also been reported.[111]

1.2 Disability

1.2.1 Overview and historical perspectives

Disability, the inability to function normally, is a major health concern, and it is not restricted to one particular organ system. Common forms of disability are that of locomotion, special senses (hearing and vision) and psychological. It is often seen as an outcome of disease as well as a diagnosis.[112] Functional status can be viewed from various aspects. In the past traditional models solely viewed function using a biomedical model.[113] This has evolved over the years into using a disablement model that focuses on the loss of function with its association to activities of daily living [114,115]

1.2.2 Definitions and concepts

Defining disability proves to be difficult due to very little consistency across definitions in literature. Most definitions mirror the conceptual elements that they are derived from.[116] Due to the broad nature of the subject a good definition should be able to encompass all the categories of disability.

A group of authors from the 'Measuring Health and Disability in Europe' consortium[117] defined disability and addressed the drawbacks of existing definitions. It was defined as " difficulty of a function/functions at the body, person, or societal level, in one or more life domains, as experienced by an individual with a health condition in interaction with contextual factors"[117].

1.2.3 Conceptual frameworks

Various conceptual frameworks have been used to define and explain disability and/ or function, as it has been the case for disease or disorder. This is likely due to the lack of a single conceptual framework among medical disciplines. The introduction of a conceptual framework for disability was made during the 1960s-1970s.[118] This allowed further scientific inquiry into medical, psychological and sociological forms of disability in order to form comprehensive rehabilitation care. This will be looked at further using the relevant conceptual frameworks that have been introduced by authors.

There are some main classes of models used to describe disability, which include the biomedical, psychosocial, functional, and integrated models. Each model views disability from a different perspective to the other, hence adding a different outlook to disability.[119]

1.2.3.1 Biomedical model

The initial understanding of disability arose from the biomedical model, which stated that disability was a direct consequence of a physical disorder, injury or disease. This model assumes that disability can be improved, cured or progression could be halted using appropriate management options.[119]

1.2.3.2 Psychosocial models

A psychological model indicates that those with a disability execute tasks because they are influenced by the same psychological processes as those without a disability. While a social model of disability demonstrates that an individual cannot partake in activities not only as a result of their disability but also because of social barriers, negative attitudes of those without a disability and environmental constraints. The social model emphasises that better access to facilities should be provided for those with disability. The removal of social barriers and constraints with a view to improving the disability status, may make it difficult to implement this model in practice. The concept behind this model has proven difficult to understand by healthcare professionals and it neglects to grasp the complexity of diseases.[119, 120]

1.2.3.3 Functional models

A functional model of disability is based on the external expression of the disease or the impairment. Thus, this model requires adaptation of any treatment provided to improve the patient function rather than to treat the underlying disease. Often those who have considerable functional limitations are classified as being disabled, while those without considerable functional limitations are classified as physically fit. Thus disability under the functional model may be viewed as dichotomous. Functional models have the advantage of adopting a pragmatic approach to assess disability. This model takes a large number and classifies them into smaller accommodating groups. This model carries the risk of being over simplistic as it turns a blind eye to the fact

that disability occurs on a continuum and in addition has the influence of physical factors and other external factors such as environmental factors.[119]

1.2.3.4 Integrated models

Most of the literature available on the conceptual frameworks defining disability has integrated two or more models from the existing conceptual frameworks.[119] These models have proven to have more success when describing disability.

1.2.4 Major disability models

An article in 2003 by Jette and Keysor[121] proposed three major disability models used in practice and research. These include the 'Disablement Model' which was developed from the theory proposed by Nagi[122], the WHO's International Classification of Impairments, Disabilities and Handicaps (ICIDH)[123] and its revision, the International Classification of Functioning, Disability and Health (ICF).[124] All these three models share a common view that disablement signifies a series of related concepts that describes the health impact or consequences of a disease. The ICF is superior to the other two models as it portrays the disability framework well.[121]

1.2.4.1 International Classification of Functioning, Disability and Health

The WHO echoed the shift in models from disability to function by revising the existing ICIDH -1980 to the updated International Classification of Functioning, Disability and Health- 2001 (ICF)[124] as it was believed to need a more comprehensive bio-psycho-social model to assess function.

This ICF model views disability and functioning as outcomes of the interactions between health conditions and contextual factors. The latter consist of personal and environmental factors. The ICF identifies different components that interplay between

person, their ability and the environment. These components can be used in a description on how disability has an effect on one's disablement process. The production of such an integrative model is vital to the multidisciplinary management of disability.

The ICF identifies three components of human functioning: (1) impairments to body structures and functions; (2) activity limitations; and (3) participation restrictions, each of which is affected by personal and environmental factors.

Definitions for each of these components were described by the ICF. Impairments are described as "difficulties and/ or deviations from body functions or structures," activity limitations are defined as "difficulties an individual may have in executing tasks," and participation restrictions as "difficulties one may experience when engaging in life situations"[124].

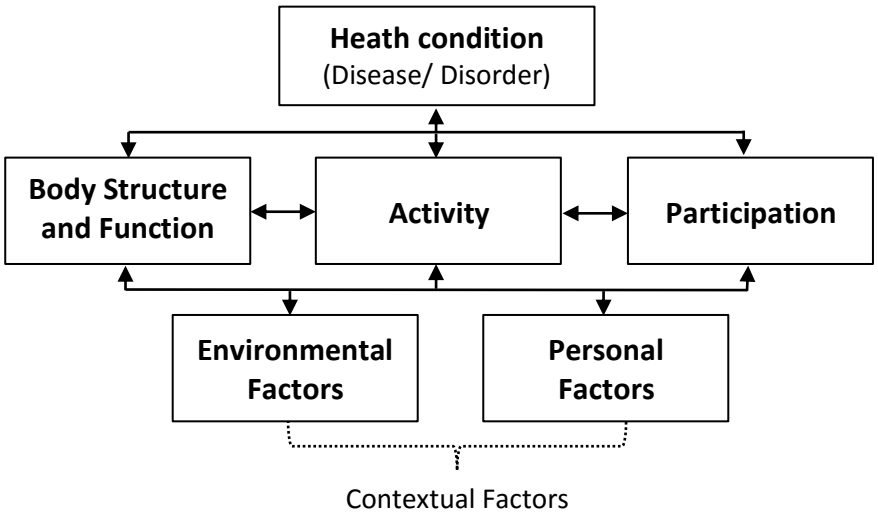

Figure 3 **Components of the ICF classification system[100]**

The ICF does not clearly distinguish 'activities' and 'participation' as separate domains. The literature on this matter acknowledge that users may wish to differentiate the terms, hence supplying the user with definitions and possible ways of using the respective domains activities and participation.[124]

This model has incorporated the biomedical model, psychological model and the social model to produce the 'biopsychosocial' view on disability.

Since its introduction, the ICF model has been widely established in professions as an organisational model of disability and as a classification platform for assessment, measurement tools and analysis and management of data.[119] The advantage of the ICF is that it provides an international language through its framework and coding systems thus allowing interpretation of data related to disability much easier. This model does come with the disadvantages such as the inability to identify causes of disability, ignores the contribution of QoL in its framework and inadequate weighting on personal and environmental factors that potentially have an impact on disability caused by disease. Yet the ICF is able to portray the dynamic course of function and disability and remains one of the best available conceptual frameworks to enable the understanding of the complex process caused by disability.[119]

1.2.5 Measurement of disability/ function

The limitation in function once acquired becomes one of many defining characteristics of the individual.[116] It has been highlighted in the literature that an accurate representation of function as such cannot be obtained from medical records.[125] In addition the need for functional health measures has been highlighted in the literature several times.[126] Impairment of function has been recognised to be an outcome of disease, but physiotherapists and those involved in rehabilitation may consider this a diagnosis, as their goal is to improve function.

Activities of Daily Living (ADL) are the standardised method of assessing disability.[127] ADL predominantly consist of self-care which is needed fundamentally for functioning; whereas instrumental activities of daily living are activities that allows for both fundamental and independent functioning.[128] These instrumental activities of daily living are what one would carry out as a representation of the population as a whole and therefore carry the fault of ecological fallacy. Results obtained from meta-analysis shows that there is a high correlation between the severity of impairment and ability to perform ADL. Severe impairment is not strongly correlated to QoL.[129]

Assessment of function enables professionals to gain insight into the patients' disability.[130] It provides invaluable information regarding the effectiveness and the quality of care provided to the patient and the outcome.

1.2.5.1 Outcome measurement

There are two methods in which functional capacity can be assessed. They are self-reported measures (patient reported outcome measures) and performance based tasks. In addition to function, disability can be assessed by evaluating the degree of joint destruction utilising physical scores or imaging scores.

1.2.5.1.1 Patient Reported Outcome Measures

1.2.5.1.1.1 Overview

Historically, the quality of medical care was assessed using measurement tools that had been developed by researchers. These instruments were based on medical models that in most instances were only understood and appropriately interpreted by healthcare professionals themselves. This method of evaluation led to the loss of crucial information about treatment effects experienced by the patients.[131] Since the advent of Patient Reported Outcome Measures (PROMs), the paradigm shifted to provide equal priority to the patients' views rather than focussing solely on

pathophysiological measurements. As patients remain the better judges of their impairment, they have a central role in managing their disease and provide feedback with regards to the quality of medical care received. This patient centred approach remains a fundamental component of patient care.[132] PROMs are defined as "a measurement based on a report that comes directly from the patient about the status of a patient's condition without amendment or interpretation of the patient's response by a clinician or anyone else"[131, 133].

Patient perspective and experience as obtained from a PROM are important factors affecting the decision-making process when monitoring the effectiveness of a treatment. PROMs provide a unique perspective into the observable concepts whilst enabling measurement of non-observable concepts.

PROMs are used in the standards of quality improvements while encouraging decision making with regards to patient care. It has the additional benefit of enabling patients to voice their opinion and interpret health status leading to higher accountability, greater patient satisfaction and autonomy when compared to other proxy measures.[134]

1.2.5.1.1.2 History

PROMs were developed for use in the National Health Service (NHS), and endorsed by the UK government department of health.[135] Initially developed for research purposes, PROMs facilitated comparison between the quality and performance offered by healthcare providers for resource allocation.[136] The use of PROMs in research is believed to be the only evidence-based tool available to doctors in obtaining a patient's perspective on disease management.[132] Since its mandatory introduction in elective procedures nationwide in the NHS in 2009, an improvement in the doctor-patient relationship, a significant change in the delivery of healthcare, disease diagnosis and monitoring have been observed.[137, 138, 139, 140] This has led to PROMs being well-

established in routine practice especially in providing regulated patient focussed management by a range of healthcare professionals including clinicians as well as physiotherapists.[134]

1.2.5.1.1.3 Typology

PROMs constitute a family of instruments that have the ability to measure a variety of features. Valderas and Alonso[141] identified three main classifications that were consistent in the PROM literature,

- Construct (objective of the measurement)

  Example:

  - Symptoms
  - Functional status
  - Health perception
  - Health related quality of life
  - Other

- Population assessed

  Examples:

  - Age,
  - Gender,
  - Disease specific versus generic
  - Culture

- Measurement model implemented

  Examples:

  - Metrics [psychometrics, econometrics, clinimetrics, other]
  - Dimensionality [index, profile, index and profile]

— Adaptability [completely standardised, partially individualised, completely individualised].[141]

In addition PROMs have been classified according to the objective of assessment: descriptive, evaluative and predictive.[142]

1.2.5.1.2 Performance measures

As the name suggests performance measures require the patient to perform certain tasks, usually in the clinic, which are objectively scored by an independent observer, who is often a healthcare professional.[143] These measures have been produced to address the recall bias associated with PROMs.[144] These instruments like any other, evaluate function by using predetermined criteria that has been validated. There are various types of performance measures. Some assess the ability to execute a task, while others subjectively assess the level of difficulty required to perform a task. Another performance measure of physical function assesses the time needed to perform a task. In addition to physical functioning, cognitive function and visual functioning can be assessed this way.

1.3 Definitions of concepts

1.3.1 Theories of methodology

Conceptual frameworks

Conceptual frameworks are based on basic scientific principles and models and guide terminology measurement and hypothesis. These form the foundation on which research and clinical care are built.[114]

Psychometrics

Psychometrics focus on the psychological aspects involved in the development of an instrument. Two main theories exist within this category namely the classical test theory and item response theory.[145-147]

Clinimetrics

This principle was introduced by Feinstein who defined it as the 'measurement of clinical phenomena' of an instrument. This view clashed with the psychometric view as its aim was to endorse clinical expertise to develop the measurement instruments.[148]

Novel approaches avoid specification of such psychometric or clinimetric terminology.[149] As such measurement properties should be assessed using the most adequate methods, models without specifications of a particular methodological theory are preferred.

1.3.2 Measurement models

Measurement models depict the causal relationships between the construct measured and its items. There are two main measurement models, reflective model and formative model.

- **Reflective model**

In a reflective model the construct manifests itself in the items. The construct is treated as a latent variable in addition to its error while the items are treated as effect indicators. A reflective model can be based on the classical test theory, reliability estimation and factor analysis. An example of a reflective model would be the construct ‘depression’. Here a person’s level of depression can be evaluated using variables such as low mood, anhedonia, and fatigue. As the construct ‘depression’

increases, so will the indicators. Therefore as items manifest itself in the construct these items are expected to correlate.[150, 151]

- **Formative model**

In a formative model the items form the construct. The theory that supports this model defines the construct as a composite variable, which is summarised by the causal indicators, the items. Incorporating the same example as above where 'depression' is the construct the following items could be used as causal indicators. These include, loss of a close relation, chronic disability or disease and loss of a job. Here each indicator will have an input to the construct and as each indicator alters so will the construct. As items do not share a common theme they do not necessarily correlate with each other.[150, 151]

The relationship between the items and the constructs are depicted in figure 3.

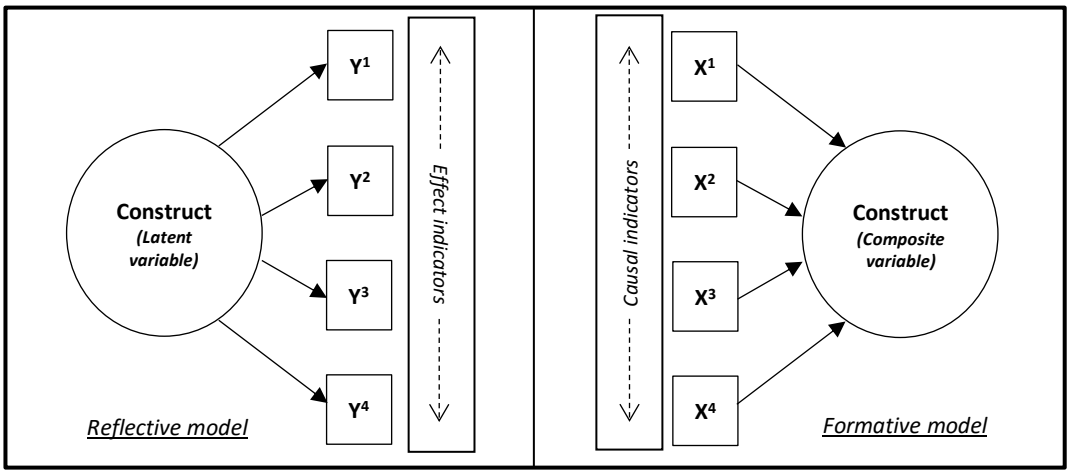

Figure 4 **Reflective model versus Formative model**

Coltman et al[150] compared both measurement models under the theoretical and practical domains.[150] Even though this study was primarily based on the business

literature, it can be applied to the medical context. Table 1 summarises these differences.

Table 1 **Reflective versus Formative measurement models[126]**

		Reflective model	**Formative model**
Theoretical differences	**Construct**	Construct exists independent of the indicators	Construct is determined as a combination of its indicators
	Causality	Causality occurs from the construct to items	Causality occurs from items to construct
	Items	Items manifests themselves in the construct therefore share a common theme, items are interchangeable, addition or removal of items do not change the conceptual framework of the construct.	Items define the construct, items not necessarily share a common theme, items are not interchangeable, addition or removal of an items will change the conceptual framework of the construct.
Practical differences	**Inter-correlation**	Items should have high internal consistency and reliability	Items can have high or low internal consistency and reliability but should have the same directional relationship
	Item relationships	Items have a significant relationship with those that precede the construct or those that follow the construct	Items may not have a significant relationship with those that precede the

			construct or those that follow the construct
	Measurement error and collinearity	Measurement error exists for each item and can be identified	Measurement error is not applicable per item but is applicable for the construct as a whole. Collinearity should be excluded.

1.3.3 COSMIN

COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) produced a critical appraisal checklist, (the COSMIN checklist) to evaluate the methodological quality of the measurement properties in a PROM.[152] This checklist was part of a larger study, the COSMIN initiative, that produced definitions, structure, and methods in the evaluation of health related patient reported outcomes. The COSMIN study was principally a Delphic study where a consensus was reached for the inclusions, definitions and assessment of measurement properties.[153] The definitions of measurement properties were produced as a part of the COSMIN taxonomy.[152, 153]

1.3.4 Measurement properties

As the scores obtained from PROMs are implemented in the decision making, investigations and management, the measures should be reliable and valid. Any failures in these would lead to imprecise assessment and biased decisions resulting in incongruous conclusions. As PROMs measure unobservable constructs it is vital that such tools contain adequate measurement properties as the consequences would be unfavourable.[152]

Measurement properties assess the qualitative nature of the measurement tool.[153] The definitions from the COSMIN taxonomy study[153] were selected as the foundation for providing definitions for terminology of the measurement properties.

The COSMIN taxonomy defined three quality domains that contained these measurement properties. The authors further went on to describe aspects of certain measurement properties.

The first domain, *reliability* was defined as the extent to which the measurement is absent from measurement error.[153] Here the extent to which scores for patients who have not changed should be the same for repeated measurements under several conditions. Definitions for the measurement properties under the reliability domain were as follows;

- **Internal consistency** was defined as the extent to which the items are interrelated.[153]
- **Reliability** was described as the proportion of the total variance in the measurement caused by changes in the consistency of the scores.[153]
- **Measurement error** was defined as error that is not attributed to true changes in the construct assessed but instead caused by the systematic and random error of a patient's score.[153]

The second domain, *validity* was defined as the extent to which a PROM relates to the construct it is expected to measure.[153] Definitions of the measurement properties of the validity domain are shown below;

- **Content validity** was defined as the ability to reflect the construct being measured in the PROM's content.[153] Face validity was described as an aspect of content validity.
    - Face validity was defined as the extent to which the items of a PROM represents the construct being measured.[153]

- **Construct validity** was described based on the assumption that the PROM validly measures the construct meant to be measured, it is the extent to which the scores of the PROM are consistent with the hypotheses.[153] Three aspects of construct validity were separately defined, structural validity, cross cultural validity and hypothesis testing.
    - Structural validity was defined as the extent to which the dimensionality of the construct being measured is reflected by the scores of the PROM.[153]
    - Cross-cultural validity was described as the ability of items to mirror the performance of the translated or culturally adapted PROM to the original PROM.[153]
    - Hypothesis testing was described as the characteristic assessing construct validity which takes the form of a hypothesis test.[153]
- **Criterion validity** was defined as the extent to which a score of a PROM reflects the score of its 'gold standard'.[153]

The third domain ***responsiveness*** stands alone as it consists of the measurement property 'responsiveness'. In both instances responsiveness was defined as "the ability of a PROM to detect change over time in the construct to be measured"[153].

The exact definitions used in the COSMIN taxonomy are available in Appendix 1.

Figure 5 **COSMIN taxonomy[129]**

1.3.5 Other properties

These properties have not been the focus of a vast majority of the literature which could possibly be due to its straightforwardness or to the fact that it is actually overlooked by most and is taken for granted.

Interpretability, was defined in the COSMIN taxonomy even though it was not classified as a measurement property. Fitzpatrick et al[154] highlighted acceptability, feasibility and precision as essential practical properties that should be possessed by a PROM.[154] Ever since, these properties have been cropping up in the literature.

The International Society for Quality Of Life research (ISOQOL) [155] recommended a minimum requirement checklist for the evaluation of measurement properties of a

PROM. Those not covered among the measurement properties in the COSMIN include conceptual and measurement models and burden.

Upon definition of each property it becomes evident that there is no consensus among these terms and frequent overlap in the definitions occur (e.g. burden and acceptability, burden and feasibility). This is in contrast to the measurement properties, which is probably due to the lack of a study similar to the COSMIN on the other properties.

Interpretability, is the extent to which a person can accredit qualitative meaning to the score of a PROM. [153]

Acceptability, is the willingness or ability of patients from the target population to complete the questionnaire.[154]

Precision as described by Fitzpatrick et al[154] indicates how precise the scores of the instrument are and its ability to make distinctions made by an instrument. Previous authors noted that this aspect has also been known as sensitivity.[156]

Feasibility, this assesses the burden on the healthcare professional. It considers the time and cost of administration, speed and ease of scoring and feedback of information and interpretation for clinicians, researchers and other staff.[154]

Conceptual and measurement model[155]

Conceptual model- Provides a framework that describes the desired construct to be included in a PROM.

The measurement model- Outlines the individual items in the PROM to the construct to depict its relationship to each other.

Responder and administrator burden-This was defined by the ISOQOL study as the time, effort, and other demands placed on those to whom the instrument is

administered (responder) or on those who administer the instrument (investigator or administrator).[157] Such factors include the questionnaire length, time constraints when filling in the questionnaire and the need for physical help when filling in the questionnaire.[131] Other ways to minimise burden include ensuring that the questionnaire is written in a way that can be comprehended by a 12 year old.[158] To assess the reading ease of a questionnaire in English, Flesch-Kincaid readability tests have been produced.[159] This includes the Flesch reading ease score which assesses the word length and the sentence length. The Flesch-Kincaid grade level translates the reading ease score into a grade level of the United States. The results are able to provide a reading age range for the questionnaire.[159] Patient burden closely ties in with acceptability of the questionnaire as described by Fitzpatrick et al.[154] Part of investigator burden is the ease of access. The requirements of copyrights and associated cost issues will all play a role in the actual practical process of obtaining a PROM.

1.4 Outcome measures in haemophilia

There are a variety of outcome measures used in haemophilia for both the paediatric and adult populations. Some of those have been solely produced for haemophilia (disease specific) whereas others were produced as generic scores for other diseases but have been used and/or validated in haemophilia.

Scores to measure outcomes have been developed in various categories and include, physical examination scores, imaging scores, functional scores, QoL measurement, mortality, bleeding frequency and cost utility measures. Boehlen et al[29] carried out a systematic review in 2014 to identify all the outcome measures used amongst

hemophiliacs over a period of ten years. This study was used as a guide to summarise the existing scores under the categories of physical scores, imaging scores and QoL measures.

1.4.1 Physical examination tools

Boehlen et al[29] identified two chief physical scores, the WFH Physical Examination score (also known as the Gilbert score)[160] and the Haemophilia Joint Health Score (HJHS)[161] .

The Gilbert Score[160] results in joint specific scores as well as an overall score. This questionnaire is available in English, Swedish and Dutch and takes about 30-45 minutes to complete. This clinician-administered measure does not require special equipment apart from a goniometer and a tape measure but does require administrator training. The Gilbert score has been tested on samples of children, but a full validation process, assessing its measurement properties such as reliability and responsiveness has not been done. This score is well adapted for the use where there is limited access to prophylaxis and factor replacement that comes with the drawback of being somewhat insensitive to mild joint disease.[29, 160]

The most up to date version of the HJHS provides joint specific scores, global gait scores and a total score. It is available in English, Dutch, Swedish and Mandarin and requires no special equipment. Administration of the HJHS by a clinician can take 45-60 minutes and the administrator needs to be adequately trained. This is predominantly designed for children with haemophilia aged 4-18 years and interpretation of values should be made according to the reference value for each age group.[162] The HJHS is more sensitive than the Gilbert score and is able to detect early signs of joint damage, hence proving useful in the monitoring of those on prophylaxis. The first version was tested and validated in children but has not been studied in adults or those with severe

disease.[163] The HJHS carries excellent reliability, correlates well with WFH score and correlates highly with radiographic features of destruction.[29, 164, 165]

1.4.2 Imaging scores

Another form of clinician reported assessment is radiological imaging scores. These enable deduction of the degree of joint destruction in haemophilic arthropathy in addition to being a mode of investigation of assessing complications. The common evaluative techniques identified in Boehlen et al[29] are X-ray, ultrasound and Magnetic Resonant Imaging (MRI).

The oldest method used to assess joint damage in haemophilia is X-ray which still remains a useful method of monitoring progressive joint disease but it is unable to detect early changes. The two chief scores implemented in practice are the Arnold-Hilgartner system[67] and the Pettersson scale [166, 167]. The Arnold-Hilgartner system is a progressive score that is simple and is more user friendly while the Pettersson score is an additive score thus more difficult. Both of these scores have good intra-observer reliability and inter-observer reliability.[168] Conventional radiology, has the disadvantage of being unable to detect osteochondral pathology in early disease.

Ultrasound allows examination of both soft tissues as well as cartilage surfaces acts. It is easily accessible but it is operator dependent. Ultrasound can be utilized to detect early arthropathy and has used as a diagnostic tool.[169] There have been scores undergoing development that look into standardising these scores.[170,171] The ultrasound scores have been known to correlate well with the bleeding episodes.[172]

There are two established MRI scores in today's practice, the Denver MRI score[173] and the European MRI score[174]. The European score gives better evaluation of soft tissue and osteochondral changes but comes with the disadvantage of being more complex.[174] There have been many more MRI scoring systems used in the literature

making it difficult to compare results across cohorts. The International Prophylaxis Study Group (IPSG) have aimed to bridge this gap by standardizing MRI interpretation.[175] Their score is a combination of a progressive and additive scale, which is highly reliable and has a low correlation with clinical parameters but does not allow discrimination between the severities of haemophilia.[174-177]

MRI is advantageous compared to other methods, as it allows better visualisation of soft tissue and cartilage changes, and being free of ionising radiation. MRI is the gold standard to detect early joint changes, staging and follow-up. MRI as a whole is shown to have good reliability.[178, 179] Drawbacks associated with MRI are that it is expensive, not easily accessible and requires sedation in young children.

In addition to the above imaging techniques computer tomography, scintigraphy and positive emission tomography have been used in haemophilic arthropathy but have proven to be of limited use in the follow up of the disease. Imaging scores although accurate in detecting subtle changes, are not reflective of disability.

1.4.3 Quality of Life measures

The subjective and personal nature of QoL was described in the WHO definition of QoL as "an individual's perceptions 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"[180]. Researchers have later gone on to develop the term Health Related Quality of Life (HR-QoL) which is defined as the impact of health on QoL as it is accepted that medical systems are unable to address all concerns inclusive in QoL.[181]

HR-QoL has become an important part of the way clinicians assess the outcome of disease. In general, HR-QoL instruments measure more relevant health outcomes such as physical, mental and social wellbeing as it is a complex multi-domain construct.[132]

There have been many generic questionnaires produced to measure a patient's HR-QoL. Not many have been produced in the English language that are validated to be used in haemophilia.

As highlighted by Boehlen et al[29] majority of the QoL measures specific to the haemophilia population are difficult to implement in the English speaking community. This is due to the fact that there is insufficient evidence on the translations and cultural validation. The Hemofilia-Qol[182] was developed in Spain and is only validated in Spanish. The Qual-Hemo[183] was developed in France and caters to a wide range of age groups and to date there appears to be no validated translation in English available. The Qual-Hemo does show sufficient convergent validity with a subscale of Short Form-36 (SF-36) questionnaire. The Haemolatin-QoL[184] questionnaire developed by a Latin-American group is available in Spanish and Portuguese but unavailable in English.

The Haem-A-QoL[185] was originally developed in Italy and consists of 46 items compiled from focus groups of patients with haemophilia in different countries. Here the Cronbach's alpha ranged from 0.74-0.88 and there was high convergent validity with the SF-36 questionnaire. The Haem-A-QoL is validated in Germany, Hungary and the UK and linguistically validated into 17 languages. It consists of good measurement properties particularly, reliability, convergent validity and discriminant validity.[185]

Haemo-QoL-A was developed and validated in 2008 by Rentz et al[186] in a multicentre-multinational study. This is the adult form of the Haemo-QoL questionnaire. Item generation was based on a literature search, and clinical input from patient and healthcare professional focus groups in North America and Europe. Cross-cultural and linguistic validation took place across 4 cultures and 3 languages. This PROM consists

of 6 subscales which are; physical functioning, role functioning, consequences of bleeding, worry, emotional impact, and treatment concerns. There are a total of 41 questions to assess the patients HR-QoL. It has good internal consistency and in addition correlates well with the SF-36.[186]

For the paediatric population the Haemo-QoL and the Children Haemophilia Outcome (CHO)-Kids Assessment Tool (KLAT) questionnaires exist. The Haemo-QoL is available in its original form[187] and also in a shorter version which is known as the Haemo-QoL index[188]. This questionnaire is available in English, French, Italian, German, Dutch and Spanish. Initial studies showed satisfactory reliability and validity testing.[187] Subsequently the questionnaire was validated in 6 other countries which showed acceptable results in reliability and construct validity.[189] The CHO-KLAT[190] was produced by a Canadian group and is available for three age groups.

A systematic review carried out by Szende et al[191] in 2003 aimed to identify HR-QoL instruments used among adult persons with haemophilia. The assessment of the results identified that the SF-36, SF-12, EuroQol-5D, health utilities index, and arthritis impact measurement scale-2 were the main general- non-specific instruments used.

1.5 Hypothesis

The hypothesis for this study is that the existing functional scores in the adult persons with haemophilia are suboptimal in their developmental methodologies and that they are inadequate in its measurement and other properties.

1.6 Aim

The aim of this study was to carry out a systematic review of all the available disease specific functional measurement instruments used in the adult English speaking haemophiliac population.

2 CHAPTER 2- Methods

2.1 Literature search

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance for systematic reviews was followed throughout this study.

Multiple databases, PubMed and Scopus databases were searched using certain Medical Subject Heading (MeSH) terms and were combined with Boolean terms to expand the search. These search terms were included in such a way that it could be expanded ensuring inclusion of relevant synonyms, alternative spellings and related terms. Backward referencing of eligible articles were carried out to include an increased number of relevant studies.

The search was not limited to a specific year range, in order to enable inclusion of a wide array of articles.

All searches carried out in both search engines were limited to human studies that are in English. The literature search was last carried out on the 2nd of August 2016.

MeSH terms used are shown in table 2. A MeSH term from column 1 was used in combination with column 2 and column 3 using the Boolean term 'AND'.

Table 2 **MeSH terms used**

MeSH 1	**MeSH 2**	**MeSH 3**
Haemophilia	Function	Score
Hemophilia		Assessment
		Measure
		Outcome

It was ensured that all mentioned instruments were developed for haemophilia (disease specific), assessed the construct function in adults and assessed overall joint function as opposed to just the function of a single joint. The severity of haemophilia are those with moderate to severe haemophilia as this is the group that have spontaneous haemarthrosis.

2.1.1 Inclusion criteria

1. Adult persons with haemophilia
2. Instruments developed for persons with haemophilia (disease specific instruments)
3. Instruments that assess the construct 'function'.
4. Data on PROMs or performance based assessments.

2.1.2 Exclusion criteria

1. Studies that assess the physical examination scores or imaging scores
2. Joint specific scores

This is further clarified in table 3 below, which explains the rationale for inclusion and exclusion.

Table 3 **Inclusion and exclusion criteria**

	Description	Rationale
Inclusion Criteria	Adult persons with haemophilia	The adult population has a variety of comorbidities affecting function with varying degrees of joint disease. Which is separate from the paediatric population who have access to prophylactic factor concentrates and less comorbidities. Type of haemophilia of interest are the congenital forms (haemophilia A and B).
	Disease (haemophilia) specific instruments	Generic instruments contain items that are irrelevant to the target population which comes at the expense of the length of the instrument.
	Instruments that assess the construct 'function'	As per research question.
	Data on PROMs or performance based assessments.	Types of measurement instruments that assess the construct 'function'
Exclusion Criteria	Studies that assess the physical examination scores	Even though physical examination may seem similar to assessment of function, these two do not assess the same construct. A good physical examination score does not necessarily mean or indicate how well a patient is able to function.
	Joint specific scores	As haemophilia is a systemic disease it tends to affect more than one joint. Therefore a questionnaire should focus on function of all limbs.

2.2 Evaluation

Studies that were analysed were those that looked at function, joint procedures, and outcomes in persons with haemophilia. The eligible studies were searched for a mention of a functional score, this functional score was later individually searched to ensure eligibility.

Once the instruments were identified from the literature, each instrument was searched on an individual basis on the PubMed and Scopus databases to identify all relevant articles produced containing that instrument. This search was carried out to ensure that no article containing vital information regarding the piloting or validation of the instrument was lost. This search had no year restriction although it was restricted to human studies available in the English language.

The articles that provided information on the pilot study and/ or validation study were located to extract data and summarise instrument characteristics. Subsequent studies that assessed properties which were not assessed in the initial validation were included. Where an initial pilot study was carried out in those with larger validation studies, data from the pilot study was included only where the validation study did not assess all properties. Access to questionnaires was of paramount importance, where open access to the questionnaires were not available this was obtained by contacting the authors. It was ensured that the article was produced in the English language and that the instrument had a version in English. Where full measurement properties in the English version were unavailable the original version was evaluated whilst summarising the results of the English instrument. Each study and questionnaire was evaluated for its development methodology, measurement properties as assessed by the COSMIN checklist and certain other properties which will be described below.

2.2.1 Development of instruments

The methodology utilised in the development of instruments is often neglected in the literature. This stage plays a vital role in the quality of items included in the instrument. Methods of development are variable depending on the discipline and is subjected to the judgement of the developers. The basic aim is to generate items which could be tested in subsequent pilot and validation studies.[149] This systematic review summarises the methods in which each instrument was developed. This included giving particular attention to the groups of participants used for item generation.

2.2.2 Measurement Properties

Measurement properties were evaluated using the COSMIN 4-point checklist. As introduced previously, the COSMIN checklist covers standards for the defined properties of health measurement instruments in terms of its design and statistical methods.[192] The content of the checklist with 4-point rating scale consists of twelve sections out of which ten sections are used to assess the methodological quality. Here the measurement properties as defined by the COSMIN taxonomy; internal consistency, reliability, measurement error, content validity, structural validity, hypothesis testing, cross cultural validity, criterion validity and responsiveness are assessed. The 4-point rating scale enables to calculate an overall methodological quality score for each measurement property which can take a value of excellent, good, fair or poor. Under each measurement property there is a variety of criteria that are rated on the 4-point rating scale. Each measurement property is scored by employing the lowest score in the criteria under the measurement property in the study. In addition to the defined measurement properties assessment of interpretability of the instrument is available in the checklist.[192] It should be noted that this is not given a quality score. Evaluation of overall characteristics include general requirements under the categories

of generalisability and item response theory methods.[192] Prior to evaluation of each questionnaire, the author read the COSMIN checklist manual which is available on the COSMIN website.[193] The author of this study evaluated the measurement properties against the COSMIN 4-point checklist and where there were uncertainties discussions were held with two statisticians in the department.

2.2.3 Other properties

Data extraction for interpretability (table 4) is supported through the COSMIN checklist. In addition, as the COSMIN checklist provides a generalisability data extraction box, this information was extracted. Thus these properties were summarised using the COSMIN checklist as a template.

As there is an overlap between acceptability and patient burden, and feasibility and investigator burden, the properties acceptability and feasibility will be assessed within the property burden. Precision of an instrument was assessed by analysing the categories and the number of items on the response scale.

A conceptual model of each instrument as described by the respective authors are summarised. Here, the approach used by each author to describe the construct 'function' was explored. A measurement model where mentioned would be stated as a formative or reflective model. Where this is ambiguous a 'thought test' is carried out to depict the direction of change of items when the construct changes. When carrying out the 'thought test' if items change when the construct changes it is suggestive of a reflective model whereas if a change in the items caused a change in the construct it was a formative model.[193]

Burden consists of patient and investigator burden. Patient burden was assessed by the time required for completion, instrument length, response rate (as well as assessing

acceptability) and readability. Readability was assessed by calculating the Flesch reading ease score and the Flesch-Kincaid reading grade level for each instrument. (table 5) It should be noted that the readability was applicable for the English versions of all of the instruments to allow for uniformity in analysis. Investigator burden (feasibility) was assessed by the ease of access to the instrument, and training required to implement the instrument in practice.

Where additional properties are evaluated in each instrument, this will be summarised as 'other properties reported', even though no such evaluation of the properties will be made.

Table 4 **Data extraction for interpretability**

Interpretability
· Percentage of missing items · Description on how missing items were handled · Distribution of the total scores · Percentage of the respondents who had the highest possible total score · Scores and change in score (i.e. means and Standard deviations) for relevant (sub) groups · Minimal Important Change or Minimal Important Clinical Difference

Table 5 **Interpretation of readability scores**

Flesch reading ease score	**Flesch-Kincaid reading grade**	**Description**
90.0-100.0	5th	Very easy
80.0-90.0	6th	Easy
70.0-80.0	7th	Fairy easy
60.0-70.0	8th-9th	Standard
50.0-60.0	10th-12th	Fairly difficult
30.0-50.0	College	Difficult
0.0-30.0	College graduate	Very difficult

3 CHAPTER 3- Results

3.1 Literature search

The databases PubMed and Scopus were used for the literature search, whereby 3752 and 1581 articles were obtained respectively. No articles were identified through backward referencing. Duplicates were identified and excluded which resulted in a balance of 2096 articles. Titles of articles were read and 1872 articles were excluded as they did not meet the inclusion criteria. Out of those 194 articles were excluded based on reading either the abstract or full article.

The process of the literature search and the results obtained are depicted in the flow diagram below (figure 5).

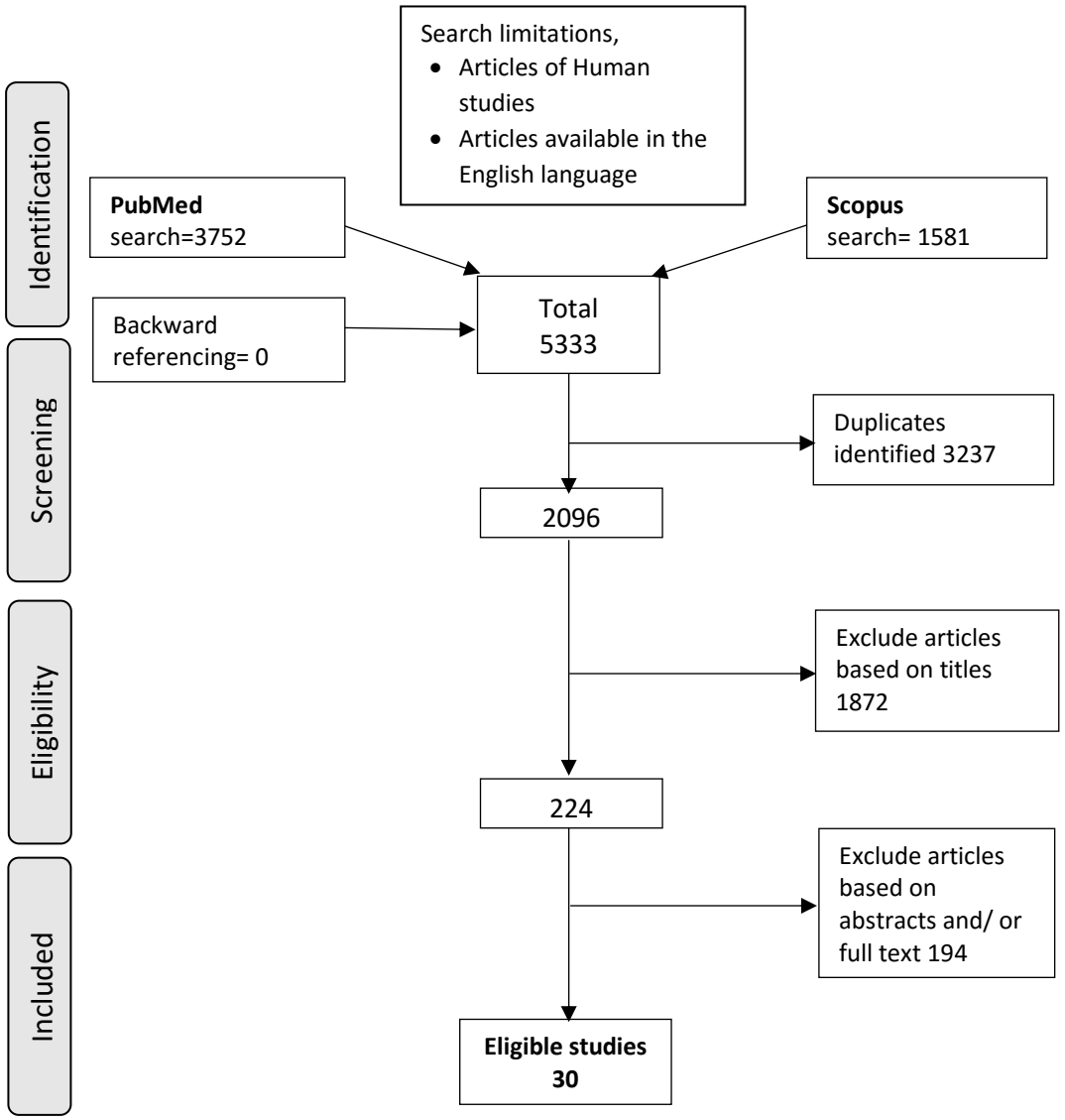

Figure 6 **Flow diagram of literature search**

The search yielded a total of 30 studies that had the mentioned eligible instruments. Among the 30 studies there were three instruments mentioned. Namely the Haemophilia Activities List (HAL), the Functional Independence Score in Haemophilia (FISH) and the Haemophilia Exercise Project- Test- Questionnaire (HEP-Test-Q).

The distribution of the studies according to the scores mentioned are shown in table 6 below.

Table 6 **Instrument distribution in literature search**

Instrument mentioned	Number of studies
HAL	9
FISH	7
HEP-Test-Q	2
HAL and FISH	10
HAL and HEP-Test-Q	2
Abbreviations: HAL – Haemophilia Activities List, FISH-Functional Independence Score in Haemophilia, HEP-Test-Q – Haemophilia Exercise Project Test Questionnaire	

The terms searched to identify all relevant articles per instrument were, "Haemophilia Activities List", "Functional Independence Score in Haemophilia" and "HEP-Test-Q". This search was last carried out on the 03 August 2016. The results of the individual searches are shown below.

Table 7 **Individual searches per instrument**

Instrument	PubMed search	Scopus search
Haemophilia Activities List	17	23
Functional Independence Score in Haemophilia	13	14
HEP-Test-Q score	5	5

The results identified that the principle scores used to assess function in adult persons with haemophilia are the HAL[194], FISH[195] and the HEP-Test-Q score[196].

In addition to the results obtained, generic scores that are not disease specific to haemophilia were identified by the initial search. Those instruments used in the adult haemophiliac population to assess function are the Western Ontario McMaster Questionnaire (WOMAC)[197], Stanford Health Assessment Questionnaire (HAQ)[198], the Arthritis Impact Measurement Scale (AIMS)[199] the Impact on Participation and Autonomy questionnaire (IPA)[200], and the Canadian Occupational Performance Measure (COPM)[201].

As the information regarding the development, and cross validation of the HAL questionnaire was unclear, the authors were contacted to obtain further information. In addition the authors of the HEP-Test-Q were contacted to obtain a copy of the questionnaire as it was unavailable in the public domain.

3.2 Evaluation

3.2.1 Haemophilia Activities List

HAL produced by Van Genderen et al[194] in 2004 is a PROM that evaluates a patient's functional status under seven domains. It is assessed by enquiring about a patient's ability to do the following tasks:

1. Lying/sitting/kneeling/standing (8 items)
2. Functions of the legs (9 items)
3. Functions of the arms (4 items)
4. Use of transportation (3 items)
5. Self-care (5 items)
6. Household tasks (6 items)
7. Leisure activities and sports (7 items)

This detailed questionnaire comprises of a total of 42 questions and the patient's answers are rated according to a 6 point Likert scale. This instrument was developed and validated in Dutch and is available in 9 other languages including English.[194] The article is available in English and provides measurement properties for the Dutch instrument, but makes no reference to the English version. This was clarified by contacting the author. This systematic review evaluates the properties of the Dutch version of the HAL as properties of the English HAL have not been tested.

The disadvantages of HAL as highlighted by previous authors include its lack of sensitivity to detect clinical changes and the fact that it is language dependent.[29, 35] The HAL pilot study and validated studies consisted of a sample size of 50[194] and 127[202] respectively. The English HAL is included in appendix 2.

3.2.1.1 Development of the HAL

The development of the HAL took place in outpatient clinics involving the participation of patients with haemophilia. Item generation took place by carrying out semi-structured interviews using the McMaster-Toronto Arthritis Patient Preference Disability Questionnaire (MACTAR) as an interview guide.[194] The MACTAR is an open ended questionnaire to identify activities that are deemed to be difficult for the patients.[203] The aim of each interview was to identify a maximum of 10 activities that the individual struggles with, subsequently ranking the top 5 activities.[194] This process was carried out on 162 patients and was halted once item saturation was reached. In those who took part in item generation, the distribution of the severity of the disease was, 73% (n=118) for severe, 10% (n=16) for moderate and 17% (n=28) for mild disease.[194] Majority of the participants had haemophilia A 86% (n=140), while 10% (n=16) had haemophilia B. In addition to those with haemophilia some participants had a disease status of Type B-Leyden disease (n=1), von Willebrand disease (n=4)

and factor VII deficiency (n=1).[194] The items generated were classified according to the ICF classification system. Items included were those that could be classified as 'Activities' in the ICF while those that were in the 'Participation' category were excluded. These items were transcribed into questions which included a response scale option on a Likert scale.[194]

3.2.1.2 Measurement properties

Internal consistency measured using Cronbach's alpha was 0.97 for the overall HAL questionnaire. The Cronbach's alpha ranged from 0.61-0.96 for the subscales of the HAL.[202] The COSMIN 4-point checklist[192] quality standard was 'poor' as the sample size contained less than five participants per number of items (item 6). The test-retest reliability was not assessed in the HAL score. As reliability was not assessed, the measurement error could not be evaluated.

The content validity of the HAL score was measured by asking patients and physical therapists if each item should be included and to rate the clarity of each item. Patients received the Dutch version of the questionnaire whilst the physical therapists received the English version of the questionnaire.[202] The HAL questionnaire proved to have good content validity evidenced by a score of 8.0 (±0.7) and 7.7 (±1.0) from patients and physical therapists respectively. This was rated on a ten point scale where 10 was very good and 1 was very bad.[194] This analysis was only carried out in the initial pilot study of the questionnaire. The content validity assessment utilising the COSMIN checklist[192] received a 'good' rating as it consisted of a moderate sample size (item 2) and as the purpose of the instrument was assumed (item 3).

The property construct validity that met the COSMIN definition was measured by validating the HAL against the Dutch–Arthritis Impact Measurement Scale 2 (Dutch-AIMS2)[204] and the Impact on Participation and Autonomy (IPA)[200]. This was referred

to as convergent validity by the developers.[202] The statistical test used to assess this variable was the Spearman rank order correlation coefficient. Construct validity with the HAL and the $AIMS2_{sum}$ (total score) and $AIMS2_{phys}$ scores (which asses the physical component of the scale) showed excellent results where correlations were 0.81 and 0.83 respectively.[202] The construct validity with the IPA_{sum} score showed a good correlation (0.71). All these results were statistically significant ($p < 0.001$).[202] Structural validity which is an aspect of the measurement property construct validity, was rated 'poor' in the COSMIN checklist[192] as it contained a sample size of less than five participants per item (item 4). Hypothesis testing which is part of the measurement property construct validity, was rated 'good.' This was because the percentage of missing items were not described (item 1), presumptions had to be made on how missing items were handled (item 2) and because there were a minimal number of hypotheses formed (item 4).

The original HAL was developed in Dutch, using the WHO's ICF classification scheme.[124] Translation of the Dutch items into English was made based on the available international versions of the ICF. Only the original HAL had undergone validation whereas the translated English version of the HAL had not undergone any formal translation or cross-cultural validation process. As the questionnaire that was evaluated was the Dutch version the cross cultural validation was not applicable to this version of the instrument. Since the HAL is a PROM there is no criterion validity assessment applicable to the instrument. Responsiveness to the change of the HAL was not assessed either at the pilot stage, at validation or any subsequent studies.

3.2.1.3 Other properties reported

The authors assessed what was described as 'construct validity' in HAL by calculating the variance extracted through explorative factor analysis. This defined the underlying

construct for the instrument.[202] The results showed that 6 items accounted for the 75.9% of the total variance. The three factors that were responsible for 50.6% of the total variance were the 'upper extremity', 'basic lower extremity' and 'complex lower extremity'. This aspect assessed by the authors tie in with the assessment of dimensionality of the internal consistency assessment.[150, 202]

The clarity and style of each question, paying close attention to the font size, number of items and sequence of domains were carefully assessed at the pilot study.[194] The results of the style of the HAL questionnaire as reported in the pilot stage by patients (Dutch version) and physical therapists (English version) was 8.2 (±0.4) and 8.1(±0.8) respectively. This was rated on a ten point scale where 10 was very good and 1 was very bad.[194]

The Spearman rank correlation coefficient for the relationships between three performance tests and the HAL questionnaire were calculated. The correlation between the button test (a performance based test that requires the patient to do up and undo the three top buttons apart from the top most button of a standard shirt) and domains of the HAL were not significant or had low correlations (r=0.39).[202] The highest correlations between the HAL domains were 0.66 for the 50 metre walking test, 0.65 for the timed up and go test and 0.61 the figure 8 walking test. The domain which had the highest correlation for each of the (latter three) performance tests was, 'lying/sitting/kneeling/standing'. [202]

3.2.1.4 Other properties evaluated

With the data provided only one item in the interpretability box of the checklist could be filled in. This item evaluated the distribution of the scores (42-252). Characteristics of the generalisability box are summarised in table 8. Assessment of the precision of the HAL specifies each activity and motion to be assessed and has 6 response options

based on the ability to perform a task. The conceptual model assessed by the HAL was to assess the functional health status. This was carried out by measuring the 'activities' category in the ICF definition of disability.[124] The 'thought test' suggests that the measurement model incorporated is a reflective model. Patient burden is higher than the investigator burden. The HAL is said to take approximately 10 minutes of the patient's time.[202] It is well laid out. As this questionnaire consists of 12 pages of text, which includes a page of an example, and how to score, this may be off-putting from a patient's perspective in practice. Response rate ranged from 87%- 93% for the HAL.[202] The Flesch reading ease score for the HAL questionnaire was 67.2 which was 'standard' and has a Flesch-Kincaid grade level of 7-8 graders.[159] The HAL has minimal investigator burden, as it comes cost free for those who require it. The questionnaire is available via the WFH website thus giving access to the public domain. (http://www.wfh.org/en/page.aspx?pid=875). This instrument requires no additional training as it is a PROM. The HAL has been in practice for almost a decade indicating that it has been taken up well by healthcare professionals.

3.2.2 Functional Independence Score in Haemophilia

In 2005, Poonnoose et al[195] produced the FISH score which is a performance based questionnaire to assess the functional capacity of patients with haemophilia. This instrument consists of 8 items and aims to assess ADL, under three categories of self-care, transfer and locomotion. The items that the patients are tested on are outlined below:

1. Eating and grooming
2. Bathing
3. Dressing
4. Chair

5. Squatting

6. Walking

7. Stairs (12-14 steps)

8. Running

These items are scored on a 4-point Likert score according to the level of assistance required when executing these actions in the presence of a healthcare professional. This instrument was not developed to assess challenging activities, and does not assess activities such as education or employment. As this is a performance based assessment, it can be used in different languages, but this does not allow for inaccuracies in translation by the administrator. The pilot study and validation studies in which the FISH was studied consisted of severe disease with sample sizes of 35[195] and 63[205] respectively. The FISH score is included in appendix 3.

3.2.2.1 Development of the FISH

Item generation stage involved the participation of patients as well as their relatives and therapists. Participants were asked to list activities that were affected by haemophilia. Generated items were classified according to the ICF classification system and only activities that could be assessed objectively or performed at an outpatient setting were included. Distribution and demographics of the participants in the development of FISH were not provided by the authors.[195]

3.2.2.2 Measurement properties

The validation study[205] of FISH score reported an internal consistency as measured by the Cronbach's alpha of 0.85. This received a 'poor' on the COSMIN checklist[192] as factor analysis was not performed to assess unidimensionality of the scale (item 5) and a Cronbach's alpha value was not calculated for each subscale separately (item 7). Reliability as assessed by intraclass correlation coefficient (ICC) was 0.98.[205] This

results in a ‘fair’ score for the reliability estimate as it is unclear how missing items were handled (item 2) and because of other methodological flaws in the study (item 10). As the reliability estimate given was not clear if it was for inter-rater or intra-rater reliability. The same reasons (apart from item 10) results in a ‘fair’ level for the measurement error property. The authors did not assess the content validity of the FISH. Nevertheless, they state that the FISH has good face validity due to the methodology of questionnaire development. Item generation incorporated both patients’ and healthcare professionals’ views. Content validity assessed in the COSMIN checklist[192] received a ‘poor’ rating as this aspect was not assessed (items 1, 2).[195]

The construct validity was assessed using the WFH score[160] and the Pettersson score[166]. The correlation coefficients were -0.61, and -0.38 respectively. This meets the criteria for structural validity and was rated ‘poor’ as no exploratory or confirmatory factor analysis was performed (item 6). Hypothesis testing was rated ‘fair’ as the hypothesis was not formulated but it was possible to deduce what was expected (item 4). The FISH is available in English and it is a performance-based assessment. There has been no cross cultural validation that has been carried out to evaluate its properties in different cultures other than the Indian culture. The authors of the FISH assessed criterion validity by using PROMs to assess function. This included the HAL[202], HAQ[198] and WOMAC[197] scores which showed correlation coefficients of, -0.66, -0.75 and -0.66 respectively. The criterion validity was given a rating of ‘good’ as it had an adequate sample size (item 3) and as it was presumed that the criterion was considered an adequate gold standard (item 4).

The responsiveness was assessed by evaluating the difference in those who underwent surgery to correct a fixed flexion deformity of the knee.[205] A Wilcox signed ranked

test was carried out and the standardised responsiveness mean of -1.93 was achieved for the FISH although the clinical score, the WFH score failed to detect any changes post- intervention. Change in scores was detected only in the PROMs the HAQ and the WOMAC.[205] The responsiveness as rated on the 4-point COSMIN checklist[192] was 'poor.' This was due to the small sample size (item 3) and due to the lack of description of the time interval (item 5) in which responsiveness to change was assessed. There were no additional properties that were reported by the authors of the FISH score.

3.2.2.3 Other properties evaluated

The evaluation of the interpretability of the scores enabled only two aspects to be completed, which were the distribution of the total scores (7-32) and the percentage of respondents with the highest possible score (9.5%). Results of the generalizability are summarised in table 8. Precision for the FISH is poor in that it only has 4 response options which allows for scoring the level of assistance required whilst performing a task.

The manuscript for the questionnaire provides moderate information regarding the conceptual model. It only defines the construct as function, in a performance based assessment context. Thus this vague description of a construct entails moderate evaluation of this property. The measurement model of the FISH was not defined but it was defined as a reflective model since the instrument assesses ADL and as the 'thought test' indicated that the items changed as the construct changed. The FISH carries more investigator burden than it does with regard to patient burden. As the FISH is a performance based instrument it takes 15 minutes to complete the 8 items and the response rate of the FISH was 100%. The Flesch reading ease score was 58.4, which means 'fairly difficult to read', and the Flesch-Kincaid grade is between 10th-12th graders.[159] The FISH questionnaire is feasible and has been accepted as it has been

implemented in practice for almost a decade. The FISH questionnaire is available to the public, and is available via the WFH website. (http://www.wfh.org/en/page.aspx?pid=875) Since the FISH is a performance based instrument it requires training of the investigator to ensure correct administration and assessment of the questionnaire. This training comes at no cost in the guidance to users of the instrument.

3.2.3 HEP-test-Q score

This score was produced by a German group led by von Mackensen et al[20] as a part of the Haemophilia and Exercise Project (HEP) in 2009. This HEP-Test-Q is PROM based on motor abilities. It can be applied in sports therapy where the focus is on prevention of injury and to rehabilitate. This assesses physical status under the categories of

1. Mobility
2. Strength and coordination
3. Endurance
4. Body perception (wellbeing, exposure of stress, self-esteem).

It consists of 25 items and the answers are filled in a 5 point Likert scale of 1 (never) to 5 (always).[20] The HEP-Test-Q is linguistically validated in German, English and Italian. Additional versions in Dutch, Greek, French and Spanish are available.[20] This systematic review evaluated the German version of the HEP-Test-Q. The HEP-Test-Q pilot validation studies had sample sizes of 24 and 43 patients respectively.[196] The English HEP-Test-Q is shown in appendix 4.

3.2.3.1 Development of the HEP-Test-Q

Item generation involved selection of items using professionals in sports medicine. Further information on these participants and methods were not available.[196]

3.2.3.2 Measurement Properties

The total value for internal consistency as calculated by Cronbach's alpha was 0.96. This varied from 0.85-0.92 across each dimension assessed.[196] Internal consistency as evaluated by the COSMIN 4-point checklist[192] received a rating of 'poor' as factor analysis was not performed or as there was no reference to another study (item 5) and the sample size was <5 participants per item for the assessment of unidimensionality (item 6). Test re-test reliability was assessed using Pearson's correlation coefficient which received a value of 0.90 ($p < 0.001$) for the total score.[196] This parameter was assigned a 'fair' due to the sample size included in its assessment (item 3) and because even though the correlation coefficient was calculated there was no evidence that systematic change had not occurred (item 11). The measurement error for absolute measures was administered a 'fair' rating due to its sample size (item 3).

A content validity study was carried out at the pilot stage where patients rated for the preliminary version of the score on a scale from 0 (not important) to 100 (very important) which gave a mean score and standard deviation of 66.8 ±19.6 (range 12-98).[196] Thus the evaluation of the pilot version of the questionnaire gave a 'good' rating for the content validity as the purpose of the instrument was not described but assumed (item 3).

Convergent validity as described by the authors was tested against the Haem-A-QoL[185], HAL[202] and the SF-36_{PCS}[206] scores. The correlation coefficients for such measurements and the total HEP-Test-Q score were -0.82, 0.76, -0.71 respectively.[196] Convergent validity defined by the authors met the COSMIN definition of structural validity.[153] This scored 'poor' as the sample size in the analysis included <5 participants per item (item 4) and no exploratory or confirmatory analysis was performed (item 6). Hypothesis testing in the COSMIN checklist[192] was rated as 'fair'

because of the moderate sample size (item 3) and the hypotheses were not formulated but it was possible to deduce what was expected (item 4). Cross cultural validity is not applicable to the original German version of the HEP-Test-Q whereas criterion validity is not applicable to this as it is a PROM.

Responsiveness was assessed in a separate study that evaluated those patients that underwent regular hydrotherapy. At 6 and 12 months follow up the physical examination score, the WFH Orthopaedic Joint Score (OJS) detected a significant change (6 months $p<0.035$,12 months $p<0.024$). However, the HEP-Test-Q was unable to detect a significant change in follow-up during both follow up times. This was rated 'poor' on the COSMIN 4-point checklist[192] because of the small sample size (item3).

3.2.3.3 Other properties reported

Authors evaluated discriminant validity in those with certain clinical characteristics. Significant differences were only associated with age (≤40 years versus >40 years), Hepatitis A (yes versus no), HBV (yes versus no) and the number of target joints (0 versus ≥1).[196]

In a subsequent study the HEP-Test-Q score was compared against objective measurements which included: ROM, one leg stand and the 12- minute walk test.[207] Spearman rank correlation coefficients were calculated for the relationships which were all significant between the total HEP-Test-Q at the level of $P<0.001$. The highest correlation was in the 12 minute walk test (0.757) while the lowest correlation was the left sided one leg stand (0.403).[207]

3.2.3.4 Other properties evaluated

Interpretability of the final HEP-Test-Q was poor as it did not provide any data for the reader to interpret values from the validation study. Results for the generalisability of

this instrument are summarised in table 8. The precision of the scores are rated on a 5 point scale based on the ability to perform a task and agreement with a statement. The conceptual model adopted in the HEP-Test-Q was depicted as a sports oriented score to assess function However, the measurement model was not clearly stated by the authors. The 'thought test' suggested that this instrument had elements of both reflective and formative models. With regards to burden this instrument takes an average of 14.4 ± 9.6 minutes to complete the 25 items. The response rate was 79%.[196] The Flesch reading ease score was 56.3, which is 'fairly difficult to read' and the Flesch-Kincaid grade level was between 10^{th}-12^{th} grades.[159] Even though this PROM is relatively new, it is feasible and accepted in those whom it has been used on. The HEP-test-Q is currently not freely available to the user thus limiting its use to those interested. However, upon contacting the author this PROM could be requested. As this instrument is a PROM it requires no training.

3.2.3.5 English Version

In the literature of the HEP-Test-Q score it is stated that the original was developed in German and that additional translated versions including the English version was validated. Thus a broader search was attempted to find this validated English version of the HEP-Test-Q. A poster that was presented at the WFH world congress in 2014 identified a partial validation study.[208] The results of this English HEP-Test-Q are presented below. The full COSMIN checklist was not applied to this PROM as it did not evaluate most measurement properties. Instead the cross cultural validation box in the 4-point COSMIN checklist[192] was applied to evaluate the integrity of cross cultural validation of this translated instrument.

The translation process undertaken was a forward/backward method and the expertise of those involved in the translation was not described. The translated instrument was

administered to English speaking communities in the United States and the UK.[208] The total of participants were 38 (total mean age 15.8 ± standard deviation 9.7) where 9 were adults (mean age 29.8 ± standard deviation 10.2) 29 were children (mean age 11.5 ± standard deviation 3.4). The internal consistency as assessed by Cronbach's alpha value was 0.916 and this ranged from 0.729-0.891 across the dimensions of the HEP-Test-Q score.[208]

When applying the cross cultural validity box in the COSMIN checklist[192] the translated version scored 'poor'. As the sample size included was < 5 participants per item (item 3), the instrument was not pre-tested in the form of cognitive interviews (item 10), the samples were not similar for all characteristics except language or culture (item 12) and confirmatory factor analysis was not performed in this population (item 14).

Table 8 **Generalisability results of studies**

Instrument/ **Author**	Year	Language evaluated	Country of development	Instrument type	Sample size	Mean age (SD)	Haemophilia		Disease severity			Study setting
							Type B (%)	**Type A (%)**	**Severe (%)**	**Moderate (%)**	**Mild (%)**	
HAL/ **Van Genderen**	2006	Dutch	Netherlands	PROM	127	42 (12.5)	17 (13)	110 (87)	127 (100)	-	-	Secondary care
FISH/ **Poonoose**	2007	English	India	Performance based assessment	63	14 (8.3)	5 (8)	58 (92)	63 (100)	-	-	Secondary care
HEP-Test-Q/ **von Mackensen**	2010	German	Germany	PROM	43	44 (11.5)	3 (7)	40 (93)	38 (88)	3 (7)	2 (5)	Secondary care

Abbreviations: SD- Standard Deviation, HAL- Haemophilia Activities List, FISH- Functional Independence Score in Haemophilia, HEP-Test-Q- Haemophilia Exercise Project-Test-Questionnaire, PROM- Patient Reported Outcome Measure

Table 9 **Evaluation of measurement properties using the COSMIN 4-point checklist**

Instrument/ **Author**	**Internal consistency**	**Reliability**	**Measurement error**	**Content validity**	**Structural validity**	**Hypothesis testing**	**Cross cultural validity**	**Criterion Validity**	**Responsiveness**
HAL/ **Van Genderen**	Poor (6)	N/A	N/A	Good (2,3)*	Poor (4)	Good (1,2,4)	-	-	N/A
FISH/ **Poonoose**	Poor (5,7)	Fair (2,10)	Fair (2,6,8)	Poor (1,2)	Poor (6)	Fair (4)	-	Good (3,4)	Poor (3,5)
HEP-Test-Q/ **von Mackensen**	Poor (5,6)	Fair (3,11)	Fair (3)	Good (3)	Poor (4,6)	Fair (3,4)	-	-	Poor (3)

The number in brackets () indicate the item number on the 4-point COSMIN checklist that corresponds to the lowest rating obtained.
*** Pilot study results**
Abbreviations: HAL- Haemophilia Activities list, FISH - Functional independence score in Haemophilia, HEP-Test-Q- Haemophilia Exercise project test Questionnaire,
N/A - not assessed ;'–' not applicable
Underlined numbers mean that the item number on the 4-point COSMIN checklist after excluding items that are affected by a small sample size

4 CHAPTER 4- Discussion

4.1 Literature search

The literature search was broad in terms to detect all the available instruments used to assess function in adult persons with haemophilia. Those selected in the systematic review were instruments that were disease specific for the adult persons with haemophilia. Disease specific instruments have the advantage over the generic instruments as they contain items that are sensitive to the disease and higher content validity thus having a high discriminating value. Whereas patients are more likely to lose interest by lengthy instruments or distracted by non-specific items, both features of generic instruments. This leads to high acceptability. In addition such disease specific measures are the most responsive to clinical changes in the disease.[209] The construct of interest was function, which resulted in the inclusion of PROMs and performance measures. The physical examination scores were excluded as they do not detect function. The construct function was decided as the focus as physical examination scores do not necessarily correspond to a patient's ability to function and perform tasks. Joint specific scores were excluded as haemophilia is a systemic disease affecting multiple joints. Therefore, it would only be right to identify instruments that evaluate function of all limbs.

4.2 Critical appraisal tool

Since the implementation of measurement instruments, there has been a large array of tools produced to assess such instruments.[154, 210] The most recent and relevant examples of such tools, are the COSMIN checklist [152] and the International Society for

Quality Of Life research (ISOQOL)[155] minimum requirements for the evaluation of measurement properties of a PROM. In 2013 ISOQOL[155] published a set of minimum requirements that should be met in PROM research indicating that if PROMs do not reach the recommended minimum standards it should not be considered appropriate for use in clinical practice. However, the authors go on to state that this does not hinder the instrument developers from adopting a maturation model to further validate and strengthen the questionnaire. The ISOQOL standards[155] were based on guidelines obtained from expert opinion and literature, which included the COSMIN checklist. This highlighted that when selecting a PROM the process of its conceptual and measurement model, reliability (internal consistency reliability, test-retest reliability), validity (content validity, construct validity, criterion validity and responsiveness); interpretability of scores and patient and investigator burden should be considered.[155] Definitions of each minimum standard as described in the ISOQOL is included in appendix 5.

The COSMIN checklist is a very comprehensive and rigorous checklist that focuses on measurement properties and methodological properties of a PROM. In addition to its checklist, the COSMIN checklist is accompanied by an associative COSMIN taxonomy, which supplements its use. It has been widely used throughout literature but this rigorous checklist comes at the expense of its complexity to the user.[152]
Since the ISOQOL was a standalone study it did not produce uniform definitions such as how the COSMIN initiative produced. In addition the ISOQOL does not produce guidance on how to evaluate the methodology of each property. Therefore the COSMIN 4-point checklist was used to evaluate the measurement properties of each instrument.

4.3 Evaluation

All three haemophilia specific functional scores the HAL, FISH and the HEP-Test-Q produced for the adult population have undergone pilot studies and a subsequent validation studies to assess their measurement properties. The HAL was validated in the largest population (n=127) which was twice as large as the population in which the FISH was validated in (n=63). The HEP-Test-Q was validated in the smallest sample (n=43).

4.3.1 Development of instruments

Concerning the development of instruments, each instrument employed different techniques in its development. The HAL provided sufficient information to the reader to be able to interpret data such as information on the semi-structured interviews for item generation, which was unavailable for other instruments. The HAL and the HEP-Test-Q involved the participation of only one group of participants (patients and healthcare professionals respectively). A patient centred approach was not incorporated to the development of the HEP-Test-Q. This was settled by seeking patients' opinion on missing items after the professionals generated the items.[196] The FISH obtained the input of patients, relatives and healthcare professionals for the item generation stage. The data presentation regarding the development of the instruments FISH and HEP-Test-Q were poor, as it did not give sources of information, such as interviewing and the distribution of participants involved in this stage. The importance of the involvement of patients in the development of PROMs was highlighted by Wiering et al.[211] They suggest that since there has been no changes in the involvement of patients over time[211] this trend should change by giving patients the preference.

4.3.2 Measurement properties of instruments

As a whole, evaluation of the instruments using the COSMIN 4-point checklist depicted that none of the instruments contained an 'excellent' in its measurement properties. As haemophilia is a rare disease authors run into challenges in obtaining large sample sizes which is a specific requirement in obtaining a high rating in the COSMIN checklist. Where the sample size has influenced a low rating, these values have been underlined in table 9.

No author of any of the instruments has clearly described the measurement model indicating if the instrument is a formative or a reflective model. This has implications on the evaluations on certain measurement properties such as internal consistency because in formative models this is not assessed.[151] The COSMIN manual advises investigators to apply a 'thought test' to decide on the measurement model.[193] This test was applied to all three instruments and the HAL and the FISH were found to be of a reflective model as they assessed ADL. Whereas the HEP-Test-Q consisted of a mixed model of both reflective and formative models as they measured ADL and factors that influence function. There is no description in the COSMIN manual on how to evaluate a mixed model. Therefore as the authors of the HEP-Test-Q score assessed internal consistency and used this as a basis for item reduction this measurement model in the HEP-Test-Q was treated as a reflective model.

With regards to internal consistency, all instruments obtained a rating of 'poor.' Both instruments HAL and HEP-Test-Q were influenced by small sample size. If the sample size items were eliminated the HAL would obtain a 'good' rating while the HEP-Test-Q would receive a rating of 'fair'. Item number 5 in the COSMIN checklist for internal consistency was rated 'poor'. This item was excluded as it focuses on factor analysis and to perform factor analysis a sample size of >100 is often required.[212] The

Cronbach's alpha value for internal consistency for the HAL and the HEP–Test-Q scores took values of 0.96 and 0.97 respectively. It has been suggested that if a Cronbach's alpha value >0.9 has the possibility of redundancy as items may be remarkably similar.[158] This suggests that the lengthy instruments could assess parameters with less items than what is included.

Reliability was not assessed in the HAL while the other two instruments received 'fair' ratings for its assessments for reliability. As FISH is a performance measure it is unclear in the reporting of data by the authors if it were for intra-rater reliability or inter-rater reliability as both parameters were assessed. Since the statistical analysis only produced a pooled ICC, it was assumed that a collective mix of both intra-rater and inter-rater reliability were reported. Data presentation of the FISH is inadequate in that it does not present much data to be able to draw proper conclusions regarding this matter. This affected the methodological integrity evaluation of the study (item 10). The HEP-Test-Q had a low sample size affecting its rating. Since the Pearson correlation coefficient was calculated (item 11), the value for overall rating of the measurement property reliability remained the same at 'fair.'

As measurement error is a measurement property that is an extension of the domain of reliability, it mirrored the results of the reliability study. It should be stated that no instrument calculated limits of agreement or standard error of measurement but this could be calculated in the HEP-Test-Q from the data provided (item 11).

Content validity was formally assessed in the HAL and the HEP-Test-Q scores, which had 'good' ratings in the COSMIN 4-point checklist. In spite of the assessments of sample size, all three instruments had the down side of not clearly stating the purpose of the instrument (item 3). The content of the HAL was assessed by administering a scale to both patients and physical therapists. A drawback in this content study was

that the patients reviewed the content of the Dutch version of the HAL whilst the physical therapists assessed the content of the English version. In the evaluation of the COSMIN 4-point checklist, the English version was not carried out, as the content validity was the only measurement property evaluated in that instrument, indicating an inadequate validation process. The content of the HAL was only evaluated at the pilot stage indicating that this content was assessed in the preliminary version of the instrument. The content study of the HEP-Test-Q only had a moderate correlation as it had a mean score of 66.8 out of 100. No content study was carried out for the FISH score, yet the COSMIN checklist was applied for this property as the authors stated that it possessed sufficient face validity. The FISH scored a 'poor' rating for the measurement property, content validity.

All instruments received a rating of 'poor' for the measurement property of structural validity. Overall the reason for this value to be taken was a low sample size (item 4) and/ or no application of factor analysis (item 6). When these items were excluded, all three instruments would score 'good' for structural validity. The HAL was validated against the Dutch-AIMS2 and the IPA. The initial version Dutch-AIMS was developed for the use in rheumatoid arthritis.[213] It has been validated for the use in severe haemophilia.[214] The IPA addresses the personal impact of disease on participation, autonomy and related experience of problems.[200, 215] This questionnaire has not been validated to be used on the haemophilic population. The authors' justification for the use of the IPA was that it consisted of similar Cronbach's alpha value as the developed HAL thus providing it to be a suitable reliability index in this disease.[194] The structural validity of the FISH was assessed using the WFH and Pettersson scores, even though both these scores are validated in the target population it is questionable as to whether it assesses the same construct. The structural validity

of the HEP-Test-Q was assessed using two validated instruments in the haemophiliac population (Haem-A-QoL and HAL) and one instrument that was not validated in a haemophiliac population, the SF-36. It should be noted that the Haem-A-QoL and SF-36 measures the construct HR-QoL as opposed to function. As the SF-36 provides a physical score (the SF-36$_{PCS}$) on function this domain was only used to validate the HEP-Test-Q.

Hypothesis testing was best assessed in the HAL which received a 'good' while the other two scores received 'fair' ratings. This indicates that the FISH and the HEP-Test-Q had not focussed on providing hypotheses.

Cross-cultural validity was not applicable to the PROMs HAL and the HEP-Test-Q instruments as their original versions (in Dutch and German) were evaluated. Even though the respective English instruments were available, cross–cultural validation was not formally assessed as the complete validation of the English instruments had not taken place. The performance based instrument FISH was developed by the authors to avoid this translator issue among different populations. Thus questioning the need for cross-cultural validation in this instrument. The methodological integrity of the translated versions of the HAL and the HEP-Test-Q should be evaluated. The authors of the HAL used the ICF classification system to classify the items in the Dutch HAL and subsequently matched the code to the English version of the ICF classification system. This was used as the basis of the translation of the instrument. No form of validation of the English instrument was performed apart from the aforementioned content study. This is seen as a major drawback of the English version of HAL. The cross cultural validation of the HEP-Test-Q obtained a 'poor' rating which would probably be due to the fact that data presentation was through a poster presentation as opposed to a published article. The English HEP-Test-Q was studied only on a small

sample of adult persons with haemophilia (n=9), as majority were children (n=29). The translated instrument was assessed only by testing for internal consistency as a measurement property. Therefore this English HEP-Test-Q has not undergone full validation.

Criterion validity is not applicable to the PROMs HAL and HEP-Test-Q as neither of these instruments has been produced from a previous longer version of the same instrument. However criterion validity is applicable to the FISH as it is a performance based instrument. This measurement property did receive a rating of 'good' but it is unclear if the criterion used was in fact an adequate 'gold standard' measurement of the construct being measured. The authors used PROMs to assess function, which is known to be a subjective assessment of function. The HAL was the only validated instrument in haemophilia whereas the WOMAC and HAQ scores have not been validated in persons with haemophilia. However, the HAL was only validated in a Dutch speaking adult population. The mean age of the validated population suggests that the FISH instrument was validated in a predominantly paediatric population. Information regarding the language version of the PROM, HAL was not provided by the authors of the FISH, which is vital information to make proper conclusions regarding this validation process.

Responsiveness to change was only assessed as a part of the original validation study in the FISH whereas the responsiveness of the HEP-Test-Q was evaluated in a subsequent study.[216] The instruments that detected change in the joints in the FISH responsiveness study, HAQ and WOMAC have not been validated for the use in haemophilia. In the HEP-Test-Q responsiveness study change was detected only by the WFH orthopaedic joint score (OJS).[216] The OJS is a physical score and does not assess the same construct, function. Other instruments (Haem-A-QoL and SF-36),

which include the HEP-Test-Q, did not show significant results post-intervention thus questioning the responsiveness of the HEP-Test-Q. As this study was carried out in 2010 prior to the presentation of results of the English HEP-Test-Q (2014) it was assumed that responsiveness was assessed in the German instrument. When applying the COSMIN checklist the HEP-Test-Q received a 'poor' rating which was due to the small sample size (item 3). If this factor were to be excluded the HEP-Test-Q would receive a rating of 'fair' for not forming proper hypotheses (item 8) and other minor flaws in the methodology of the study (items 13,16). The assessment of the methodology of the study is irrelevant as it shows that the HEP-Test-Q is not responsive. The HAL has not been assessed for responsiveness to change. There was a case series of three patients, which evaluated ankle distraction. Nevertheless this study was excluded in the analysis of responsiveness as the indication for intervention was not genuine haemophilic arthropathy but secondary causes such as trauma in haemophililacs.[217] The two instruments that assessed responsiveness, the FISH and HEP-Test-Q do not assess functional status if it were to get worse (i.e. following haemarthrosis), or more pronounced intervention required (i.e. joint replacements). Responsiveness is an important aspect especially if an instrument needs to be used in longitudinal studies and assess outcomes.[153, 218] The evidence of responsiveness remains crucial to make adequate assessments using a PROM.

4.3.3 Other properties

Interpretability of scores in all three instruments were poorly presented. This is an aspect that should be clearly improved in each instrument even though it is not a measurement property. Attention should be given by authors to the missing items and to how these were handled. In addition the distribution of scores, the percentage of respondents with the lowest and highest scores, and distribution of scores for each

subgroup should be presented. Another critical value that should be presented is the minimal important change or minimal important clinical difference of the instrument. This allows accurate interpretation to be made regarding what is important to the patient.[219] The advantage of using a disease specific instrument is that clinically important changes can be assessed[209], however if the instruments do not provide information to interpret these values the instrument fails to stay true to its purpose.

The generalisability results suggested that there was an important difference between the populations in which the instruments were validated. The mean age for the HAL 42 years (standard deviation ±13) and the HEP-Test-Q 44 years (standard deviation ± 12) were similar. Whereas the mean age for the FISH score was 14 years (standard deviation ±8) indicating that this instrument was validated in a mixed population, skewed more towards the paediatric population. In addition the FISH was developed and validated in India, a developing nation that has a more severe form of joint disease than those in the western world due to the limited treatment modalities and resources available. This suggests that the FISH score is sensitive to a more progressive arthropathy. This limitation was acknowledged by the developers of the FISH.[195]

The HAL had the best precision out of all instruments while the HEP-Test-Q was next in line. The applicability of a more precise response category is questionable to a performance based instrument such as the FISH. It should be noted that the precision of the items on the instrument may have been confounded by the degree of joint disease seen in the validated population of the FISH.

The construct assessed in all these instruments were function. The conceptual model varied between each developer indicating that each instrument was subjected to changes by the developer. The HAL and the FISH are similar in that they categorised items into the ICF classification. The HAL only included those under the 'activities'

domain while FISH included any activity only excluding those that could not be assessed in a clinic setting. HEP-Test-Q score is a sport oriented score that focuses mainly on the lower limbs. The measurement model for the HAL and the FISH score was a reflective model as all these scores assessed ADL. The HEP-Test-Q was a mixed model of both formative and reflective models as its items possess both indicators and effectors. It is indeed necessary to incorporate both the formative and reflective aspects of function, as the reflective model of function will only measure the homogenous task specific actions. This limitation calls for a mixed model incorporating both formative and reflective models. The HEP-Test-Q assesses a wider construct of function as opposed to the other two instruments.

Evaluating patient and investigator burden revealed that even though the HAL was quite detailed it came at the expense of its length. To ascertain the willingness of a patient to complete a questionnaire, the items should be relevant and the list as short as possible as it poses a threat to the acceptability. As the FISH is a performance based questionnaire it takes longer to be filled in by the investigators than the HEP-Test-Q, which is short, hence takes less time to fill in. All instruments had good response rates noting that the FISH is not a PROM. It should be noted that acceptability as assessed by response rate could be influenced by other causes of disability (physical or mental), current health status and the mode of administration.[154, 220] The Flesch reading ease score for each of the existing scores came above the recommended standard indicating that each was too high. Since the FISH is administered by the clinician, and not a PROM the Flesch reading ease score was disregarded. The high reading ages in the HAL and the HEP-Test-Q scores could be regarded to be the cause of the fact that neither of the scores was properly translated and cross culturally validated in an English speaking population. All instruments were well accepted by the population in

which it was studied on and were practicable instruments. The HAL and FISH have been in practice for at least a decade and are freely accessible. However to obtain the HEP-Test-Q it is mandatory to contact the developer. This fine analysis suggests that HAL has more patient burden than investigator burden. In contrast the FISH has more investigator burden than patient burden. Both of these facts are self-explanatory due to the nature of each instrument. The HEP-Test-Q has a relative burden on both parties, which may be seen as a shortcoming of this instrument.

4.3.3.1 Health literacy and language

The ability to understand and the capacity to process and obtain basic health information, and access to services needed to make appropriate health decisions is known as the health literacy of a patient.[221] This encompasses skills such as reading, comprehension, listening, analysing and decision making. Functional health literacy is crucial as it enables effective communication between the healthcare providers and patients.[222] To address this issue healthcare professionals have attempted to bridge this gap through communicating. However, when implementing a self-administered questionnaire, this may deem arduous . This is by far the biggest challenge encountered by even the best developed PROM that consist of excellent measurement properties.[223] Most of the existing PROMs require a considerable amount of cognitive work by patients when selecting responses to each item, the main reason behind this being the phrasing of each item.[224] This is seen as a principal cause for non-response.[154]

Although individuals with high levels of health literacy have the necessary skills to make decisions using health information, inaccuracies arise when those with low levels of health literacy struggle in this process and hence are more likely to contribute to poor outcomes.[221] The literature has highlighted that cohorts that are more likely to have low levels of health literacy are those in the older age range[225-229], minorities[221,

[230, 231], low socioeconomic status[221, 228] and low levels of education[221, 228, 232]. Thus it is vital that PROMs accurately address these health literacy levels by ensuring that the language of such instruments are not complex and has the ease of understanding. Ambiguous items in a questionnaire may result in responses that are inaccurate representations of the participants' views or may even lead to missing responses. Health literacy is not the only culpable factor in this case but also other circumstances such as using phrases with alternative meanings, which are unfamiliar to the subject. Thus at item generation and questionnaire design, one should use plain, lay language to address the varying levels of education and experience of respondents. This is the only way to ensure that a vast majority of respondents would be able to understand the items. Throughout the literature it has been cited that items should be written in a way that could be comprehended by a 12 year old child.[158] To assess the reading ease of a questionnaire in English Flesch-Kincaid readability tests have been produced.[233]

Other tactics to be included whilst developing a questionnaire are avoiding any negative wording, including items that are specific, and ensuring that each item contains only one question.[149]

The information that is obtained using a PROM is only valid if the participants can understand what is being asked of them which enables the healthcare professional to obtain data about the patient that reflect subject experiences, health status or perspective. Excess cognitive efforts of the patient, results in confusion. This leads to an inaccurate answer be in given to the intended question or a longer response time as the patient attempts to comprehend what the investigator intends on measuring. Therefore, it is of paramount importance to ensure that the items on the questionnaires are clear and easy to understand.

4.4 Assessment of joint disease

It is evident that even though the mentioned instruments assess joint function they are very different in what they assess. The HAL and the FISH assess ADL whereas the HEP-Test-Q assesses ADL and factors that influence function. The HAL and HEP-Test-Q are PROMs whereas the FISH is a performance based assessment.

A study[234] showed that the inter-rater reliability between different healthcare professionals that assessed disability using performance based assessments werepoor.[234] This indicates that performance tasks lack uniformity in practice as different professions may have varying levels of expectations depending on to their healthcare role. In order to minimise these careful instructions, and/or induction courses on how to use such tools are carried out but this is time consuming and expensive. Other disadvantages in these performance measures would be the practicality of such tests in a clinical environment that may be rushed for time, and space required to perform the tasks. Pinheiro et al[143] evaluated PROMs and performance based measures in patients with arthritis. They revealed that PROMs have the capacity to unravel domains that a performance task is unable to do. In addition PROMs accurately represent overall function as these provide a platform for the patient to discuss their personal functional experiences.[143]

PROMs are an essential way of assessing function, as it has proven useful in routine practice and clinical research to inform the effectiveness of interventions. A well-developed PROM, provides a good foundation for healthcare professionals in making an informed decision about patient care and to evaluate the safety and effectiveness of interventions.[136] In the absence of PROMs it would be difficult to weigh up the risks and benefits of management options on outcomes other than mortality.[132] PROMs do carry the risk of the 'normative social effect' where often the patients fill in items based

on what the healthcare professional would want to hear. This overclouds the actual functional status of the patient. PROMs are used to make comparisons between the outcome of care received and efficacy to allow clinical decision making as investigations would do.[137] Studying the HAL and the HEP-Test-Q showed that even though both these PROMs assess the same construct they are very different instruments in terms of what they assess. The broader construct of function, which the HEP-Test-Q assesses, proves to be vital even though the ADL that it assesses remain vague. Close attention should be paid to the personal and environmental factors that influence one's function which is a part of disability model described by the ICF. A broader construct often requires a mix of the formative and reflective measurement models. Inclusion of a broader construct comes with its own problems, which revolves around the psychometric analysis of the measurement properties. With haemophilia being a rare disease obtaining the necessary sample size for complex statistical analysis such as factor analysis for such mixed model proves to be the biggest hurdle.

Other methods of evaluation of joint disease such as physical examination tools, imaging scores and QoL measures prove important as they provide unique aspects of joint disease to be assessed by the healthcare professional as these measure different constructs. PROMs to date remain a preferable method to obtain a good representation of joint health and a well refined method of prognostic information. Nevertheless, it should be emphasised that a PROM will not be able to solely measure joint disease and should be used to complement other instruments that assess different constructs of joint disease as each instrument has different influencing factors.[235, 236]

4.5 Limitations

The PRISMA guidance was followed throughout this study wherever possible. The only instance in which this could not be followed was data collection stage in the methods. The input of an additional investigator would have been beneficial in the study selection. However, this did not compromise the detection of available instruments that assess function in adult haemophilia.

A limitation in this study is that analysis of the instruments was done using the COSMIN checklist, which is a PROM guidance. It should be reiterated that the FISH in fact is not a PROM. This was still used for uniformity as there is no rigorous appraisal tool to evaluate the methodological quality of performance based instruments. It should be noted that the COSMIN group is currently in the process of producing such a guidance tool to be utilised by researchers.

The COSMIN checklist has the drawback of only evaluating measurement properties. To minimise this effect other properties mentioned in the literature were identified and evaluated. However this is affected by measurement bias as there is no standardised method to evaluate such properties. It is thought that it would be better if the author assessed these 'other properties' as opposed to not assessing thus this was carried out.

The readability formulae only take into account the word length and neglects vocabulary. As a result, one should not rely on this method exclusively. However as there is no other way to assess readability and as these formulae are the tools most widely used in medical literature, they were used to assess readability.[237]

5 CHAPTER 5- Further Study and Conclusions

5.1 Further study

This systematic review revealed that there is a need for a new PROM to assess function in the English speaking adult persons with haemophilia. One may question why not fully cross-culturally validate or improve the existing PROMs to an English speaking population. This is definitely a feasible option as highlighted above, the construct and the items of the existing instruments assess a narrow niche. Therefore it deems feasible that a new PROM is developed for this population to assess function in a holistic manner. A three part study has been designed addressing the drawbacks of the existing instruments, to develop and validate a novel PROM. This study has been presented for an ethical review under the research ethics committee reference 16/NW/0532 and has been given ethical approval. It is currently undergoing the Health Research Authority (HRA) process.

5.2 Conclusion

The existing functional scores consist of satisfactory measurement and other properties but not all properties have been assessed by the developers. Applying the COSMIN 4-point checklist provided a comprehensive approach to investigate the measurement properties indicating that the methodological quality varied significantly between each property and instrument. Even though the COSMIN checklist was applied to the FISH the results should be interpreted with caution as it is a performance based instrument.

The HAL had the least amount of measurement properties assessed, but had better ratings on the checklist than the FISH and the HEP-Test-Q. The English versions of the PROMs HAL and HEP-Test-Q had not undergone full cross-cultural validation and had high readability scores. The HAL and the FISH assessed function only as ADL whereas the HEP-Test-Q assessed function broadly. None of these models have taken into account the full disability model of ICF.

The narrow niche measured should be broadened to assess function in a holistic manner to incorporate the full disability model and the views of patients. PROMs play an important role in obtaining the perspective of a patient with a chronic disease such as haemophilia. There is a need for a validated PROM to assess function for the English speaking persons with haemophilia. This systematic review and critical appraisal was carried out as a part of preliminary work for the development of a novel instrument that aims to look at a disease specific PROM for the adult population. The shortcomings of the existing scores will be addressed in this novel PROM that will assess function in this population.

6 REFERENCES

1. Rodriguez-Merchan EC. Management of the orthopaedic complications of haemophilia. The Journal of bone and joint surgery British volume. 1998;80(2):191-6.

2. Mannucci PM, Tuddenham EGD. The Hemophilias — From Royal Genes to Gene Therapy. New England Journal of Medicine. 2001;344(23):1773-9.

3. Konkle BA JN, Nakaya Fletcher S. . Hemophilia B. In: Pagon RA AM, Ardinger HH, et al., editors. , editor. GeneReviews® [Internet]: Seattle (WA): University of Washington, Seattle; 1993-2016. ; 2000.

4. Plug I, Mauser-Bunschoten EP, Bröcker-Vriends AHJT, van Amstel HKP, van der Bom JG, van Diemen-Homan JEM, et al. Bleeding in carriers of hemophilia. Blood. 2006;108(1):52.

5. Raabe M. Genes & Disease: Hemophilia. NSTA Recommends. 2009:3-.

6. Gomez K, Bolton-Maggs P. Factor XI deficiency. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14(6):1183-9.

7. Giangrande P. ACQUIRED HEMOPHILIA Revised edition. World Federation of Haemophilia: 2012.

8. Franchini M. Acquired hemophilia A. Hematology (Amsterdam, Netherlands). 2006;11(2):119-25.

9. Rodriguez-Merchan EC. Prevention of the Musculoskeletal Complications of Hemophilia. Advances in Preventive Medicine. 2012:1-8.

10. Ingram GI. The history of haemophilia. Journal of clinical pathology. 1976;29(6):469-79.

11. Rogaev EI, Grigorenko AP, Faskhutdinova G, Kittler EL, Moliaka YK. Genotype analysis identifies the cause of the "royal disease". Science (New York, NY). 2009;326(5954):817.

12. World Federation of Hemophilia. ANNUAL GLOBAL SURVEY 2014. 2015.

13. Centers for Disease Control and Prevention. Hemophilia (Minimal) Data Set,Federally Funded Hemophilia Treatment Centers. National Summary. Atlanta: Centers for Disease Control and Prevention 2002.

14. Soucie JM, Evatt B, Jackson D. Occurrence of hemophilia in the United States. American Journal of Hematology. 1998;59(4):288-94.

15. White GC, 2nd, Rosendaal F, Aledort LM, Lusher JM, Rothschild C, Ingerslev J. Definitions in hemophilia. Recommendation of the scientific subcommittee on factor VIII and factor IX of the scientific and standardization committee of the International Society on Thrombosis and Haemostasis. Thrombosis and haemostasis. 2001;85(3):560.

16. Berntorp E, Shapiro AD. Modern haemophilia care. The Lancet.379(9824):1447-56.

17. White G, Rosendaal F, Aledort L, Lusher J, Rothschild C, Ingerslev J. Recommendation of the scientific subcommittee on factor VIII and factor IX of the

scientific and standardization committee of the International Society on Thrombosis and Haemostasis. Thrombosis and haemostasis. 2001;85:560.

18. Price VE, Hawes SA, Chan AKC. A practical approach to hemophilia care in children. Paediatrics & Child Health. 2007;12(5):381-3.

19. The Haemophilia Alliance. A NATIONAL SERVICE SPECIFICATION FOR HAEMOPHILIA AND OTHER INHERITED BLEEDING DISORDERS. 2006.

20. Hemophilia WFo. Hemophilia 2012 [updated May 2016; cited 2016 21/01]. Treatment]. Available from: http://www.wfh.org/en/page.aspx?pid=642.

21. Shastry SP, Kaul R, Baroudi K, Umar D. Hemophilia A: Dental considerations and management. Journal of International Society of Preventive & Community Dentistry. 2014;4(Suppl 3):S147-S52.

22. Keely ML, Taylor N, Chard RL, Jr. Spinal cord compression as a complication of haemophilia. Archives of disease in childhood. 1972;47(255):826-8.

23. Trueta J. The Orthopædic Management of Hæmophilia. Proceedings of the Royal Society of Medicine. 1962;55(12):1058-9.

24. Rodriguez-Merchan EC. Musculoskeletal complications of hemophilia. HSS Journal: The Musculoskeletal Journal Of Hospital For Special Surgery. 2010;6(1):37-42.

25. Pettersson H, Gilbert MS. The Hemophilic Pseudotumor. Diagnostic Imaging in Hemophilia: Musculoskeletal and Other Hemorrhagic Complications. London: Springer London; 1985. p. 89-102.

26. Gilbert MS. The hemophilic pseudotumor. Progress in clinical and biological research. 1990;324:257-62.

27. Caviglia H LM, Galatro G, Candela M and Neme D. . Pseudotumors of the Limbs in Patients with Hemophilia. J Blood Disord 2015;2(2).

28. Fehily M, Fleming P, O'Shea E, Smith O, Smyth H. Total knee arthroplasty in patients with severe haemophilia. International orthopaedics. 2002;26(2):89-91.

29. Boehlen F, Graf L, Berntorp E. Outcome measures in haemophilia: a systematic review. European journal of haematology Supplementum. 2014;76:2-15.

30. Bolton-Maggs PH, Pasi KJ. Haemophilias A and B. Lancet (London, England). 2003;361(9371):1801-9.

31. Den Uijl IEM, Fischer K, Van Der Bom JG, Grobbee DE, Rosendaal FR, Plug I. Clinical outcome of moderate haemophilia compared with severe and mild haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2009;15(1):83-90.

32. Bussing R, Johnson SB. Psychosocial issues in hemophilia before and after the HIV crisis: a review of current research. Gen Hosp Psychiatry. 1992;14(6):387-403.

33. Sterndale H. Haemarthrosis and haemophilia. Proceedings of the Royal Society of Medicine. 1967;60(1):37-8.

34. Mackman N. Tissue-specific hemostasis in mice. Arterioscler Thromb Vasc Biol. 2005;25(11):2273-81.

35. Monahan PE, Doria AS, Ljung R, Jimenez-Yuste V. Optimizing joint function: new knowledge and novel tools and treatments. Haemophilia : the official journal of the World Federation of Hemophilia. 2012;18 Suppl 5:17-26.

36. Ramgren O. Haemophilia in Sweden. III. Symptomatology, with special reference to differences between haemophilia A and B. Acta medica Scandinavica. 1962;171:237-42.

37. Centers for Disease Control and Prevention. Report on the Universal Data Collection Program. The Universal Data Collection Program; 2005. p. 1-39.

38. Knobe K, Berntorp E. Haemophilia and joint disease: pathophysiology, evaluation and management. 2011. 2011;1(1):9.

39. US Department of Health and Human Services. Report on the Universal Data Collection Program. Atlanta, Georgia: Centers for Disease Control and Prevention, 2000.

40. Ota S, McLimont M, Carcao MD, Blanchette VS, Graham N, Paradis E, et al. Definitions for haemophilia prophylaxis and its outcomes: the Canadian consensus study. Haemophilia : the official journal of the World Federation of Hemophilia. 2007;13(1):12-20.

41. Donadel-Claeyssens S. Current co-ordinated activities of the PEDNET (European Paediatric Network for Haemophilia Management). Haemophilia : the official journal of the World Federation of Hemophilia. 2006;12(2):124-7.

42. Blanchette VS, Key NS, Ljung LR, Manco-Johnson MJ, van den Berg HM, Srivastava A. Definitions in hemophilia: communication from the SSC of the ISTH. J Thromb Haemost. 2014;12(11):1935-9.

43. World Federation of Hemophilia (WFH). Chronic Haemophilic Synovitis: The role of Radiosynovectomy 2004 [cited 2015 21 September]. Available from: http://www1.wfh.org/publication/files/pdf-1176.pdf.

44. Lack CH. Chondrolysis in arthritis. The Journal of bone and joint surgery British volume. 1959;41-B(2):384-7.

45. Tan AH, Mitra AK, Chang PC, Tay BK, Nag HL, Sim CS. Assessment of blood-induced cartilage damage in rabbit knees using scanning electron microscopy. Journal of orthopaedic surgery (Hong Kong). 2004;12(2):199-204.

46. Aigner T, Soeder S, Haag J. IL-1beta and BMPs--interactive players of cartilage matrix degradation and regeneration. Eur Cell Mater. 2006;12:49-56; discussion

47. Roosendaal G, Vianen ME, Wenting MJ, van Rinsum AC, van den Berg HM, Lafeber FP, et al. Iron deposits and catabolic properties of synovial tissue from patients with haemophilia. The Journal of bone and joint surgery British volume. 1998;80(3):540-5.

48. Valentino LA, Hakobyan N, Rodriguez N, Hoots WK. Pathogenesis of haemophilic synovitis: experimental studies on blood-induced joint damage. Haemophilia : the official journal of the World Federation of Hemophilia. 2007;13 Suppl 3:10-3.

49. Lotz M. Cytokines in Cartilage Injury and Repair. Clinical Orthopaedics and Related Research. 2001;391:S108-S15.

50. Hakobyan N, Kazarian T, Jabbar AA, Jabbar KJ, Valentino LA. Pathobiology of hemophilic synovitis I: overexpression of mdm2 oncogene. Blood. 2004;104(7):2060-4.

51. Lafeber FP, Miossec P, Valentino LA. Physiopathology of haemophilic arthropathy. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14 Suppl 4:3-9.

52. Roosendaal G, Lafeber FPJG. Blood-Induced Joint Damage in Hemophilia. Semin Thromb Hemost. 2003;29(01):037-42.

53. Jansen NW, Roosendaal G, Bijlsma JW, Degroot J, Lafeber FP. Exposure of human cartilage tissue to low concentrations of blood for a short period of time leads to prolonged cartilage damage: an in vitro study. Arthritis Rheum. 2007;56(1):199-207.

54. Jansen NWD, Roosendaal G, Bijlsma JWJ, DeGroot J, Theobald M, Lafeber FPJG. Degenerated and healthy cartilage are equally vulnerable to blood-induced damage. Annals Of The Rheumatic Diseases. 2008;67(10):1468-73.

55. Hooiveld MJJ, Roosendaal G, van den Berg HM, Bijlsma JWJ, Lafeber FPJG. Haemoglobin‐derived iron‐dependent hydroxyl radical formation in blood‐induced joint damage: an in vitro study. Rheumatology. 2003;42(6):784-90.

56. Valentino LA, Hakobyan N, Kazarian T, Jabbar KJ, Jabbar AA. Experimental haemophilic synovitis: rationale and development of a murine model of human factor

VIII deficiency. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10(3):280-7.

57. Roosendaal G, Vianen ME, Marx JJM, Van Den Berg HM, Lafeber FPJG, Bijlsma JWJ. Blood-induced joint damage: A human in vitro study. Arthritis & Rheumatism. 1999;42(5):1025-32.

58. Roosendaal G, van Rinsum AC, Vianen ME, van den Berg HM, Lafeber FP, Bijlsma JW. Haemophilic arthropathy resembles degenerative rather than inflammatory joint disease. Histopathology. 1999;34(2):144-53.

59. Ahlberg Å. Haemophilia in Sweden VII. Incidence, treatment and prophylaxis of arthropathy and other musculo-skeletal manifestations of haemophilia A and B. Acta Orthopaedica. 1965;36(S77):3-132.

60. Soucie JM, Cianfrini C, Janco RL, Kulkarni R, Hambleton J, Evatt B, et al. Joint range-of-motion limitations among young males with hemophilia: prevalence and risk factors. Blood. 2004;103(7):2467-73.

61. CDC editor. Summary Report of UDC Activity National- Treatment and Outcome by Severity (Hemophilia). Atlanta: Department of Health and Human Services Centers for Disease Control and Prevention, 2008.

62. Rodriguez-Merchan EC. Pathogenesis, early diagnosis, and prophylaxis for chronic hemophilic synovitis. Clin Orthop Relat Res. 1997(343):6-11.

63. Simpson ML, Valentino LA. Management of joint bleeding in hemophilia. Expert review of hematology. 2012;5(4):459-68.

64. Blanchette P, Rivard G, Israels S, Robinson S, Ali K, Walker I, et al. A survey of factor prophylaxis in the Canadian haemophilia A population. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10(6):679-83.

65. Miners, Sabin, Tolley, Jenkinson, Kind, Lee. Assessing health-related quality-of-life in individuals with haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 1999;5(6):378-85.

66. DePalma AF. Hemophilic arthropathy. Clin Orthop Relat Res. 1967;52:145-65.

67. Arnold WD, Hilgartner MW. Hemophilic arthropathy. Current concepts of pathogenesis and management. The Journal of bone and joint surgery American volume. 1977;59(3):287-305.

68. Nilsson IM, Berntorp E, Lofqvist T, Pettersson H. Twenty-five years' experience of prophylactic treatment in severe haemophilia A and B. J Intern Med. 1992;232(1):25-32.

69. Cassis FRMY, Querol F, Forsyth A, Iorio A, On Behalf Of The Hero International Advisory B. Psychosocial aspects of haemophilia: a systematic review of methodologies and findings. Haemophilia : the official journal of the World Federation of Hemophilia. 2012;18(3):e101-e14.

70. Waters L. Haemophilia. Nursing standard (Royal College of Nursing (Great Britain) : 1987). 1995;9(47):33-7; quiz 8-9.

71. Sulser E, editor A patient's perspective on hemophilia. Seminars in hematology; 2006: Elsevier.

72. Barlow JH, Stapley J, Ellard DR. Living with haemophilia and von Willebrand's: a descriptive qualitative study. Patient Educ Couns. 2007;68(3):235-42.

73. Browne WJ, Mally MA, Kane RP. PSYCHOSOCIAL ASPECTS OF HEMOPHILIA: A STUDY OF TWENTY-EIGHT HEMOPHILIC CHILDREN AND THEIR FAMILIES*. American Journal of Orthopsychiatry. 1960;30(4):730-40.

74. Ross J. Perspectives of haemophilia carriers. Haemophilia : the official journal of the World Federation of Hemophilia. 2000;6 Suppl 1:41-5.

75. Cassis F. PSYCHOSOCIAL CARE FOR PEOPLE WITH HEMOPHILIA. TREATMENT OF HEMOPHILIA. 2007;44.

76. Bottos AM, Zanon E, Sartori MT, Girolami A. Psychological aspects and coping styles of parents with Haemophilic child undergoing a programme of counselling and psychological support. Haemophilia : the official journal of the World Federation of Hemophilia. 2007;13(3):305-10.

77. Goldstein G, Kenet G. The impact of chronic disease on the family. Haemophilia : the official journal of the World Federation of Hemophilia. 2002;8(3):461-5.

78. Ross. Perspectives of haemophilia carriers. Haemophilia : the official journal of the World Federation of Hemophilia. 2000;6:41-5.

79. duTreil S. Physical and psychosocial challenges in adult hemophilia patients with inhibitors. Journal of Blood Medicine. 2014;5:115-22.

80. Berntorp E, Astermark J, Bjorkman S, Blanchette VS, Fischer K, Giangrande PL, et al. Consensus perspectives on prophylactic therapy for haemophilia: summary statement. Haemophilia : the official journal of the World Federation of Hemophilia. 2003;9 Suppl 1:1-4.

81. Aledort LM. History of haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2007;13 Suppl 5:1-2.

82. Mannucci PM. Back to the future: a recent history of haemophilia treatment. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14 Suppl 3:10-8.

83. Ludlam CA. New-variant Creutzfeldt-Jakob disease and treatment of haemophilia. Executive Committee of the UKHCDO. United Kingdom Haemophilia Centre Directors' Organisation. Lancet (London, England). 1997;350(9092):1704.

84. Darby SC, Kan SW, Spooner RJ, Giangrande PL, Lee CA, Makris M, et al. The impact of HIV on mortality rates in the complete UK haemophilia population. Aids. 2004;18(3):525-33.

85. Darby SC, Ewart DW, Giangrande PL, Spooner RJ, Rizza CR, Dusheiko GM, et al. Mortality from liver cancer and liver disease in haemophilic men and boys in UK given blood products contaminated with hepatitis C. UK Haemophilia Centre Directors' Organisation. Lancet (London, England). 1997;350(9089):1425-31.

86. Rumi MG, De Filippi F, Santagostino E, Colombo M. Hepatitis C in haemophilia: lights and shadows. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10 Suppl 4:211-5.

87. Wilde JT, Lee CA, Darby SC, Kan SW, Giangrande P, Phillips AN, et al. The incidence of lymphoma in the UK haemophilia population between 1978 and 1999. Aids. 2002;16(13):1803-7.

88. Sabin CA, Yee TT, Devereux H, Griffioen A, Loveday C, Phillips AN, et al. Two decades of HIV infection in a cohort of haemophilic individuals: clinical outcomes and response to highly active antiretroviral therapy. Aids. 2000;14(8):1001-7.

89. Guidelines on the selection and use of therapeutic products to treat haemophilia and other hereditary bleeding disorders. Haemophilia : the official journal of the World Federation of Hemophilia. 2003;9(1):1-23.

90. Keeling D, Tait C, Makris M. Guideline on the selection and use of therapeutic products to treat haemophilia and other hereditary bleeding disorders. A United Kingdom Haemophilia Center Doctors' Organisation (UKHCDO) guideline approved by the British Committee for Standards in Haematology. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14(4):671-84.

91. Richards M, Williams M, Chalmers E, Liesner R, Collins P, Vidler V, et al. A United Kingdom Haemophilia Centre Doctors' Organization guideline approved by the British Committee for Standards in Haematology: guideline on the use of prophylactic factor VIII concentrate in children and adults with severe haemophilia A. British Journal of Haematology. 2010;149(4):498-507.

92. Kempton CL, White GC. How we treat a hemophilia A patient with a factor VIII inhibitor. Blood. 2009;113(1):11.

93. Hay CR, Baglin TP, Collins PW, Hill FG, Keeling DM. The diagnosis and management of factor VIII and IX inhibitors: a guideline from the UK Haemophilia Centre Doctors' Organization (UKHCDO). Br J Haematol. 2000;111(1):78-90.

94. Weigel N, Carlson BR. Physical activity and the hemophiliac: yes or no? Am Correct Ther J. 1975;29(6):197-205.

95. Herbsleb M, Puta C, Hilberg T. Hemophilia and Exercise Project (HEP) Conception and Contents of a “Programmed Sports Therapy” for Hemophilic Patients. In: Scharrer I, Schramm W, editors. 37th Hemophilia Symposium: Springer Berlin Heidelberg; 2008. p. 45-59.

96. Beeton K, Cornwell J, Alltree J. Muscle rehabilitation in haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 1998;4(4):532-7.

97. Buzzard BM. Proprioceptive training in haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 1998;4(4):528-31.

98. Mulder K, Cassis F, Seuser DR, Narayan P, Dalzell R, Poulsen W. Risks and benefits of sports and fitness activities for people with haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10 Suppl 4:161-3.

99. Hanley J, McKernan A, Creagh MD, Classey S, McLaughlin P, Goddard N, et al. Guidelines for the management of acute joint bleeds and chronic synovitis in haemophilia: A United Kingdom Haemophilia Centre Doctors' Organisation (UKHCDO) guideline. Haemophilia : the official journal of the World Federation of Hemophilia. 2017.

100. Beeton K, Buzzard B. Physiotherapy in the Management of Patients with Inhibitors. Inhibitors in Patients with Haemophilia: Blackwell Science Ltd; 2008. p. 160-8.

101. Stephensen D, Tait RC, Brodie N, Collins P, Cheal R, Keeling D, et al. Changing patterns of bleeding in patients with severe haemophilia A. Haemophilia : the official journal of the World Federation of Hemophilia. 2009;15(6):1210-4.

102. Cuesta-Barriuso R, #xe9, #xf3, mez-Conesa A, #xf3, pez-Pina J, et al. Physiotherapy Treatment in Patients with Hemophilia and Chronic Ankle Arthropathy: A Systematic Review. Rehabilitation Research and Practice. 2013;2013:10.

103. Wallny T, Brackmann HH, Semper H, Schumpe G, Effenberger W, Hess L, et al. Intra-articular hyaluronic acid in the treatment of haemophilic arthropathy of the knee. Clinical, radiological and sonographical assessment. Haemophilia : the official journal of the World Federation of Hemophilia. 2000;6(5):566-70.

104. Rampal V, Odent T, Torchet MF, Rothschild C, Elie C, Glorion C, et al. Surgical synovectomy of the knee in young haemophiliacs: long-term results of a monocentric series of 23 patients. Journal of Children's Orthopaedics. 2010;4(1):33-7.

105. Manco-Johnson MJ, Nuss R, Lear J, Wiedel J, Geraghty SJ, Hacker MR, et al. 32P Radiosynoviorthesis in children with hemophilia. Journal of pediatric hematology/oncology. 2002;24(7):534-9.

106. Wiedel JD. Arthroscopic synovectomy of the knee in hemophilia: 10-to-15 year followup. Clin Orthop Relat Res. 1996(328):46-53.

107. Rodriguez-Merchan EC. Methods to treat chronic haemophilic synovitis. Haemophilia : the official journal of the World Federation of Hemophilia. 2001;7(1):1-5.

108. Geraghty S. Orthopedic Complications and Treatment Related to Chronic Hemarthrosis: National Haemophilia Foundation; 2012 [cited 2015 28 September]. Available from: https://www.hemophilia.org/sites/default/files/document/files/Nurses-Guide-Chapter-9-Orthopedic-Complications-Treatment.pdf.

109. Karthaus RP, Novakova IR. Total knee replacement in haemophilic arthropathy. The Journal of bone and joint surgery British volume. 1988;70(3):382-5.

110. Thomason HC, 3rd, Wilson FC, Lachiewicz PF, Kelley SS. Knee arthroplasty in hemophilic arthropathy. Clin Orthop Relat Res. 1999(360):169-73.

111. Houghton GR, Dickson RA. Lower limb arthrodeses in haemophilia. The Journal of bone and joint surgery British volume. 1978;60-b(3):387-9.

112. Stucky B, Pereira C. Henrica C. W. de Vet, Caroline B. Terwee, Lidwine B. Mokkink, and Dirk L. Knol: Measurement in medicine: a practical guide. Quality of Life Research. 2012;21(2):371-3.

113. Engel GL. The need for a new medical model: a challenge for biomedicine. Science (New York, NY). 1977;196(4286):129-36.

114. Verbrugge LM, Jette AM. The disablement process. Social science & medicine (1982). 1994;38(1):1-14.

115. Riva S, Bullinger M, Amann E, von Mackensen S. Content comparison of haemophilia specific patient-rated outcome measures with the international classification of functioning, disability and health (ICF, ICF-CY). Health and quality of life outcomes. 2010;8:139.

116. Altman BM. Definitions, concepts, and measures of disability. Annals of Epidemiology. 2014;24(1):2-7.

117. Leonardi M, Bickenbach J, Ustun TB, Kostanjsek N, Chatterji S. The definition of disability: what is in a name? The Lancet.368(9543):1219-21.

118. Whiteneck G, editor Conceptual models of disability: past, present, and future. Workshop on disability in America: A new look; 2006: The National Academies Press Washington DC.

119. Sage Publications i, Mullner RM. Health and Medicine (The Sage Reference Series on Disability: Key Issues and Future Directions) [electronic book]: Sage Publications.

120. Drum CE, Krahn GL, Bersani HA. Disability and public health. [electronic book]: Washington, D.C. : American Public Health Association, 2009.; 2009.

121. Jette AM, Keysor JJ. Disability models: implications for arthritis exercise and physical activity interventions. Arthritis Rheum. 2003;49(1):114-20.

122. Nagi SZ. Some conceptual issues in disability and rehabilitation. In: SUSSMAN MB, editor. SOCIOLOGY AND REHABILITATION Washington, EUA : American Sociological Association.; 1965.

123. World Health Organisation. The International Classification of Impairments, Disabilities, and Handicaps—a manual relating to the consequences of disease. World Health Organization; : Geneva: World Health Organization; , 1980.

124. World Health Organisation. International Classification of Functioning, Disability and Health (ICF). . World Health Organisation: 2001.

125. Bierman AS. Functional Status: The Sixth Vital Sign. Journal of General Internal Medicine. 2001;16(11):785-6.

126. Crossing the quality chasm. [electronic book] : a new health system for the 21st century: Washington, D.C. : National Academy Press, c2001.; 2001.

127. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. Archives of Physical Medicine and Rehabilitation.81:S21-S9.

128. Lawton MP. The functional assessment of elderly people. Journal of the American Geriatrics Society. 1971;19(6):465-81.

129. Dijkers M. Quality of life after spinal cord injury: a meta analysis of the effects of disablement components. Spinal cord. 1997;35(12):829-40.

130. de Kleijn P, van Genderen FR, van Meeteren NL. Assessing functional health status in adults with haemophilia: towards a preliminary core set of clinimetric instruments based on a literature search in Rheumatoid Arthritis and Osteoarthritis. Haemophilia : the official journal of the World Federation of Hemophilia. 2005;11(4):308-18.

131. US Food and Drug Administration. Guidance for Industry. Patient-reported outcome measures: Use in medical product development to support labeling claims.

2009 [cited 2015 27/10]. Available from: http://www.fda.gov/downloads/Drugs/.../Guidances/UCM193282.pdf.

132. European Haematology Association. Guidelines- Patient Reported Outcomes in Haematology 2012 [27/10/2015.]. Available from: http://www.ehaweb.org/assets/documents/Guidelines-PRO-SWG-QoL.pdf.

133. Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ. 2010;340.

134. Kyte DG, Calvert M, van der Wees PJ, ten Hove R, Tolan S, Hill JC. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. Physiotherapy. 2015;101(2):119-25.

135. Department of Health. Equity and Excellence:Liberating the NHS. London2010.

136. Haywood KL. Patient-reported outcome I: Measuring what matters in musculoskeletal care. Musculoskeletal Care. 2006;4(4):187-203.

137. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013;346.

138. NHS Information Centre. National Patient Reported Outcomes Programme [cited 2015 02 November]. Available from: http://www.hscic.gov.uk/proms.

139. Marshall S, Haywood K, Fitzpatrick R. Impact of patient-reported outcome measures on routine practice: a structured review. Journal of evaluation in clinical practice. 2006;12(5):559-68.

140. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Quality of Life Research. 2008;17(2):179-93.

141. Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. Qual Life Res. 2008;17(9):1125-35.

142. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis. 1985;38(1):27-36.

143. Pinheiro LC, Callahan LF, Cleveland RJ, Edwards LJ, Reeve BB. The Performance and Association Between Patient-reported and Performance-based Measures of Physical Functioning in Research on Individuals with Arthritis. J Rheumatol. 2016;43(1):131-7.

144. Curb JD, Ceria-Ulep CD, Rodriguez BL, Grove J, Guralnik J, Willcox BJ, et al. Performance-based measures of physical function for high-function populations. Journal of the American Geriatrics Society. 2006;54(5):737-42.

145. Lord FM, Novick MR, Birnbaum A. Statistical theories of mental test scores: Reading (Mass.): Addison-Wesley, 1968.; 1968.

146. Nunnally JC, Bernstein IH. Psychometric theory: New York ; McGraw-Hill, 1994.
3rd ed.; 1994.

147. Embretson SE, Reise SP. Item response theory for psychologists multivariate. Mahwa: Lawrence Erlbaum Associates, Inc; 2000.

148. R. FA. Clinimetrics. . Psychological Medicine. 1988;18(01):248-.

149. De Vet HCW. Measurement in medicine. [electronic book] : a practical guide: Cambridge : Cambridge University Press, 2011.; 2011.

150. Coltman T, Devinney TM, Midgley DF, Venaik S. Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research. 2008;61:1250-62.

151. Edwards JR. The Fallacy of Formative Measurement. Organizational Research Methods. 2011;14(2):370-88.

152. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.

153. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of clinical epidemiology. 2010;63(7):737-45.

154. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health technology assessment (Winchester, England). 1998;2(14):i-iv, 1-74.

155. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures

used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8):1889-905.

156. Testa MA, Simonson DC. Assesment of quality-of-life outcomes. N Engl J Med. 1996;334(13):835-40.

157. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11(3):193-205.

158. Streiner DL, Norman GR. Health measurement scales. [electronic book] : a practical guide to their development and use: Oxford : Oxford University Press, 2008.
4th ed.; 2008.

159. Flesch R. A new readability yardstick. Journal of Applied Psychology. 1948;32(3):221-33.

160. Gilbert MS. Prophylaxis: musculoskeletal evaluation. Seminars in hematology. 1993;30(3 Suppl 2):3-6.

161. Feldman BM, Funk S, Lundin B, Doria AS, Ljung R, Blanchette V. Musculoskeletal measurement tools from the International Prophylaxis Study Group (IPSG). Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14 Suppl 3:162-9.

162. Soucie JM, Wang C, Forsyth A, Funk S, Denny M, Roach KE, et al. Range of motion measurements: reference values and a database for comparison studies.

Haemophilia : the official journal of the World Federation of Hemophilia. 2011;17(3):500-7.

163. Feldman BM, Funk SM, Bergstrom BM, Zourikian N, Hilliard P, van der Net J, et al. Validation of a new pediatric joint scoring system from the International Hemophilia Prophylaxis Study Group: validity of the hemophilia joint health score. Arthritis care & research. 2011;63(2):223-30.

164. Hilliard P, Funk S, Zourikian N, Bergstrom BM, Bradley CS, McLimont M, et al. Hemophilia joint health score reliability study. Haemophilia : the official journal of the World Federation of Hemophilia. 2006;12(5):518-25.

165. P. H, VS. B, A. D. The Hemophilia Joint Health Score (HJHS) correlates highly with radiographic damage. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14(Suppl.2):80 (Poster 12PO56).

166. Pettersson H, Ahlberg A, Nilsson IM. A radiologic classification of hemophilic arthropathy. Clin Orthop Relat Res. 1980(149):153-9.

167. Pettersson H, Nilsson IM, Hedner U, Norehn K, Ahlberg A. Radiologic evaluation of prophylaxis in severe haemophilia. Acta paediatrica Scandinavica. 1981;70(4):565-70.

168. Silva M, Luck JV, Jr., Quon D, Young CR, Chin DM, Ebramzadeh E, et al. Inter- and intra-observer reliability of radiographic scores commonly used for the evaluation of haemophilic arthropathy. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14(3):504-12.

169. Foppen W, van der Schaaf IC, Fischer K. Value of routine ultrasound in detecting early joint changes in children with haemophilia using the 'Haemophilia Early Arthropathy Detection with UltraSound' protocol. Haemophilia : the official journal of the World Federation of Hemophilia. 2016;22(1):121-5.

170. Zukotynski K, Jarrin J, Babyn PS, Carcao M, Pazmino-Canizares J, Stain AM, et al. Sonography for assessment of haemophilic arthropathy in children: a systematic protocol. Haemophilia : the official journal of the World Federation of Hemophilia. 2007;13(3):293-304.

171. Keshava S, Gibikote S, Mohanta A, Doria AS. Refinement of a sonographic protocol for assessment of haemophilic arthropathy. Haemophilia : the official journal of the World Federation of Hemophilia. 2009;15(5):1168-71.

172. Melchiorre D, Linari S, Innocenti M, Biscoglio I, Toigo M, Cerinic MM, et al. Ultrasound detects joint damage and bleeding in haemophilic arthropathy: a proposal of a score. Haemophilia : the official journal of the World Federation of Hemophilia. 2011;17(1):112-7.

173. Nuss R, Kilcoyne RF, Geraghty S, Shroyer AL, Rosky JW, Mawhinney S, et al. MRI findings in haemophilic joints treated with radiosynoviorthesis with development of an MRI scale of joint damage. Haemophilia : the official journal of the World Federation of Hemophilia. 2000;6(3):162-9.

174. Lundin B, Ljung R, Pettersson H, The European Paediatric Network for Haemophilia M. MRI scores of ankle joints in children with haemophilia – comparison with clinical data. Haemophilia : the official journal of the World Federation of Hemophilia. 2005;11(2):116-22.

175. Lundin B, Babyn P, Doria AS, Kilcoyne R, Ljung R, Miller S, et al. Compatible scales for progressive and additive MRI assessments of haemophilic arthropathy. Haemophilia : the official journal of the World Federation of Hemophilia. 2005;11(2):109-15.

176. Jelbert A, Vaidya S, Fotiadis N. Imaging and staging of haemophilic arthropathy. Clinical radiology. 2009;64(11):1119-28.

177. Fotiadis N, Ekonomou I, Haritanti A, Tsatra I, Athanassiou-Metaxa M, Dimitriadis AS. The compatible MRI scoring system for staging of haemophilic arthropathy. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14(4):866-7.

178. Doria AS, Babyn PS, Lundin B, Kilcoyne RF, Miller S, Rivard GE, et al. Reliability and construct validity of the compatible MRI scoring system for evaluation of haemophilic knees and ankles of haemophilic children. Expert MRI working group of the international prophylaxis study group. Haemophilia : the official journal of the World Federation of Hemophilia. 2006;12(5):503-13.

179. Lundin B, Pettersson H, Ljung R. A new magnetic resonance imaging scoring method for assessment of haemophilic arthropathy. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10(4):383-9.

180. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Social science & medicine (1982). 1995;41(10):1403-9.

181. Kaplan RM, Bush JW. Health-related quality of life measurement for evaluation research and policy analysis. Health Psychology. 1982;1(1):61-80.

182. Arranz P, Remor E, Quintana M, Villar A, Diaz JL, Moreno M, et al. Development of a new disease-specific quality-of-life questionnaire to adults living with haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10(4):376-82.

183. Trudeau E, Dias C, Lambert T, Chambost H, Peynet J, Fressinaud E, et al. PHL11 DEVELOPMENT OF THE HAEMOPHILIA AGEGROUP SPECIFIC QUALITY OF LIFE QUESTIONNAIRE. Value in Health. 2003;6(6):744.

184. Remor E. Development of a disease-specific measure to the assessment of quality of life in adult patients living with hemophilia in Latin America: The Hemolatin-QoL. Interamerican Journal of Psychology. 2005;39(2):211-20.

185. Chevallet L, Weatherall JH, Von Mackensen S. PSY44 LINGUISTIC VALIDATION OF THE HAEMO-QOL AND HAEM-A-QOL FOR USE IN INTERNATIONAL STUDIES. Value in Health.11(3):A165.

186. Rentz A, Flood E, Altisent C, Bullinger M, Klamroth R, Garrido RP, et al. Cross-cultural development and psychometric evaluation of a patient-reported health-related quality of life questionnaire for adults with haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2008;14(5):1023-34.

187. Bullinger M, von Mackensen S, Fischer K, Khair K, Petersen C, Ravens-Sieberer U, et al. Pilot testing of the 'Haemo-QoL' quality of life questionnaire for haemophiliac children in six European countries. Haemophilia : the official journal of the World Federation of Hemophilia. 2002;8 Suppl 2:47-54.

188. Pollak E, Muhlan H, S VONM, Bullinger M. The Haemo-QoL Index: developing a short measure for health-related quality of life assessment in children

and adolescents with haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2006;12(4):384-92.

189. von Mackensen S, Bullinger M. Development and testing of an instrument to assess the Quality of Life of Children with Haemophilia in Europe (Haemo-QoL). Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10 Suppl 1:17-25.

190. Young NL, Bradley CS, Blanchette V, Wakefield CD, Barnard D, Wu JK, et al. Development of a health-related quality of life measure for boys with haemophilia: the Canadian Haemophilia Outcomes--Kids Life Assessment Tool (CHO-KLAT). Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10 Suppl 1:34-43.

191. Szende A, Schramm W, Flood E, Larson P, Gorina E, Rentz AM, et al. Health-related quality of life assessment in adult haemophilia patients: A systematic review and evaluation of instruments. Haemophilia : the official journal of the World Federation of Hemophilia. 2003;9(6):678-87.

192. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC medical research methodology. 2010;10:22.

193. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. COSMIN checklist manual 2012 [cited 2016 03 August]. Available from: http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf.

194. van Genderen FR, van Meeteren NL, van der Bom JG, Heijnen L, de Kleijn P, van den Berg HM, et al. Functional consequences of haemophilia in adults: the development of the Haemophilia Activities List. Haemophilia : the official journal of the World Federation of Hemophilia. 2004;10(5):565-71.

195. Poonnoose PM, Manigandan C, Thomas R, Shyamkumar NK, Kavitha ML, Bhattacharji S, et al. Functional Independence Score in Haemophilia: a new performance-based instrument to measure disability. Haemophilia : the official journal of the World Federation of Hemophilia. 2005;11(6):598-602.

196. von Mackensen S, Czepa D, Herbsleb M, Hilberg T. Development and validation of a new questionnaire for the assessment of subjective physical performance in adult patients with haemophilia--the HEP-Test-Q. Haemophilia : the official journal of the World Federation of Hemophilia. 2010;16(1):170-8.

197. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-40.

198. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. J Rheumatol. 1982;9(5):789-93.

199. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum. 1992;35(1):1-10.

200. Cardol M, de Haan RJ, van den Bos GA, de Jong BA, de Groot IJ. The development of a handicap assessment questionnaire: the Impact on Participation and Autonomy (IPA). Clinical rehabilitation. 1999;13(5):411-9.

201. Law M, Baptiste S, McColl M, Opzoomer A, Polatajko H, Pollock N. The Canadian occupational performance measure: an outcome measure for occupational therapy. Can J Occup Ther. 1990;57(2):82-7.

202. van Genderen FR, Westers P, Heijnen L, de Kleijn P, van den Berg HM, Helders PJ, et al. Measuring patients' perceptions on their functional abilities: validation of the Haemophilia Activities List. Haemophilia : the official journal of the World Federation of Hemophilia. 2006;12(1):36-46.

203. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire--an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. J Rheumatol. 1987;14(3):446-51.

204. RIEMSMA RP, TAAL E, RASKER JJ, HOUTMAN PM, VAN PAASSEN HC, WIEGMAN O. EVALUATION OF A DUTCH VERSION OF THE AIMS2 FOR PATIENTS WITH RHEUMATOID ARTHRITIS. Rheumatology. 1996;35(8):755-60.

205. Poonnoose PM, Thomas R, Keshava SN, Cherian RS, Padankatti S, Pazani D, et al. Psychometric analysis of the Functional Independence Score in Haemophilia (FISH). Haemophilia : the official journal of the World Federation of Hemophilia. 2007;13(5):620-6.

206. Ware JE, Snow KK, Kosinski M, Gandek B, Institute NEMCHH. SF-36 health survey: manual and interpretation guide: The Health Institute, New England Medical Center; 1993.

207. Czepa D, Von Mackensen S, Hilberg T. Haemophilia & Exercise Project (HEP): subjective and objective physical performance in adult haemophilia patients--results of a cross-sectional study. Haemophilia : the official journal of the World Federation of Hemophilia. 2012;18(1):80-5.

208. von Mackensen SC, D. Khair, K. Valenion, L. Hilberg, T. Poster no 168: Validation of the HEP-Test-Q for the Assessment of Subjective Physical Performance in Engilsh-speaking Haemophilia Patients. World Federation of Haemophilia 11-15 May; Melbourne2014.

209. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care. 1989;27(3 Suppl):S217-32.

210. Globe D, Young NL, Von Mackensen S, Bullinger M, Wasserman J, And On Behalf Of The Healthrelated Quality Of Life Expert Working Group Of The International Prophylaxis Study G. Measuring patient-reported outcomes in haemophilia clinical research. Haemophilia : the official journal of the World Federation of Hemophilia. 2009;15(4):843-52.

211. Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. Health Expectations. 2016:n/a-n/a.

212. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychological methods. 1999;4(1):84.

213. Riemsma RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. British journal of rheumatology. 1996;35(8):755-60.

214. De Joode EWP, Van Meeteren NLU, Van Den Berg HM, De Kleijn P, Helders PJM. Validity of health status measurement with the Dutch Arthritis Impact Measurement Scale 2 in individuals with severe haemophilia. Haemophilia : the official journal of the World Federation of Hemophilia. 2001;7(2):190-7.

215. Cardol M, de Haan RJ, de Jong BA, van den Bos GA, de Groot IJ. Psychometric properties of the Impact on Participation and Autonomy Questionnaire. Arch Phys Med Rehabil. 2001;82(2):210-6.

216. Czepa D, von Mackensen S, Hilberg T. Haemophilia & Exercise Project (HEP): the impact of 1-year sports therapy programme on physical performance in adult haemophilia patients. Haemophilia : the official journal of the World Federation of Hemophilia. 2013;19(2):194-9.

217. Van Meegeren ME, Van Veghel K, De Kleijn P, Van Roermund PM, Biesma DH, Lafeber FP, et al. Joint distraction results in clinical and structural improvement of haemophilic ankle arthropathy: a series of three cases. Haemophilia : the official journal of the World Federation of Hemophilia. 2012;18(5):810-7.

218. Hamilton DF, Giesinger JM, MacDonald DJ, Simpson AH, Howie CR, Giesinger K. Responsiveness and ceiling effects of the Forgotten Joint Score-12 following total hip arthroplasty. Bone Joint Res. 2016;5(3):87-91.

219. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials. 1989;10(4):407-15.

220. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Quality of Life Research.4(4):293-307.

221. Nielsen-Bohlman L. Health literacy. [electronic book] : a prescription to end confusion: Washington, D.C. : National Academies Press, c2004.; 2004.

222. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun. 2012;17 Suppl 3:141-59.

223. Hildon Z, Allwood D, Black N. Making data more meaningful: patients' views of the format and content of quality indicators comparing health care providers. Patient Educ Couns. 2012;88(2):298-304.

224. Donovan JL, Frankel SJ, Eyles JD. Assessing the need for health status measures. J Epidemiol Community Health. 1993;47(2):158-62.

225. Baker DW, Gazmararian JA, Williams MV, Scott T, Parker RM, Green D, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees. Am J Public Health. 2002;92(8):1278-83.

226. Gazmararian JA, Baker DW, Williams MV, et al. HEalth literacy among medicare enrollees in a managed care organization. JAMA. 1999;281(6):545-51.

227. Howard DH, Gazmararian J, Parker RM. The impact of low health literacy on the medical costs of Medicare managed care enrollees. Am J Med. 2005;118(4):371-7.

228. Sudore RL, Mehta KM, Simonsick EM, Harris TB, Newman AB, Satterfield S, et al. Limited literacy in older people and disparities in health and healthcare access. Journal of the American Geriatrics Society. 2006;54(5):770-6.

229. Sudore RL, Yaffe K, Satterfield S, Harris TB, Mehta KM, Simonsick EM, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med. 2006;21(8):806-12.

230. Bennett CL, Ferreira MR, Davis TC, Kaplan J, Weinberger M, Kuzel T, et al. Relation between literacy, race, and stage of presentation among low-income patients with prostate cancer. J Clin Oncol. 1998;16(9):3101-4.

231. Cooper-Patrick L, Gallo JJ, Gonzales JJ, Vu HT, Powe NR, Nelson C, et al. Race, gender, and partnership in the patient-physician relationship. JAMA. 1999;282(6):583-9.

232. Arozullah AM, Lee SYD, Khan T, Kurup S, Ryan J, Bonner M, et al. The roles of low literacy and social support in predicting the preventability of hospital admission. Journal of General Internal Medicine. 2006;21(2):140-5.

233. Farr JN, Jenkins JJ. Tables for use with the Flesch readability formulas. J Appl Psychol. 1949;33(3):275-8.

234. Johnston M, Bromley I, Boothroyd-Brooks M, Dobbs W, et al. Behavioural assessments of physically disabled patients: Agreement between rehabilitation therapists and nurses. International Journal of Rehabilitation Research. 1987;10(4, Suppl 5):205-13.

235. Lee CE, Simmonds MJ, Novy DM, Jones S. Self-reports and clinician-measured physical function among patients with low back pain: a comparison. Arch Phys Med Rehabil. 2001;82(2):227-31.

236. Hoeymans N, Wouters ER, Feskens EJ, van den Bos GA, Kromhout D. Reproducibility of performance-based and self-reported measures of functional status. J Gerontol A Biol Sci Med Sci. 1997;52(6):M363-8.

237. Badarudeen S, Sabharwal S. Assessing Readability of Patient Education Materials: Current Role in Orthopaedics. Clinical Orthopaedics and Related Research. 2010;468(10):2572-80.

7 APPENDICES

7.1 Appendix 1- COSMIN table of definitions[129]

Term			Definition
Domain	**Measurement property**	**Aspect of a measurement property**	
Reliability			The degree to which the measurement is free from measurement error
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health related-patient reported outcomes (HR-PRO) (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)
	Internal consistency		The degree of the interrelatedness among the items
	Reliability		The proportion of the total variance in the measurements which is due to 'true'† differences between patients
	Measurement error		The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Validity			The degree to which a PROM instrument measures the construct(s) it purports to measure
	Content validity		The degree to which the content of an PROM instrument is an adequate reflection of the construct to be measured
		Face validity	The degree to which (the items of) an PROM instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of an PROM instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an PROM instrument are an adequate reflection of the dimensionality of the construct to be measured
		Hypotheses testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted PROM instrument are an adequate reflection of the performance of the items of the original version of the PROM instrument
	Criterion validity		The degree to which the scores of an PROM instrument are an adequate reflection of a 'gold standard'
Responsiveness			The ability of an PROM instrument to detect change over time in the construct to be measured
	Responsiveness		Idem Responsiveness

Interpretability*			Interpretability is the degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations – to an instrument's quantitative scores or change in scores.

† The word 'true' must be seen in the context of the CTT, which states that any observation is composed of two components – a true score and error associated with the observation. 'True' is the average score that would be obtained if the scale were given an infinite number of times. It refers only to the consistency of the score, and not to its accuracy

* Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument.

7.2 Appendix 2- Haemophilia Activities List

Hemophilia Activities List

Date	: .....
Patient	: .....

Version 2005

Van Creveldkliniek / Dep. Rehab. Med.
University Medical Centre Utrecht

Introduction

This is the Hemophilia Activities List, or HAL. In this questionnaire several activities are listed that could be difficult for people with hemophilia. The aim of this questionnaire is to see how easy it is for you to do these activities

General comments

When answering the questions, it is only **your own** experience that counts. You should tick the box behind the question that best reflects your own situation.
For every activity, you are asked whether you had any difficulty in performing that activity **due to hemophilia**. There are six different response options. Answer each question by ticking the box that describes your situation.

Example:

In the past month, did you have any difficulty **due to hemophilia** with:

	n/a	Impossible	Always	Mostly	Sometimes	Rarely	Never
Using public transportation (bus, train, subway)	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

For every question you are required to tick one box. The "n/a" response option ("not applicable") can be used if you never (have to) perform that specific activity. The "n/a" option is only available for some activities. The difference between the "Impossible" and "Always" response option, is that with "Always" you are in fact able to perform that activity, but with problems and with "Impossible" you are unable to perform that activity. It is very important that you answer all questions. Even when a question seems irrelevant to you, or when you have no opinion relating to the question, please tick the box that describes your situation most closely.

It will take 5-10 minutes to finish this questionnaire.

Lying down/ sitting / kneeling / standing

In the previous month, did you have any difficulty, due to hemophilia, with:

	Impossible	Always	Mostly	Sometimes	Rarely	Never
Sitting down (e.g. on a chair or couch)	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Rising from a chair with armrests	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Rising from a chair without armrests	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Kneeling / squatting	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Bending forward	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Kneeling for a longer period of time	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Squatting for a longer period of time	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Standing for a longer period of time	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

Functions of the legs

In the previous month, did you have any difficulty, due to hemophilia, with:

	Impossible	Always	Mostly	Sometimes	Rarely	Never
Walking short distances (less than 1 kilometer / 15 minutes)	❑1	❑2	❑3	❑4	❑5	❑6
Walking long distances (more than 1 kilometer / 15 minutes)	❑1	❑2	❑3	❑4	❑5	❑6
Walking on a soft surface (e.g. on the beach or through the woods)	❑1	❑2	❑3	❑4	❑5	❑6
Walking on an uneven surface (e.g. cobblestones, high sidewalks)	❑1	❑2	❑3	❑4	❑5	❑6
Strolling / (window-)shopping	❑1	❑2	❑3	❑4	❑5	❑6
Climbing up the stairs	❑1	❑2	❑3	❑4	❑5	❑6
Climbing down the stairs	❑1	❑2	❑3	❑4	❑5	❑6
Running (e.g. in order to catch the bus)	❑1	❑2	❑3	❑4	❑5	❑6
Jumping	❑1	❑2	❑3	❑4	❑5	❑6

Functions of the arms

In the previous month, did you have any difficulty, due to hemophilia, with:

	Impossible	Always	Mostly	Sometimes	Rarely	Never
Lifting heavy objects	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Carrying heavy objects in the arms	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Fine hand movements (e.g. closing buttons)	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Reaching above your head (to pick something up from a high shelf)	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

Use of transportation

In the previous month, did you have any difficulty due to hemophilia with:

	n/a	Impossible	Always	Mostly	Sometimes	Rarely	Never
Riding a bicycle	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Getting in and out of a car	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Using public transportation (bus, train, subway)	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

Self care

In the previous month, did you have any difficulty, due to hemophilia, with:

	Impossible	Always	Mostly	Sometimes	Rarely	Never
Drying your whole body	❑1	❑2	❑3	❑4	❑5	❑6
Putting on a shirt, sweater etc.	❑1	❑2	❑3	❑4	❑5	❑6
Putting on sock and shoes	❑1	❑2	❑3	❑4	❑5	❑6
Putting on a tie or closing the top button of a shirt	❑1	❑2	❑3	❑4	❑5	❑6
Going to the toilet	❑1	❑2	❑3	❑4	❑5	❑6

Household tasks

In the previous month, did you have any difficulty, due to hemophilia, with:

	n/a	Impossible	Always	Mostly	Sometimes	Rarely	Never
Going out shopping (for food, drink etc.)	❑8	❑1	❑2	❑3	❑4	❑5	❑6
Washing the dishes, cleaning the sink	❑8	❑1	❑2	❑3	❑4	❑5	❑6
Cleaning the house	❑8	❑1	❑2	❑3	❑4	❑5	❑6
Other household tasks (ironing, making the beds)	❑8	❑1	❑2	❑3	❑4	❑5	❑6
Doing odd jobs (both in and around the house)	❑8	❑1	❑2	❑3	❑4	❑5	❑6
Gardening	❑8	❑1	❑2	❑3	❑4	❑5	❑6

Leisure activities and sports

In the previous month, did you have any difficulty, due to hemophilia, with:

	n/a	Impossible	Always	Mostly	Sometimes	Rarely	Never
Playing games (outdoors, e.g. with your children)	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Sports	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Going out (theatre / museum / movie theatre / bar)	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Hobbies	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Dancing	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Going on a holiday (active)	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
Going on a holiday ("passive"; beach-/hotel holiday)	❑ 8	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

Adaptations and using an aid

To do some activities, you might need some adaptations or an aid. This does not apply to acute bleeding episodes, when you or more or less forced to use crutches to be able to walk. In the following questions we ask you some things regarding those adaptations or aids.

Do you own a car with adaptations?

- ❑ No, I don't have a car
- ❑ No, I don't have adaptations in my car

Yes, I own a car with (multiple responses are allowed):

- ❑ Electronic windows
- ❑ Power steering
- ❑ Automatic gearbox
- ❑ The ability to sit in a wheelchair inside your car
- ❑ Brake and/or accelerator on the steering column
- ❑ Other, namely: .....
- ❑ Other, namely: .....
- ❑ Other, namely: .....

Do you use aids when performing certain activities?

- ❑ No, I don't use any aids

Yes, I use (multiple responses are allowed):

- ❑ A crutch (1 crutch / cane)
- ❑ Crutches (two)
- ❑ Wheelchair
- ❑ Rollator
- ❑ Other, namely: .....
- ❑ Other, namely: .....
- ❑ Other, namely: .....

Thank you for completing the questions on activities. To finish this questionnaire, please provide us with some personal information in the box below. The information you provide will be handled strictly confidentially.

Today's date : .....

Your date of birth : .....

What type of haemophilia do you have?

Haemophilia type* ❑ 1 Haemophilia A

❑ 2 Haemophilia B

Severity* ❑ 1 Mild

❑ 2 Moderate

❑ 3 Severe

* Please tick the appropriate box

Thank you very much for your cooperation

Scoring system

Scores can be calculated for each of the seven domains of the HAL. Additionally, three component scores can be calculated (Activities involving the Upper Extremities, Basic activities involving the Lower Extremities and Complex activities involving the Lower Extremities) as well as an overall score. Before summarizing the individual item scores, recoding is required (see Table 1); a higher raw score represents more functional limitations; possible scoring ranges are given (Table 2).

Normalized scores for the domains, components, and the full questionnaire can also be obtained. Missing values are controlled for and the possible scores range from 0 to 100, where 0 represents the worst possible functional status and 100 the best possible functional status (Table 3).

Tabel 1: Recoding

Score	Recode	Meaning
8	0	N/A
1	6	Impossible
2	5	Always problems
3	4	Mostly problems
4	3	Sometimes problems
5	2	Rarely problems
6	1	Never problems

Table 2: Scores

Score		**Items**	**Score range**
Lying / sitting / kneeling / standing	LSKS	1-8 (8)	8 - 48
Functions of the legs	LEGS	9-17 (9)	9 - 54
Functions of the arms	ARMS	18-21 (4)	4 - 24
Use of transportation	TRANS	22-24 (3)	3 - 18
Self care	SELFC	25-29 (5)	5 - 30
Household tasks	HOUSEH	30-35 (6)	6 - 36
Leisure activities and sports	LEISPO	36-42 (7)	7 - 42
Upper Extremity Activities	UPPER	* (9)	9 - 54
Basic Lower Extremity Activities	LOWBAS	** (6)	6 -36
Complex Lower Extremity Activities	LOWCOM	*** (9)	9 - 54
Sum score	SUM	1-42 (42)	42 - 252

* Items for UPPER-component: 18, 19, 20, 21, 25, 26, 27, 28, 29. (9 items)
** Items for LOWBAS-component: 8, 9, 10, 11, 12, 13. (6 items)
*** Items for LOWCOM-component: 3, 4, 6, 7, 14, 15, 16, 17, 22. (9 items)

Table 3: Normalization

Score	**Normalisatie**
LSKS	100 - ((Σ_{1-8} - valid) * (100/(5 * valid)))
LEGS	100 - ((Σ_{9-17} - valid) * (100/(5 * valid)))
ARMS	100 - ((Σ_{18-21} - valid) * (100/(5 * valid)))
TRANS	100 - ((Σ_{22-24} - valid) * (100/(5 * valid)))
SELFC	100 - ((Σ_{25-29} - valid) * (100/(5 * valid)))
HOUSEH	100 - ((Σ_{30-35} - valid) * (100/(5 * valid)))
LEISPO	100 - ((Σ_{36-42} - valid) * (100/(5 * valid)))
UPPER	100 - (($\Sigma_{18-21;25-29}$ - valid) * (100/(5 * valid)))
LOWBAS	100 - ((Σ_{8-13} - valid) * (100/(5 * valid)))
LOWCOM	100 - (($\Sigma_{3-7;14-17;22}$ - valid) * (100/(5 * valid)))
SUM	100 - ((Σ_{1-42} - valid) * (100/(5 * valid)))

"valid" = number of items scored within the specific domain/component.
Items with "n/a"-response are to be considered ***NOT*** valid

Example:

A patient completed the domain of Leg Functions as follows:

	Impossible	Always	Mostly	Sometimes	Rarely	Never
Walking short distances (less than 1 kilometer / 15 minutes)	☐ 1	☐ 2	☐ 3	☒ 4	☐ 5	☐ 6
Walking long distances (more than 1 kilometer / 15 minutes)	☐ 1	☐ 2	☒ 3	☐ 4	☐ 5	☐ 6
Walking on a soft surface (e.g. on the beach or through the woods)	☐ 1	☐ 2	☐ 3	☐ 4	☒ 5	☐ 6
Walking on an uneven surface (e.g. cobblestones, high sidewalks)	☐ 1	☐ 2	☐ 3	☒ 4	☐ 5	☐ 6
Strolling / (window-)shopping	☐ 1	☐ 2	☐ 3	☒ 4	☐ 5	☐ 6
Climbing up the stairs	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
Climbing down the stairs	☐ 1	☐ 2	☐ 3	☒ 4	☐ 5	☐ 6
Running (e.g. in order to catch the bus)	☒ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
Jumping	☒ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

Item	Score	Recode
Item 9	4	3
Item 10	3	4
Item 11	5	2
Item 12	4	3
Item 13	4	3
Item 14	Invalid	Invalid
Item 15	4	3
Item 16	1	6
Item 17	1	6

Based on the recoded scores, the raw domain score for the LEGS domain is 30 points.

The LOWBAS component encompasses the items 8-13. Item 8 scored 6 points (i.e. "Impossible"), which results in a raw component score of 6+3+4+2+3+3 = 21 points.

To normalize the scores (both domain and component scores), the formulas presented in Table 3 are used. This results in the following:

LEGS Normalized: Within the domain, 1 item is invalid, resulting in 8 valid responses out of a possible 9. This results in the following formula:

$$100-((\Sigma_{9-17}\text{-valid})*(100/(5*\text{valid}))) = 100-((30-8)*(100/(5*8))) = 100-55 = \textbf{45 points}$$

LOWBAS Normalized: Within the component, no items are invalid, resulting in 6 valid responses out of a possible 6. This results in the following formula:

$$100-((\Sigma_{8-13}\text{-valid})*(100/(5*\text{valid}))) = 100-((21-6)*(100/(5*6))) = 100 - 50 = \textbf{50 points}$$

7.3 Appendix 3- Functional Independence Score in Haemophilia

FUNCTIONAL INDEPENDENCE SCORE IN HEMOPHILIA (FISH)

Performance based instrument

Patient Name:	**Patient Code:**
	Today (dd/mm/yyyy):___/___/____.
A.Self Care	
1. Eating and grooming	○ 1 ○ 2 ○ 3 ○ 4
2. Bathing	○ 1 ○ 2 ○ 3 ○ 4
3. Dressing	○ 1 ○ 2 ○ 3 ○ 4
B. Transfers	
4. Chair	○ 1 ○ 2 ○ 3 ○ 4
5. Squatting	○ 1 ○ 2 ○ 3 ○ 4
C. Locomotion	
6. Walking	○ 1 ○ 2 ○ 3 ○ 4
7. Stairs (12 - 14 steps)	○ 1 ○ 2 ○ 3 ○ 4
8. Running	○ 1 ○ 2 ○ 3 ○ 4
Total Score	

Scores range from 1 - 4 depending on the degree of independence (see scoring key)

Comments:

7.4 Appendix 4- HEP-Test-Q

Questionnaire removed for confidential reasons

7.5 Appendix 5- ISOQOL minimum standard definitions[131]

Measurement Property	*Recommended minimum standard*
Conceptual and Measurement model	A PROM should define and describe the concept(s) included as well as the participants it targets. In addition, there should be documentation of how the concept(s) are organised into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts included in the PROM.
Reliability	The reliability of a PROM should preferably be at or above 0.70 for group-level comparisons, but may be lower if appropriately justified. Reliability can be estimated using a variety of methods including internal consistency reliability, test–retest reliability, or item response theory. The use of each method should be justified.
Validity • *Content Validity*	Should have necessary evidence supporting content validity. This should comprise of evidence that patients and experts consider the content of the PROM relevant and comprehensive for the concept, population, and aim of the measurement application. Documentation of the following should be included: (1) qualitative and/or quantitative methods used to solicit and confirm attributes (i.e., concepts measured by the items) of the PROM relevant to the measurement application; (2) the characteristics of participants recruited (e.g., race/ethnicity, culture, age, gender, socio-economic status, literacy level) with an emphasis on similarities or differences with respect to the target population; and (3) justification for the recall period for the measurement application
• *Construct validity*	Should have evidence supporting its construct validity, including documentation of empirical findings that support predefined hypotheses on the expected associations among measures similar or dissimilar to the used PROM.
Responsiveness	For use in longitudinal research study should have evidence of responsiveness. This should include empirical evidence of changes in scores

	consistent with predefined hypotheses regarding changes in the measured PROM in the target population for the research application.
Interpretability of scores	Documentation to support interpretation of scores, including what low and high scores represent for the measured concept.
Translation of the PROM	If translated to one or more languages a PROM should have documentation of the methods used to translate and evaluate in each language. Studies include evidence from qualitative methods (e.g., cognitive testing) to evaluate the quality of translations.
Patient and Investigator Burden	Must not be overly burdensome for patients or investigators to complete. The length of the PROM should be considered in the context of other PROMs included in the assessment, the frequency of PROM data collection, and the characteristics of the study population. The level of literacy require for participants to full comprehend the contents of a PROM, it should usually be at a 6th grade education level or lower (i.e., 12 year old or lower); however, it should be appropriately justified for the context of the proposed application.