From THE INSTITUTE OF ENVIRONMENTAL MEDICINE
Karolinska Institutet, Stockholm, Sweden

THE HEALTH ECONOMICS OF DUCHENNE MUSCULAR DYSTROPHY

Erik Landfeldt

Karolinska Institutet

Stockholm 2016

Published by Karolinska Institutet.
Printed by Eprint AB 2016

ISBN 978-91-7676-365-0

The health economics of Duchenne muscular dystrophy

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Erik Landfeldt

Principal Supervisor:
Associate Professor Peter Lindgren
Karolinska Institutet
Department of Learning, Informatics, Management and Ethics
Medical Management Centre

Co-supervisors:
Professor Katharine Bushby
Newcastle University
Institute of Genetic Medicine
John Walton Muscular Dystrophy Research Centre

Professor Lars Alfredsson
Karolinska Institutet
Institute of Environmental Medicine
Unit of Cardiovascular Epidemiology

Opponent:
Professor Per Carlsson
Linköping University
Department of Medical and Health Sciences
Division of Health Care Analysis

Examination Board:
Professor Finn Rasmussen
Karolinska Institutet
Department of Public Health Sciences
Child and Adolescent Public Health Epidemiology

Docent Britt-Marie Anderlid
Karolinska Institutet
Department of Molecular Medicine and Surgery
Clinical Genetics

Docent Martin Henriksson
Linköping University
Department of Medical and Health Sciences
Division of Health Care Analysis

ABSTRACT

Duchenne muscular dystrophy (DMD) is a rare neuromuscular disease characterized by progressive muscle degeneration resulting in loss of independent ambulation, serious multisystem complications, and a mean life expectancy at birth of about 25 years. Several therapies are on the horizon for DMD and it has thus become urgent to better understand the health economic context of the disease to enable assessments of cost-effectiveness and inform decisions regarding reimbursement of new health technologies. However, to date, little is known of the burden of DMD, including cost of illness and impact on health-related quality of life (HRQoL). The aim of this thesis was to estimate and describe the health economics of DMD, develop a tool measuring disease progression, and construct a model framework for economic evaluation of new treatments.

Paper I to IV were based on data collected in a cross-sectional, observational study. Patients with DMD from Germany, Italy, the UK, and the US were identified and recruited through the Translational Research in Europe – Assessment and Treatment of Neuromuscular Diseases (TREAT-NMD) network and invited with a caregiver to complete a study questionnaire online. Paper V was a modelling study synthesizing data from the preceding papers. In paper I to III, we estimated direct and indirect costs of DMD and measured patient and caregiver HRQoL and burden using several generic and disease-specific instruments. In paper IV, we developed the Duchenne muscular dystrophy Functional Ability Self-Assessment Tool (DMDSAT), a new rating-scale measuring functional ability in DMD. Item selection for the DMDSAT was led by neuromuscular specialists and the psychometric properties of the instrument were examined using Rasch analysis. In paper V, we synthesized our health economic evidence and developed a decision-analytic model framework for the assessment of the cost-effectiveness of treatments for DMD based on the DMDSAT. For comparison, we also developed two models based on stages of disease as defined in the international DMD clinical care guidelines and patients' ventilation status, respectively.

In paper I, we show that DMD is associated with a substantial cost burden to society and affected households. Across countries, the mean per-patient annual direct cost of illness was found to be 7 to 16 times higher than the mean per-capita health expenditure. Indirect and informal care costs of DMD were substantial, each component constituting between 18% and 43% of total costs.

In paper II, we show that HRQoL (as valued by the general population) in DMD is substantially impaired in relation to general population reference values, strongly negatively associated with disease progression, and in good agreement with the caregivers' subjective rating of patients' current health and mental status. Still, most children and young adults with DMD were perceived as happy and in good health by their caregivers, indicating that some domains of HRQoL remain intact through the progression of the disease.

In paper III, we show that caring for a person with DMD can be associated with a substantial burden and impaired HRQoL. Our findings suggest that caregivers to patients with DMD should be screened for depression and emphasize the need for a holistic approach to family mental health in the context of chronic childhood disease.

In paper IV, the administered version of the new rating-scale, the DMDSAT, comprised a total of eight questions covering four domains (arm function, mobility, transfers, and ventilation status). Results from the psychometric analysis show that the DMDSAT is an instrument fit for purpose to measure functional ability in ambulant and non-ambulant patients with DMD.

In paper V, we found that model structure and perspective of analysis have a substantial impact on assessments of cost-effectiveness of treatments for DMD. Our results show that the DMDSAT represents a sensitive and clinically relevant option for modelling DMD across the entire trajectory of disease in economic evaluations compared with frameworks based on conventional staging of disease progression.

In summary, this thesis provides a description of the previously unknown health economic context of DMD, including a portfolio of cost and utility data, a new tool designed to measure DMD disease severity, and a fully populated decision-analytic model framework for cost-effectiveness analysis. These data and tools should be helpful to inform health technology assessments and health economic programmes of new treatments for DMD.

LIST OF SCIENTIFIC PAPERS

This doctoral thesis is based on the following publications, referred to in the text by their Roman numerals.

I. **Landfeldt E**, Lindgren P, Bell CF, Schmitt C, Guglieri M, Straub V, Lochmüller H, Bushby K. The burden of Duchenne muscular dystrophy: an international, cross-sectional study. *Neurology* 2014; 83(6): 529–536.

II. **Landfeldt E**, Lindgren P, Bell CF, Guglieri M, Straub V, Lochmüller H, Bushby K. Health-related quality of life in patients with Duchenne muscular dystrophy: a multinational, cross-sectional study. *Dev Med Child Neurol* 2016; 58(5): 508–515.

III. **Landfeldt E**, Lindgren P, Bell CF, Guglieri M, Straub V, Lochmüller H, Bushby K. Quantifying the burden of caregiving in Duchenne muscular dystrophy. *J Neurol* 2016; 263(5): 906–915.

IV. **Landfeldt E**, Mayhew A, Eagle M, Lindgren P, Bell CF, Guglieri M, Straub V, Lochmüller H, Bushby K. Development and psychometric analysis of the Duchenne muscular dystrophy Functional Ability Self-Assessment Tool (DMDSAT). *Neuromuscul Disord* 2015; 25(12): 937–944.

V. **Landfeldt E**, Alfredsson L, Straub V, Lochmüller H, Bushby K, Lindgren P. A model for the economic evaluation of treatments for Duchenne muscular dystrophy based on the Duchenne muscular dystrophy Functional Ability Self-Assessment Tool (DMDSAT). Submitted to *Pharmacoeconomics*.

CONTENTS

LIST OF ABBREVIATIONS

6MWT	Six-Minute Walk Test
ASD	Autism-spectrum disorder
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CI	Confidence interval
CK	Creatine kinase
CMA	Cost-minimization analysis
CTT	Classical test theory
CUA	Cost-utility analysis
DMD	Duchenne muscular dystrophy
EQ-5D-3L	EuroQol EQ-5D-3L
EV	Expected value
GLM	Generalized linear model
HRQoL	Health-related quality of life
HTA	Health technology assessment
HTR	Health technology regulation
HUI	Health Utilities Index Questionnaire
ICC	Intraclass correlation
ICER	Incremental cost-effectiveness ratio
ICF	Informed Consent Form
MCS	Mental Health Component Summary Score
MD	Muscular dystrophy
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NSAA	NorthStar Ambulatory Assessment
OCD	Obsessive-compulsive disorder
OECD	Organisation for Economic Co-operation and Development
OR	Odds ratio
PCS	Physical Component Summary Score
PedsQL	Pediatric Quality of Life Inventory

PedsQL GCS	Pediatric Quality of Life Inventory 4.0 Generic Core Scales
PedsQL NMM	Pediatric Quality of Life Inventory 3.0 Neuromuscular Module
PIS	Participant Information Sheet
PPP	Purchasing power parity
PSI	Person Separation Index
PSS	Personal Social Services
PUL	Performance of the Upper Limb
QALY	Quality-adjusted life-year
QoL	Quality of life
SF-12	Short Form (12) Health Survey
SF-36	Short Form (36) Health Survey
SoC	Standard of care
TLV	Dental and Pharmaceutical Benefits Agency
TREAT-NMD	Translational Research in Europe – Assessment and Treatment of Neuromuscular Diseases
Trt	Treatment
TTO	Time trade-off
VAS	Visual analogue scale
WHO	World Health Organization
WPAI	Work Productivity and Activity Impairment Questionnaire
WTP	Willingness-to-pay
ZBI	Zarit Caregiver Burden Interview

1 INTRODUCTION

A rare disease is a medical condition that affects a very small proportion of the population (defined as one in 2,000 in the EU [1] and one in 1,250 in the US [2]). There are between 5,000 and 8,000 distinct rare diseases, the majority chronically debilitating and life-threatening, prevalent among an estimated 30 million people throughout the EU and the US, respectively [1,2]. Thus, although individually uncommon, rare diseases are associated with a substantial global health burden and constitute an important medical and social issue [3].

Up until the 1980s, rare diseases received little attention by academia and the pharmaceutical industry, both with respect to basic research and drug discovery, in part due to low expected profitability but also because of challenges associated with conducting studies of small populations [4,5]. As a result, these conditions became known as orphan diseases, and their treatments as orphan drugs, referring to their neglected position in relation to more common illnesses. In response, policy-makers in several jurisdictions, including the US, Japan, Australia, and the EU, adopted legislation such as tax credits and marketing exclusivity rights to stimulate development of orphan drugs [5,6]. In addition, global register networks and similar infrastructure have been implemented to help pool resources across different rare indications, recruit patients to research, and strengthen collaboration between advocacy groups, academia, and the pharmaceutical industry.

In the wake of these initiatives, treatments for a range of rare diseases have received marketing authorization or are currently being tested in trials. However, because of the small number of patients, many orphan drugs are associated with high prices to recoup manufacturers' research and development costs [7-9]. As a consequence, policy-makers have found it increasingly difficult to contain healthcare costs within budget constraints and simultaneously meet public demands of equal access to the most efficacious, state-of-the-art, health technologies [10]. Therefore, it has become urgent to better understand the health economic context of specific orphan drugs and their position within the conventional health technology assessment (HTA) paradigm [8,10].

A rare disease for which several promising treatments are on the horizon is Duchenne muscular dystrophy (DMD), a genetic, neuromuscular condition characterized by progressive muscle degeneration resulting in loss of ambulation and severe multi-system complications [11]. Without intervention, mean age at death is around 19 years, but following improvements in the medical management of the disease, some patients now live beyond their third or even fourth decade [12].

To date, little is known of the health economics of DMD, including the burden of disease in terms of costs and impaired health-related quality of life (HRQoL). The aim of this thesis was to estimate and describe the health economic context of DMD, develop a tool measuring disease progression, and construct a model framework for economic evaluation of new treatments.

2 BACKGROUND

2.1 DUCHENNE MUSCULAR DYSTROPHY

This section provides a summary description of the pathophysiology, epidemiology, and medical management of DMD.

2.1.1 Pathophysiology

DMD is an inherited, terminal neuromuscular disease [13]. The condition was first described in 1830 by English physician Charles Bell (1774–1842) but named after Guillaume Benjamin Amand Duchenne (1806–1875), a French neurologist who established its basic clinical and pathological features [14]. DMD is characterized by muscle weakness caused by mutations in the gene that produces dystrophin, a cell membrane protein required to maintain muscle integrity. Deficiency or complete absence of dystrophin causes plasma membrane leakage and muscle fibre degeneration leading to progressive muscle weakening. DMD is inherited in an X-linked recessive pattern from female carriers to males, although females with the disease have been described in the literature [11,15]. Approximately two-thirds of all cases have genetic origin, with the rest accounted for by sporadic mutations [16,17].

Patients with DMD are diagnosed around the age of five years, but many boys show symptoms earlier due to proximal muscle weakness resulting in delayed physical milestones (e.g. walking, running, and climbing stairs), a waddling gait, and the use of Gowers' manoeuvre when arising from the floor. Creatine kinase (CK) concentrations are markedly increased in patients with DMD and the pathway to diagnosis typically starts with testing for CK on dried blood spots. Muscle biopsy may also be performed to assess if the dystrophin protein is absent, but gene analysis is usually conducted to establish a definitive diagnosis of the disease [11,15].

As the disease progresses, patients' functional ability diminish rapidly and affected children become non-ambulatory usually in their early teens and eventually need assistance to carry out most activities of daily living, including feeding, dressing, and toileting. It is also common for patients with DMD to have some degree of mental impairment and learning difficulties, and many boys suffer from serious mental health comorbidities, including autism-spectrum disorder (ASD) and obsessive-compulsive disorder (OCD) [11]. Untreated, the mean age at death is around 19 years, but the introduction of glucocorticoid therapy, proactive cardiac management, and nocturnal ventilatory support has prolonged life expectancy into the third and sometimes fourth decade of life. Respiratory failure and cardiomyopathy are common causes of death [11,12].

2.1.2 Epidemiology

Data on the incidence of DMD is scarce and varies across studies depending on diagnostic criteria. Based on observations from new-born screening programs in Belgium, Canada, Cyprus, France, Germany, New Zealand, the UK, the US, and Wales, comprising cases of

disease confirmed through genetic testing or muscle biopsy, the incidence of DMD has been estimated at between 1 in 3,802 and 6,291 live male births [18,19]. Three additional studies, in which it is uncertain if all cases where confirmed, reported an incidence of between 1 in 5,319 and 9,337 live male births [20].

To date, only one study, a retrospective patient chart review conducted in Northern England [21], has estimated the prevalence of DMD based exclusively on cases confirmed through genetic testing or muscle biopsy. The reported point prevalence among males was 8.29 (95% confidence interval [CI]: 6.90–9.88) per 100,000. A recent longitudinal study [22] comprising six sites in the US estimated the prevalence of definite or probable cases of DMD (which also comprise cases not confirmed through genetic testing or muscle biopsy) at 10.20 (9.20-11.20) per 100,000 male individuals. Point prevalence estimates from other studies using different diagnostic criteria in which cases were not confirmed (or information about disease confirmation is missing) range from 0.95 (0.80-1.12) in South Africa [23] to 7.13 (5.16–9.60) in Japan [24] per 100,000 male individuals. In addition, the prevalence in Swedish male children less than 16 years old has been estimated at 16.80 (11.40–23.80) [25], Chinese male children less than 19 years old at 9.81 (7.52–12.58) [26], and Estonian male children less than 20 years old at 12.76 (8.26–18.84) [27] per 100,000.

As noted above, the prognosis for survival in DMD has improved dramatically during the last decades. Mean life expectancy at birth for UK patients who died in the 1960s has been estimated at 14 years, markedly lower than the median survival (i.e. the point in time when 50% of patients are still alive) of between 25 and 35 years recorded for patients who received ventilation support (introduced in most settings in the late 1980s or early 1990s) [12,28,29]. However, given that these median estimates were derived for patients who had in fact survived up until they required and received ventilation support (due to stratification in the analysis phase or restriction of the study sample population), they would be expected to be biased and should be interpreted with caution. To date, data on the mean life expectancy for patients receiving ventilation support is not available due to insufficient follow-up time.

2.1.3 Medical management

DMD is associated with extensive morbidity that necessitates a multidisciplinary approach to manage the disease. International clinical care guidelines for DMD [11,30] were published in 2010 and provide integrated treatment recommendations for both preventive and active interventions to address the primary and secondary manifestations and complications throughout the disease progression. Table 1 provides a summary overview of the key components of the medical management of DMD.

At present, there is no cure for DMD. Glucocorticoids (e.g. prednisone and deflazacort) have been shown to slow down the rate of deterioration in muscle strength and are considered for patients between four to six years of age when motor function reaches a plateau. Indications for glucocorticoids in non-ambulatory patients are more relative than absolute due to insufficient efficacy and safety evidence [11]. However, several new therapeutic strategies

have been developed in animal models and human trials of some of these interventions have started. Examples of treatment approaches that have been investigated include exon skipping (e.g. drisapersen and eteplirsen), mutation suppression, utrophin upregulation, myostatin inhibition, and insulin growth factors [31,32].

Table 1: Summary of DMD care guidelines

Care component	Frequency (minimum)	Practitioner	Aims
Neuromuscular management	Every six months	Neuromuscular specialist	Evaluation of disease status and progression (e.g. muscle strength, function, and range of movements); anticipatory planning of future developments and prevention of complications (e.g. scoliosis); ensuring immunisation schedule (e.g. varicella zoster, pneumococcal vaccine, and flu jab); planning pharmacological interventions (e.g. glucocorticoid therapy); efficacy and side-effect management of pharmacological treatments; evaluation of psychological issues and daily activities.
Cardiac management	At diagnosis; annually in ambulatory patients and every six months in non-ambulatory patients	Cardiac specialist	Evaluation of cardiac function (using e.g. electrocardiography and echocardiogram) to allow timely prevention and management of complications (e.g. dyskinesia, left ventricle dysfunction and, dilated cardiomyopathy).
Respiratory management	Annually in ambulatory patients and every six months in non-ambulatory patients	Respiratory specialist	Evaluation of respiratory function (using e.g. measurement of forced vital capacity and peak cough flow) to allow timely prevention and management of complications and trigger interventions (e.g. respiratory physiotherapy, cough assist machine, non-invasive ventilation, and tracheostomy).
Orthopaedic management	As indicated	Orthopaedist	Evaluation of surgical options for joint contractures (e.g. Achilles tendons and hips); monitoring for scoliosis and need for other interventions (e.g. spinal fusion).
Physiotherapy	Every six months	Physiotherapist	Assessment of disease progression and complications (e.g. joint contractures and spinal deformities); trigger interventions for management of complications (e.g. orthoses and referral to orthopaedic surgeon); provision of advice about stretching exercises.
Psychosocial therapy	As indicated	Therapist and/or psychologist	Provision of family support; early evaluation and timely interventions for learning and behavioural issues; evaluation of coping strategies; promoting independency and social development.
Speech and language therapy	As indicated	Speech and language therapist	Evaluation of speech developmental delay and establishment of interventions; assessment of dysphagia.

Note: Table adapted from [33].

2.2 HEALTH ECONOMICS

Health economics studies how resources are allocated and health is produced in society and the healthcare sector through the application of economic theories, methods, and concepts [34-36]. A central focus in health economics is to help understand how to make the best use of the money and other resources spent on health and healthcare. This may involve decisions about e.g. the reimbursement of a more expensive but also more efficacious pharmaceutical drug or whether or not to implement a screening program to prevent a specific disease. Moreover, interventions that represent value for money may only do so for certain patient groups with specific disease characteristics (e.g. risk of disease complications). To study these topics, and help decision-makers formulate evidence-based health policy, health economists conduct various analyses, one of the most common being economic evaluations [34]. The outcomes of health economic research inform decision-making in healthcare and health policy predominantly through HTAs.

2.3 HEALTH TECHNOLOGY ASSESSMENT

HTA may be defined as "the systematic evaluation of the properties, effects, and/or other impacts of healthcare technology" [37]. In this context, a healthcare technology is "any intervention that may be used to promote health, to prevent, diagnose or treat disease or for rehabilitation or long-term care" [38]. The main purpose of HTA is to inform policy-making, including pricing and reimbursement decisions, for existing and new healthcare technologies. Specifically, a HTA synthesizes different pieces of evidence, including but not limited to efficacy, safety, cost, and cost-effectiveness data, to assess clinical, economic, ethical, and social aspects of healthcare technologies. As a result, HTA has been described as "the bridge between evidence and policy-making" [39]. HTA is complementary to health technology regulation (HTR), which is mandatory for marketing authorization and conducted to guarantee the safety and efficacy of healthcare technologies with the objective to prevent harm [38].

Since the origin of HTA in the US in the late 1970s [40], agencies responsible for HTA activities have been established in most countries throughout the Western world. However, the mandate and responsibilities of HTA agencies, as well as the processes through which HTA outcomes are disseminated and implemented, vary greatly across jurisdictions. For example, the HTA body in the UK, the National Institute for Health and Care Excellence (NICE), provides recommendations on the use of new and existing medicines and treatments within the National Health Service (NHS). However, the NHS is legally obliged to fund and resource health technologies recommended by NICE's appraisals. In Sweden, on the other hand, the Dental and Pharmaceutical Benefits Agency (TLV) (one of two national HTA agencies) determines the reimbursement status of medicines, medical devices, and dental care treatments (i.e. if they should be subsidized by the central government). However, in contrast to the UK, there is no guarantee that a particular technology will be used locally even if it is reimbursed by TLV, because the different county councils in Sweden have mandate to make their own recommendations. In addition, there are notable inter-country differences regarding HTA evidence requirements and assessment criteria [38].

2.3.1 Health technology assessment of orphan drugs

As mentioned in the Introduction (Section 1), legislation has been implemented in several jurisdictions to help stimulate the development of orphan drugs, that is, medicines treating rare diseases (defined as one in 2,000 in the EU and one in 1,250 in the US) [5,6]. Examples include tax credits for research and clinical development, reduced fees for regulatory applications, and marketing exclusivity rights. These interventions are based on principles of equity in access to healthcare, where individuals suffering from orphan diseases should be entitled to the same opportunity of receiving treatment as patients with more prevalent conditions [7,8]. It is well-demonstrated that these incentives have led to increased product availability of orphan drugs in both the EU and the US [9]. However, although harmonized legislation for drug development and marketing authorization exists in some regions, decisions regarding pricing and reimbursement are made at the national level [6]. As a

consequence, policies governing reimbursement of orphan drugs differ substantially between health systems, creating inequality in access to and utilization of orphan drugs between citizens of different countries. For example, only 65% of all marketed orphan drugs in 2009 were reimbursed in Sweden, compared with 100% in France [41].

A recent literature review of orphan drug legislation, regulations, and policies [6] found that in most countries (29 of 35), HTA of orphan drugs comprises evidence of cost-effectiveness from economic evaluations. However, because costs of orphan drugs usually are very high, frequently exceeding €100,000 per patient and year [42] to recoup research and development costs from small patient populations, these medicines are typically not found to be cost-effective [7-9]. Therefore, a range of other factors are also taken into consideration in HTAs of orphan drugs, including therapeutic value, budget impact, impact on clinical practice, pricing and reimbursement practices globally, patient organizations, economic importance, ethical arguments, and the political climate [6]. In Sweden, for example, reimbursement decisions for all drugs are based on three criteria: "the cost-effectiveness principle", "the human dignity principle" (all individuals have equal value and rights), and "the needs-solidarity principle" (resources should be allocated and first utilized where the need is largest) [43].

A key component of HTAs in general, and assessments of the value for money of new health technologies in particular, is evidence from economic evaluations.

2.4 ECONOMIC EVALUATION

Economic evaluations compare the costs and benefits (also referred to as consequences, outcomes, or health effects) of two or more alternatives, for example two different pharmaceutical drugs. Costs in this context refer to opportunity costs (i.e. the value of the best alternative use of the resources), as opposed to financial costs associated with monetary transactions for goods and services in the marketplace [35]. Examples of health benefits include life-years gained, reduced incidence rate of a disease complication, or improved HRQoL. There are four main methods of economic evaluation [34]:

- Cost-minimization analysis (CMA)
- Cost-effectiveness analysis (CEA)
- Cost-utility analysis (CUA)
- Cost-benefit analysis (CBA)

The listed methods all include costs in monetary units measured as opportunity costs. The main difference instead concerns how benefits are taken into account in the analysis. In CMA, evaluated alternatives are assumed to be associated with identical benefits and only costs are compared. This simple method is appropriate for comparing e.g. clinically equivalent drugs with different costs. In CEA, benefits of evaluated alternatives are measured in the same natural unit (e.g. life-years). CUA is a special case of CEA where benefits are measured in terms of quality-adjusted life-years (QALYs) (described in Section 2.7), which

enables comparison of alternatives with different type of benefits. Lastly, in CBA, both costs and benefits are expressed in terms of monetary units, which allows for the comparison of any type of benefits, however originally quantified [34].

As noted above, in all types of economic evaluation, costs constitute a core analysis input. Thus, in order to perform an economic evaluation, cost data must be collated or estimated. The sections below provide a summary overview of how costs are estimated to inform economic evaluations, as well as a description of cost of illness studies.

2.5 COST STUDIES

The process of estimating costs associated with a disease can be broken down into several steps. The first step is to identify all resources used or consumed as a consequence of the disease (e.g. hospitals, physicians, and medications) [34,35]. Which resource items to include depend on the perspective of the analysis, and two main viewpoints exist: (i) the health service/healthcare perspective, in which only costs carried by the healthcare system are considered, and (ii) the societal perspective, comprising all costs irrespective of who carry them (including so called indirect costs, typically referring to costs associated with lost production due to absenteeism from work, described below).

The second step in assessing disease costs involves measuring the quantities of identified resources used or consumed as a consequence of the disease (e.g. number of hospital admissions, physician visits, or filled prescriptions) [34,35]. Resource data may be extracted from registers, patient charts, or case report forms from clinical trials, or collected directly from patients using questionnaires. However, many rare diseases, including DMD, do not have a unique disease classification code (in e.g. ICD-9 or ICD-10). As a result, it is usually not possible to identify resource use for patients with DMD in e.g. national administrative registers or claims databases.

The third step in estimating costs is to collate price data for each resource type [34,35]. All resources are valued at their economic value, that is, their opportunity cost. A number of approaches can be used to estimate the opportunity cost depending on data availability and required level of precision. However, in most studies, national price lists from hospitals and/or market prices are typically used, although they may not convey the true opportunity cost due to e.g. distortions in the market place such as monopoly competition, taxes, and trade barriers.

For non-tradable goods (i.e. resources not sold in the marketplace), e.g. lost production due to absenteeism from work and informal care by family members, which may be particularly relevant for paediatric diseases, other valuation techniques are required. The main approach to estimate production losses in health economic research is the human capital approach, in which the loss in production is valued according to the cost of employment (i.e. the national mean gross income plus employer’s costs and social fees) [34,35]. An alternative valuation method includes the friction cost method, described elsewhere [44]. In addition, several methods exist to estimate the economic value of informal care [45,46]. However, for this

resource, it is often necessary to differentiate between the time spent providing informal care instead of working and the number of hours of leisure time devoted to informal care. The former is typically valued using the human capital approach. The latter, on the other hand, could be valued in a number of ways, e.g. at the cost of a nurse visit or the value of travel time savings. Ultimately, the choice of method depends on the analysis objective, perspective, and audience.

The final step to obtain the cost associated with a resource is to multiply each resource use quantity for a relevant time unit, e.g. a month or a year, with the estimated price. Included resource costs are then summed to obtain the total cost of illness.

In addition to estimating costs for use in economic evaluations, cost studies may also be conducted to assess the total cost of a disease to society. These studies, typically referred to as cost of illness or burden of illness studies, do not provide any guidance in terms of resource allocation (for which data on benefits also are included) but may be useful for highlighting the importance of a disease and help identify drivers or categories of costs. Cost of illness studies are either incidence-based or prevalence-based [34,35]. In the former approach, lifetime costs are estimated for a cohort of patients (e.g. for all patients diagnosed in a specific year). In a prevalence-based cost study, annual costs are estimated for all patients in given year. In practice, prevalence-based studies are more common because they require less data. Moreover, cost of illness studies can be conducted using a "top-down" or "bottom-up" approach. The former involves using aggregated disease cost data to estimate total costs for a given prevalence sample, seldom utilized due to lack of precision and data availability. In the much more common "bottom-up" approach, per-patient costs are estimated from a representative sample, using data collected retrospectively (from e.g. registries or patient charts) or prospectively [34,35].

2.6 QUALITY OF LIFE

A central aim of the palliative medical management of incurable diseases, such as DMD, is to promote and maintain patient quality of life [47]. Quality of life is also a key concept in health economics and economic evaluation [34]. The World Health Organization (WHO) defines quality of life as "individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards, and concerns" [48]. Accordingly, quality of life is a subjective trait that is unique both in its definition and attainment to each person. If you ask two individuals what they regard as quality of life, and what primarily determines their quality of life, you are likely to receive two different answers. Individuals' perception of quality of life also change over time, e.g. with age, as their health changes, or as they find themselves in new cultural settings. In addition, quality of life depends on past experiences and beliefs, including previous perceptions of what constitute quality of life [49-51].

For these reasons, measuring quality of life is challenging. Yet, quality of life has emerged as a key outcome in many settings, e.g. as endpoints in clinical trials, as well as in health

economic research including economic evaluations. To facilitate measurement and interpretation of quality of life in the context of medical science, the concept HRQoL has been proposed to distinguish between quality of life in the wider sense and quality of life concerning health and illness.

2.6.1 Health-related quality of life

HRQoL is a multidimensional construct referring specifically to the individual's perception of the impact of health and illness on three aspects (also known as "domains") of life: (i) physical, (ii) mental, and (iii) social [49,50]. These aspects originate from WHO's definition of health: "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [52].

There are two main categories of rating-scales (also referred to as "instruments" or "tools") developed to measure HRQoL: (i) generic, and (ii) disease-specific. The former are designed to be broadly applicable across conditions regardless of severity or treatments, whereas the latter usually focuses on measuring HRQoL in patients with a specific disease. A range of different generic HRQoL instruments exist, but the most common include the EuroQol EQ-5D-3L (EQ-5D-3L) [53], the Short Form (36) Health Survey (SF-36) [54], the Health Utilities Index Questionnaire (HUI) [55], and the Pediatric Quality of Life Inventory 4.0 Generic Core Scales (PedsQL GCS) [56] (Table 2).

Table 2: Examples of generic health-related quality of life rating-scales

	The EQ-5D-3L [53]	The SF-36 [54]	The HUI [55]	The PedsQL GCS [56]
Age restriction[a]	≥12 years	≥18 years	≥13 years	≥5 years
Domains	Mobility; self-care; usual activities; pain/discomfort; anxiety/depression.	Vitality; physical functioning; bodily pain; general health perceptions; physical role functioning; emotional role functioning; social role functioning; mental health.	Vision; hearing; speech; ambulation; dexterity; emotion; cognition; pain.	Physical; emotional; social; school functioning.
No. of items	5	36	18	23
No. of levels	3	3-6	5-6	5
Recall period	None (time of completion)	1 or 4 weeks	1, 2, or 4 weeks	4 weeks

Note: EuroQol EQ-5D-3L (EQ-5D-3L). Short Form (36) Health Survey (SF-36). Health Utilities Index Questionnaire (HUI). Pediatric Quality of Life Inventory 4.0 Generic Core Scales (PedsQL GCS).
[a] *Self-assessment.*

Only one disease-specific rating-scale measuring HRQoL in DMD, a module to the PedsQL, has been published. The instrument, referred to as the PedsQL 3.0 DMD Module Scale [57], comprises 18 items (i.e. questions), each described in five levels (i.e. response categories), across four domains: (i) daily activities, (ii) treatment, (iii) worry, and (iv) communication. Questions within the PedsQL 3.0 DMD Module Scale are transformed to an ordinal scale ranging from 0 to 100, where a higher score indicates higher HRQoL. A validation article [57] was published in 2012 but to date, the module has not been used in any published research.

The PedsQL 3.0 Neuromuscular Module (PedsQL NMM) [58] is a rating-scale not specific to DMD but neuromuscular diseases in general. The instrument consists of 25 items, each

described in five levels, covering three domains: (i) About my/my child's neuromuscular disease (17 questions related to the disease process and associated symptomatology), (ii) Communication (three questions related to the patient's ability to communicate with healthcare providers and others about his/her illness), and (iii) About our family resources (five questions related to family financial and social support systems). Specific formats are available for ages 5 to 7, 8 to 12, and 13 to 18 years. Analogously to the PedsQL 3.0 DMD Module Scale, instrument items are transformed to an ordinal scale ranging from 0 to 100, where a higher score indicates higher HRQoL.

Due to its subjective nature, by using rating-scales such as those described above, it is possible to measure aspects of life that may or may not be associated with HRQoL, but not HRQoL *per se*. For example, although it is possible to assess if an individual experience problems with pain (using e.g. the EQ-5D-3L), it is not possible to assess if, or to what extent, pain has an impact on HRQoL. To help bridge this gap, researchers have developed a methodological framework where the concept of utility is a central topic.

2.6.2 Utilities

The term utility has different meanings in different contexts, but usually relate in some way or another to the notion of preference, where higher utility is associated with more preferred outcomes, and *vice versa* [34]. In health economics, utilities are units of measure (ordinal or interval) that reflect an individual's preferences for specific health states (also referred to as "vignettes"). A health state may refer to the health status associated with a specific disease, one or several disease complications, or general aspects of life (e.g. physical, mental, and social). Utilities usually range between 0 (interpreted as being in a health state that is equal to being dead) and 1 (interpreted as being in a health state of perfect health) [59]. Preferences for health states may be estimated using various methods, of which the most common include the standard gamble, the time trade-off (TTO), and the visual analogue scale (VAS) [34].

2.6.2.1 Standard gamble

The standard gamble, first introduced by von Neumann and Morgenstern [60], is the classic method of measuring cardinal preferences under conditions of uncertainty. Preferences derived using the standard gamble method are consistent with the axioms of rational behaviour and expected utility theory, and are referred to as "utilities" [59].

In the standard gamble approach for chronic states, participants are asked to make a choice, or gamble, between two alternatives:

(1) Reside in health state i (associated with e.g. a disease) for t years; or
(2) Reside in a health state with perfect health for t years at probability p or immediate death at probability 1- p.

The probability, p, is then varied until the respondent is indifferent between the two alternatives. The utility score for health state i for time t is p, measured on an interval scale

from 0=immediate death to 1=perfect health. Alternative specifications exist for temporary health states [59].

2.6.2.2 Time trade-off

The TTO method [61] was developed as a complement to the standard gamble. In the TTO approach for chronic states, participants are asked to make a choice between two alternatives:

(1) Reside in health state i (associated with e.g. a disease) for t years and then die; or
(2) Reside in a health state with perfect health for $v<t$ years, and then die.

Time v is then varied until the respondent is indifferent between the two alternatives. The utility score for health state i is v/t, measured on an interval scale from 0=immediate death to 1=perfect health. Utilities derived through the TTO method represent cardinal "values" based on value theory (rather than expected utility theory, the basis of the standard gamble method) because there is no element of uncertainty or risk [34,59].

2.6.2.3 Visual analogue scale

A VAS consists of a line, or scale, typically ranging from 0 (reflecting death) to 1 (or 100) (reflecting perfect health), onto which participants are asked to locate the HRQoL they associate with a particular health state. The VAS approach generate "values" rather than "utilities", as it involves neither any element of choice nor decision-making under uncertainty [34,59].

Comparing these methods, most studies report that the standard gamble yields higher estimates than the TTO, which yields higher values than the VAS [62].

2.6.3 Whose preferences should be measured?

Irrespective of method employed for estimating utilities for health states, a central decision concerns whose preferences should be measured. Two common sources include: (i) patients (i.e. individuals who currently experience the health state), and (ii) the general population (i.e. healthy individuals, *ex ante* experiencing the health state).

Research has shown that patients and the general population value health states differently. Severely ill patients have for example been found to be unwilling to trade any life-expectancy to become well, and patients with severe disease have reported similar HRQoL as their healthy counterparts [34]. In a recent comparison of patient and general public preferences in a Swedish setting, experience-based preference values were higher "in virtually all health conditions" [63]. There are several possible explanations for these differences, four of which are discussed below. First, patients and the general population may interpret health states differently based on their own experiences and information. Second, individuals from the general population may be subject to a process known as "focalism" or "focusing illusion" [64,65], where positive aspects of a health state are underestimated or neglected because the attention is focused on negative consequences. Third, patients may learn to cope with their

illness, adjust their perception and expectations of HRQoL, and adapt to their health state, a phenomenon known as "response shift" or "well-being paradox" [49]. Lastly, patients and the general population may have different vantage points, with patients assessing their health state in terms of the benefits that would result from regaining health, while member of the general population may view the health state in terms of the costs associated with losing good health [64].

Proponents of patient preferences argue that it is not certain to what degree healthy individuals may comprehend and rate hypothetical health states. Those who are in favour of using preferences of members of the general population instead argue that in a publicly funded healthcare system, preferences of the general public are most relevant [59,66]. It should be noted that given evidence of non-trivial discrepancies in preferences, as discussed above, the choice of population would be expected to have considerable consequences for applications of utilities in e.g. economic evaluations. Currently, NICE in the UK states that HRQoL in adults should preferably be measured using public preferences, whereas TLV in Sweden now prefers experience-based utilities [67,68].

2.6.4 Multi-attribute health status classification systems

Of convenience to researchers, preferences have been estimated (using the methods discussed above) for health states as defined by rating-scales measuring HRQoL, including the EQ-5D-3L, the SF-36, and the HUI. These sets of preferences are referred to as multi-attribute health status classification systems, or value sets, and measures linked to such systems as preference-based instruments. For example, the EQ-5D-3L consists of five items each described in three levels generating a total of 243 combinations of possible answers (or health states). Utilities for 42 of these health states have been estimated using the TTO method based on a sample from the general population, and these valuations have subsequently been used to predict preferences for all 243 health states described by the rating-scale [69]. Accordingly, based on the attributes included in the EQ-5D-3L, it is possible to describe the health state and corresponding utility (from the perspective of the sample population from which the utility values were estimated) for anyone who completes the instrument. A summary of the most commonly employed health status classification systems for the EQ-5D-3L [69], the SF-36 [70], and the HUI [71] are presented below in Table 3.

Table 3: Common health status classification systems

	The EQ-5D-3L UK value set [69]	**The SF-6D [70]**	**The HUI Mark 3 [71]**
Rating-scale	EQ-5D-3L	SF-36	HUI
Total no. of health states	243	18,000	972,000
No. of health states valued	42	249	25
Valuation method	TTO	Standard gamble	Standard gamble and VAS
Year of estimation	1997	1999	1994
Sample population	2,997 randomly selected members (data on age and sex not available) of the non-institutionalized adult general population in England, Scotland, and Wales.	611 randomly selected members (≥16 years of age, mean age 46 years; 61% female) of the general population in the UK.	256 randomly selected members (16-86 years of age, mean age 43 years; 61% female) of the general population in Hamilton, Ontario, Canada.
Utility scale range	-1 to 1	0 to 1	-0.36 to 1

Note: EuroQol EQ-5D-3L (EQ-5D-3L). Short Form-6 dimension (SF-6D). Short Form (36) Health Survey (SF-36). Health Utilities Index Questionnaire (HUI). Time trade-off (TTO). Visual analogue scale (VAS).

Evident from the table, value sets for the EQ-5D-3L and the HUI comprise values below 0, interpreted as health states rated worse than being dead. The methods employed for deriving these preference estimates are described in the respective articles [69-71]. However, when using utilities in economic evaluations, preference weights are usually anchored on measures of dead (utility=0) and perfect health (utility=1) [59].

2.6.5 Health-related quality of life in paediatric populations

Children, adolescents, and adults have different reference systems and thus differ in their perception of HRQoL and its determinants [49]. For example, the meaning of physical, mental, and social aspects of life to adults may not be relevant to children because of differences in activities of daily living (e.g. working versus playing), needs, expectations, etc. In addition, other domains, e.g. cognitive function, autonomy, self-esteem, body image, etc. may be equally relevant elements of HRQoL in children. Moreover, the format of adult measures may not be appropriate for children due to incomplete cognitive and emotional development [49]. For these reasons, age-specific formats of existing adult scales, as well as new instruments for measuring HRQoL in children and adolescents, have been developed. A summary of three of the most popular instruments, in addition to the HUI and the PedsQL GCS, are presented in Table 4. Among the listed rating-scales used to measure HRQoL in paediatric populations, currently only the HUI is linked to utilities.

Table 4: Examples of paediatric health-related quality of life rating-scales

	The Child Health Questionnaire [72]	**The KIDSCREEN-52 [73]**	**The EQ-5D-Y [74]**
Age restriction[a]	≥10 years	≥8 years	≥8 years
Domains	Behaviour; bodily pain; family activity; mental health; general health; role function behaviour; physical function; role emotion; role physical; self-esteem; global behaviour; global general health; family cohesion.	Physical well-being; psychological well-being; moods and emotions; self-perception; autonomy; parent relations and home life; social support and peers; school environment; social acceptance; financial resources.	Mobility; looking after myself; doing usual activities; having pain or discomfort; feeling worried, sad or unhappy.
No. of items	87	52	5
No. of levels	4-6	5	3
Recall period	Mixed	1 week	None (time of completion)

[a] *Self-assessment.*

There is no consensus regarding at which age children are capable of providing reliable self-assessments of their HRQoL, and different thresholds, e.g. five years [75] and eight years [49,74], have been proposed in the literature. It has been shown that 75% of six-year-olds use terms for feeling comfortable, excited, upset, glad, unhappy, calm, embarrassed, hateful, nervous, and cheerful. Yet, it is not until adolescence that cognitive skills become more complex, allowing the individual to formulate ideas, contemplate their future more systematically, and engage in deduct reasoning [49].

Historically, in research of HRQoL in paediatric populations, proxy-reports by e.g. caregivers or physicians were often used instead of patient self-assessments. However, the level of agreement between proxy-reports and child self-reports has been shown to be questionable. Thus, following the development of dedicated rating-scales, the consensus is that children can and should report their own HRQoL, although the use of proxy-reports also is encouraged to collect complementary HRQoL data [49,76].

2.7 QUALITY-ADJUSTED LIFE-YEARS

In the context of health economics, the primary use of utilities is in economic evaluation, more specifically CUA, where benefits of all compared alternatives are expressed in terms of QALYs. A QALY incorporates both the quality and quantity of life and is constructed by multiplying every life year with a utility weight reflecting HRQoL. As mention in Section 2.6.2, utilities usually range between 0=dead and 1=perfect health. This is to ensure that QALYs reflect units of perfect health and that the absence of a year is equivalent to a year at zero weight [59]. As an example, if an individual lives for ten years and assigns a utility weight of 0.80 to each of those years, her total number of QALYs would be 8 (10×0.80).

2.8 DECISION-RULES OF COST-EFFECTIVENESS ANALYSIS

The main outcome of an economic evaluation (comparing e.g. alternative A and B) is the expected differences in costs ($C_A - C_B = \Delta C$) and benefits ($E_A - E_B = \Delta E$). In CUA, as benefits are measured in terms of QALYs, the incremental benefit equals the number of QALYs gained (or lost) from implementing alternative A instead of alternative B. These increments (ΔC and ΔE) are subsequently used to calculate the incremental cost per incremental QALY, known as the incremental cost-effectiveness ratio (ICER):

$$\frac{\Delta C}{\Delta E}$$

Possible results of a CEA of two mutually exclusive and independent alternatives may be illustrated in a cost-effectiveness plane (Figure 1). In sum, there are four outcomes, each covered by a quadrat in the figure.

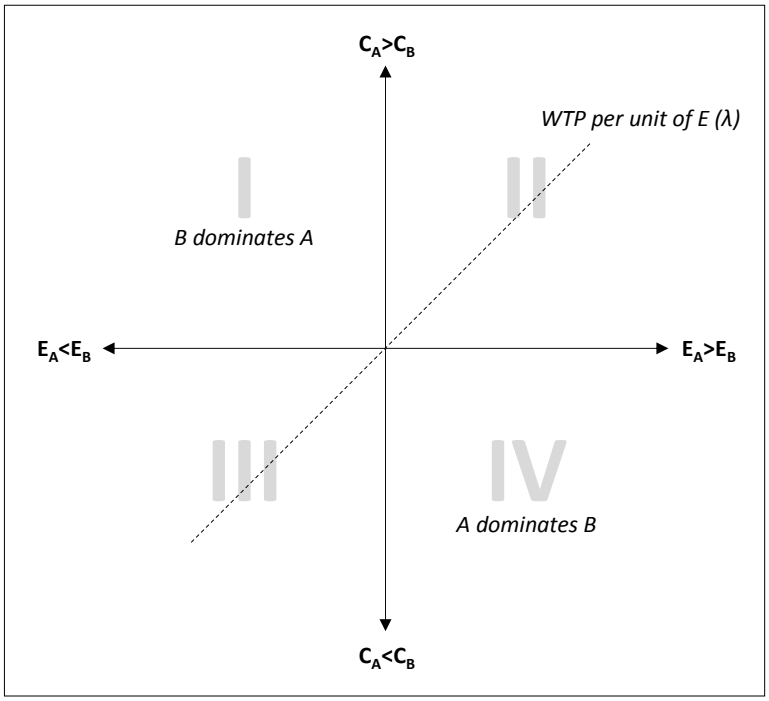


Figure 1: Example of a cost-effectiveness plane

Note: Willingness-to-pay (WTP).

In quadrant I, alternative A is dominated by B ($C_A > C_B$ and $E_A < E_B$), and in quadrant IV, alternative B is dominated by A ($C_A < C_B$ and $E_A > E_B$). If the evaluation result lies in quadrant II ($C_A > C_B$ and $E_A > E_B$) or quadrant III ($C_A < C_B$ and $E_A < E_B$), the estimated ICER must be compared against a monetary willingness-to-pay (WTP) for a QALY (γ) to test for cost-effectiveness. Alternative A is cost-effective compared with alternative B if:

$$\frac{\Delta C}{\Delta E} < \gamma$$

Alternatively, rearranged to fit the cost-benefit analysis framework, alternative A should be implemented instead of alternative B if:

$$\Delta E \times \gamma - \Delta C > 0$$

The societal WTP for a QALY varies between settings and jurisdictions, but in health economic evaluations cited values often lie between approximately €50,000 and €100,000 [77]. However, as noted in Section 2.3, HTA agencies in many jurisdictions also consider factors other than cost-effectiveness in their reimbursement appraisals.

2.9 CAREGIVER BURDEN

Throughout the Western world, including the US, Australia, Japan, and the EU, unpaid family caregivers are responsible for the vast majority of the long-term care of disabled or chronically ill patients [78-81]. Although emotionally rewarding, it is well-documented that the provision of informal care can be associated with serious adverse health effects for the caregiver, including anxiety and depression, impaired immune system function, and coronary heart disease, as well as social isolation, financial deprivation, and even premature death [78,82-84]. Compared with caregivers to elderly patients (e.g. a son caring for his father with dementia), parents raising a child with a chronic illness may face even greater challenges as they normally live together with the patient and have no choice but to fully take on the caregiver role. In addition, as a result of improved survival in many childhood diseases [85,86], including DMD [12], the duration of informal caregiving has increased considerably, in some indications from years to several decades, with increased morbidity and care needs in later stages of the patient's life.

Caregiver burden has been defined as "the extent to which caregivers perceive that caregiving has had an adverse effect on their emotional, social, financial, physical, and spiritual functioning." [87] Similarly to HRQoL, caregiver burden is thus a multidimensional construct. Accordingly, investigations of caregiver burden commonly cover assessment of overall HRQoL (e.g. using a generic HRQoL rating-scale such as the EQ-5D-3L), specific aspects of health (e.g. depression), as well as examinations of more tangible aspects of caregiving (e.g. the number of hours devoted to informal care, or loss of household income). In addition, a wide variety of rating-scales have been developed specifically to measure caregiver burden. In fact, a recent review [88] of instruments measuring the impact of informal caregiving of the elderly identified a total of 93 different instruments. The most

widely applied measure of caregiver burden is the Zarit Caregiver Burden Interview (ZBI) [87,88]. The ZBI was initially developed to measure subjective burden among caregivers of adults with dementia, but has been employed in many different indications. It contains 22 questions formulated as statements (response options range from "Never" to "Nearly Always") and each question is scored on a Likert scale (global score range from 0="low burden" to 88="high burden").

2.10 DECISION-ANALYTIC MODELLING

Decision-analysis may be described as a systematic approach to inform decision-making under conditions of uncertainty. It encompasses ways and means for dealing with complex decisions. Decision-analytic models are tools within decision-analysis. In the context of health economics and economic evaluation, decision-analytic models provide a framework for comparative analysis, synthesizing evidence from a multiple sources, translating analysis inputs into relevant outputs (e.g. costs and QALYs), and assessment of uncertainty (Table 5) [34,89].

Table 5: Features of decision-analytic models

Feature	Description
Analysis structure	Decision-analytic models provide a framework for the analysis problem and comparative evaluation (e.g. CEA of two alternative health technologies).
Evidence synthesis	A central principle of evidence-based medicine is to consider all relevant evidence. However, a single trial seldom provide all data needed for economic evaluation (e.g. efficacy, mortality, cost, and utility data for the included alternatives). Instead, data from multiple sources must be collated (from e.g. clinical trials, observational studies, and clinical experts). Decision-analytic models help synthesize and structure the many different data components that are required for the evaluation, and also provides a framework for data inter- and extrapolation (e.g. efficacy beyond the duration of the trial).
Estimation	Decision-analytic models translate inputs into analysis outputs relevant to decision-makers (e.g. the ICER in CUA).
Assessment of uncertainty	Decision-analytic models allow for assessment and variability of uncertainty associated with the decision problem, e.g. probabilities associated with treatment response, disease complications, and mortality. In addition, decision-analytic models allow for deterministic and probabilistic sensitivity analysis (described in Section 2.10.3), which informs decision-makers how model assumptions and uncertainty around the model inputs affect the outcomes (e.g. the ICER)

Note: Cost-utility analysis (CUA). Incremental cost-effectiveness ratio (ICER).

2.10.1 Decision trees

Decision trees are one of the most basic types of decision-analytic models. A decision tree models decisions and their possible consequences as branches which may be illustrated in a tree-like figure (Figure 2) [34,89]. In addition to the branches, a decision tree usually consists of three types of nodes:

- Decision nodes (squares)
- Chance nodes (circles)
- Terminal nodes (triangles)

The decision node usually marks the start of the decision tree and shows the decision point between the evaluated alternatives (e.g. drug *A* vs. drug *B*, as shown in Figure 2). The pathways that follow each alternative represent a sequence of possible events, denoted by branches emanating from chance nodes. The alternatives at each chance node are mutually exclusive (i.e. their probabilities sum to exactly one). In Figure 2, two possible events have

been modelled for each drug and each of these events have been assigned drug-specific probabilities. The end points of each pathway are denoted by terminal nodes to which pay-offs, such as QALYs or costs, are assigned.

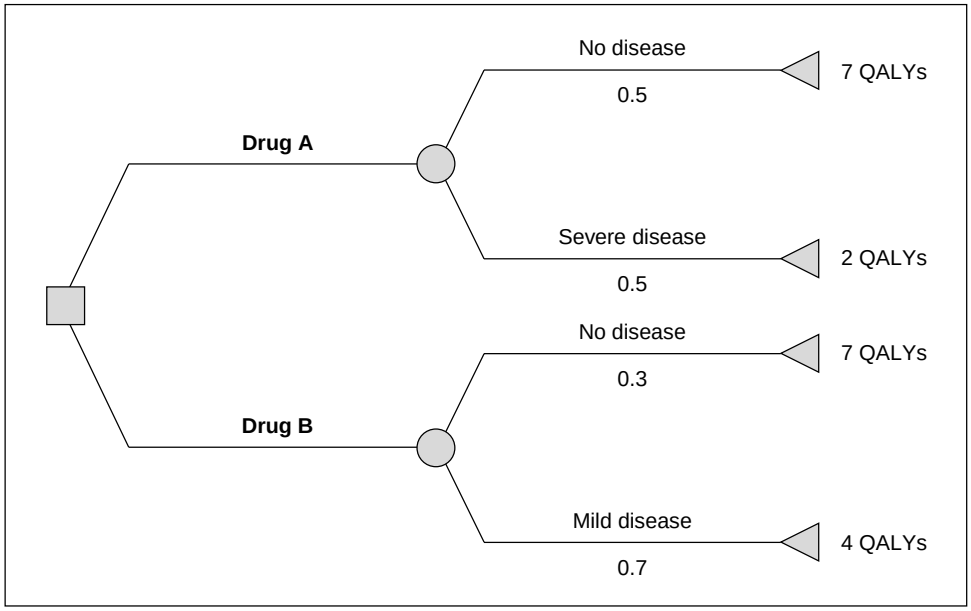


Figure 2: Example of a decision tree

Once all model data has been specified, the expected value (EV) for each alternative can be calculated. In the example in Figure 2, the EV for drug A is $0.5 \times 7 + 0.5 \times 2 = 4.5$ QALYs, and for drug B $0.3 \times 7 + 0.7 \times 4 = 4.9$ QALYs. Thus, drug B is preferred.

2.10.2 Markov models

Basic models, such as decision trees, are useful to inform simple decisions. Complex decisions, on the other hand, typically require more sophisticated models. The most common model framework in economic evaluations, such as CUA of health technologies, is the Markov model [34,89].

In a Markov model, in contrast to decision trees, a hypothetical cohort (or individual patients) are divided between a finite set of predetermined, mutually exclusive and collectively exhaustive "Markov states". These Markov states may for example represent disease complications, stages of disease progression, or other events central to the decision problem. Transitions between the Markov states take place over a series of discrete time intervals, known as cycles. The cycle length vary depending on the decision problem, but common durations include one month, six months, and one year. The distribution of patients in each Markov state in each cycle is determined by transition probabilities, and the matrix of probabilities is known as a "Markov trace". Compared with decision trees, Markov models thus allow for the explicit incorporation of a time dimension into the analysis [34,89]. Figure 3 shows an example of a Markov model of a monotonically progressive disease with four Markov states:

(1) Mild disease;
(2) Moderate disease;

(3) Severe disease; and
(4) Dead (an absorbing state in which patients stay for the remainder of the simulation).

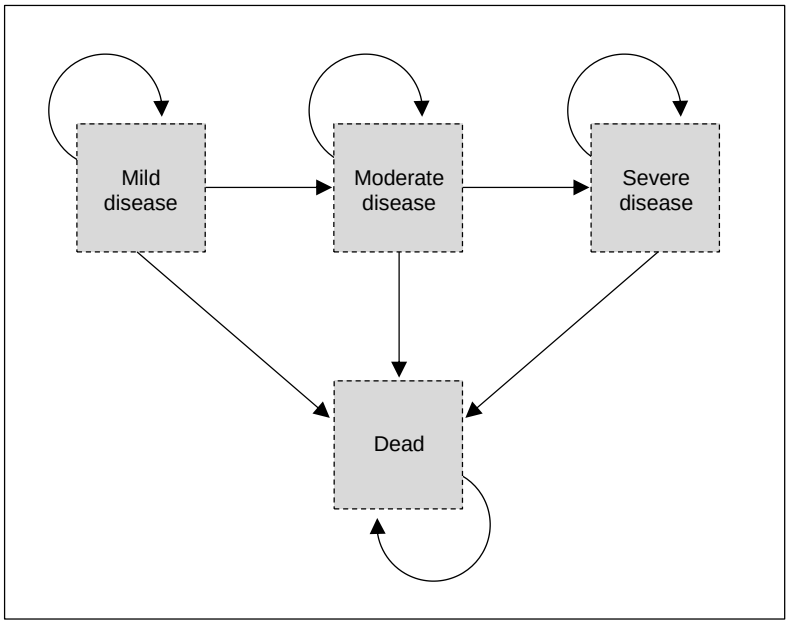


Figure 3: Example of a Markov model of a progressive disease

In a Markov model, costs and benefits (e.g. QALYs) are not attributed at terminal nodes, but instead allocated as a value per Markov state and cycle. Accordingly, at the end of the simulation, instead of calculating EVs for each evaluated alternative, the main model output is the sum of cycle-specific total costs and QALYs, usually discounted to present values. The model simulation terminates once a pre-specified criterion is realized, for example that all patients in the modelled cohort has died or a specified proportion has transition to a certain Markov state.

An important feature of Markov models is the "Markovian property", which stipulates that future events only are dependent on the current Markov state. As a result, all patients within each Markov state are treated the same irrespectively of their history within the model. In the example in Figure 3, the probability of having moderate disease is independent of the number of cycles (i.e. length of time) the patient has had mild disease, as well as the number of cycles the patient has had moderate disease in the past. It is possible, however, to include memory into the model by evaluating patients individually instead of a cohort. In such a model, sometimes referred to as a microsimulation model, transition probabilities, as well as costs and benefits, may be dependent on e.g. the number of cycles a patient has been in a specific Markov state and her history in other Markov states [34,89].

2.10.3 Uncertainty in decision-analytic models

As described in Section 2.10, assessment of uncertainty is a key feature of decision-analytic models. In the context of economic evaluations, uncertainty is usually assessed through deterministic and probabilistic sensitivity analysis (PSA) (although several other types of uncertainty have been described in the literature [90]). Deterministic sensitivity analysis explores the impact on the analysis outcome (e.g. the ICER) of changing one (one-way) or

several (multi-way) model parameters. One-way sensitivity analysis is commonly executed to analyse different "scenarios", e.g. different efficacy profiles, treatment durations, or discount rates. In addition, one-way sensitivity analysis may also be applied to better understand to which model parameter the evaluation outcome (e.g. the ICER) is most sensitive by changing all parameters by a fix amount, e.g. ±25% or ±50%. The results from such an analysis may be presented in a Tornado diagram, as shown in Figure 4.

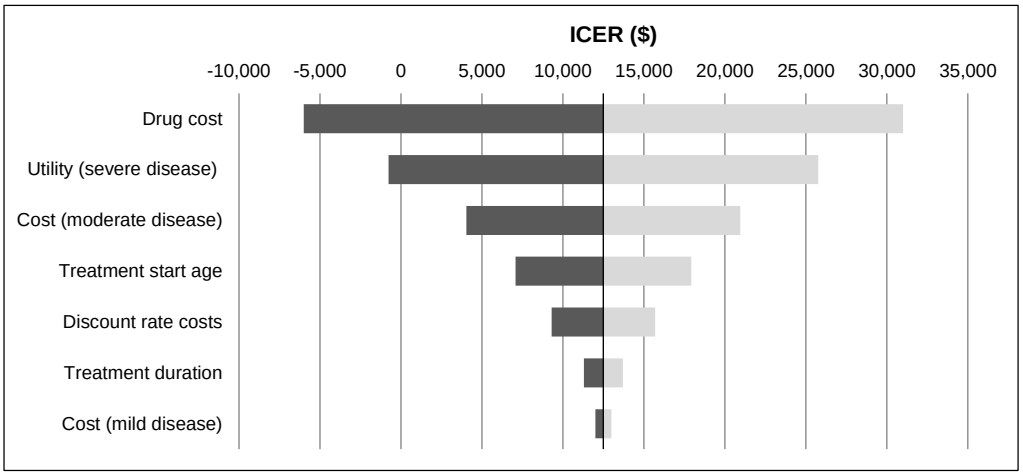


Figure 4: Example of a Tornado diagram

Note: Incremental cost-effectiveness ratio (ICER).

In contrast to deterministic sensitivity analysis, PSA is conducted to understand how uncertainty around the input data (commonly referred to as second-order uncertainty), e.g. the uncertainty around the point estimate of the mean drug efficacy, impacts the evaluation outcome. In PSA, "all parameters are varied simultaneously, with multiple sets of parameter values being sampled from *a priori* defined probability distributions" [90]. Results from PSA can be illustrated in e.g. a cost-effectiveness acceptability curve (Figure 5), which presents the number of sets that resulted in a cost-effective outcome given different WTP thresholds, or in a cost-effectiveness plane (Figure 1 in Section 2.8).

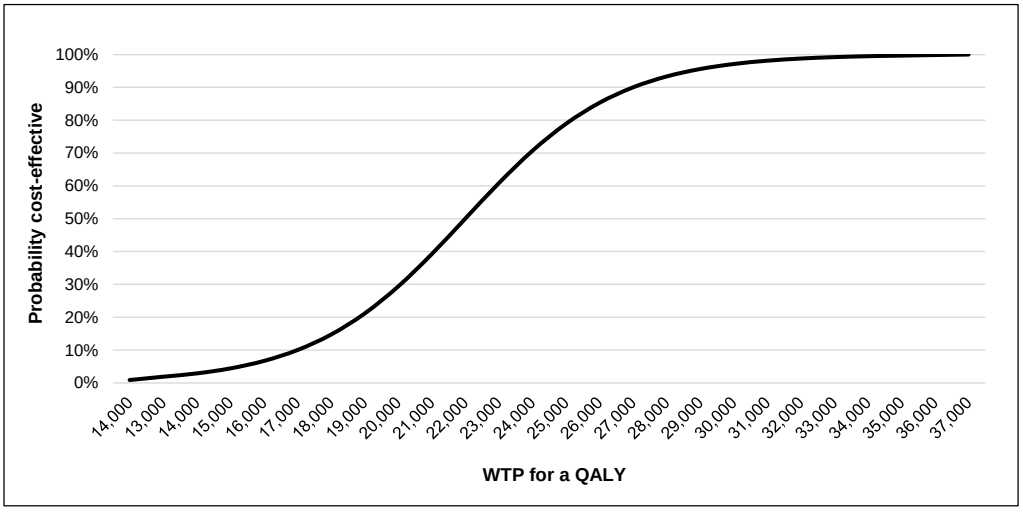


Figure 5: Example of a cost-effectiveness acceptability curve

Note: Willingness-to-pay (WTP). Quality-adjusted life-year (QALY).

2.11 THE SCIENCE OF MEASUREMENT

2.11.1 Introduction and terminology

Measurement may be defined as the assignment of numbers to categories of observations. The first stable measures originated from the fields of engineering and included units such as length, volume, and weight [91]. These measures were constructed for directly observable objects. However, some things that we wish to quantify cannot be measured directly, e.g. functional ability. Such hidden (or latent) aspects (or traits) must instead be measured indirectly through their observable manifestations [92].

A rating-scale consists of one or several items, each described in two or more levels, put together to measure a latent trait. Specifically, rating-scales seek to map out, or conceptualize, a latent trait on an abstract line, known as a "scale" or "continuum", described by its items. The purpose of this process, known as "scaling" or "abstraction", is to create a scale onto which people, or things, can be located, and is common to all measurement [91,92]. Figure 6 shows an example of a continuum of weight (i.e. a variable that is directly observable) and functional ability (i.e. a latent trait).

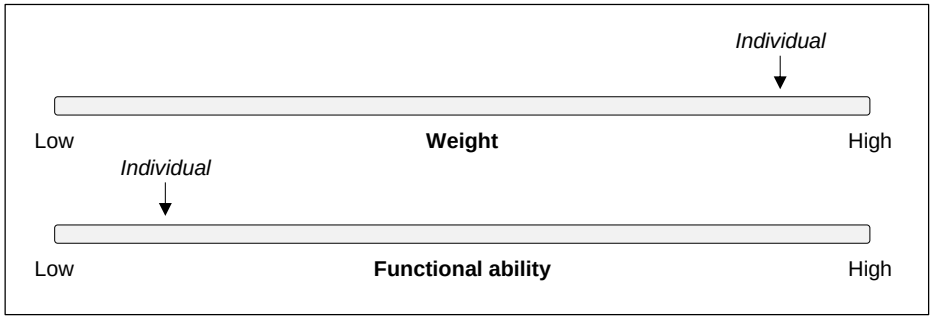


Figure 6: Example of continuums of weight and functional ability

2.11.2 Likert's method of summated ratings

Most measures of social variables construct scores by counting responses to items using Likert's method of summated ratings [92,93]. An example of a Likert item comprising five levels with level rank order in parentheses is presented below.

- Strongly disagree (0)
- Disagree (1)
- Neither agree nor disagree (2)
- Agree (3)
- Strongly agree (4)

By Likert's method, response options are ordered and allocated sequential integers and item scores are subsequently summed to produce total scores. Accordingly, instruments scored using this method are ordinal-level scales that fail to adhere to the requirements for sound measures, outlined in the section below.

2.11.3 Requirements for measures

Between 1925 and 1932, engineer and psychologist Louis Thurstone published a set of epistemological requirements for stable, objective measures of social variables. These requirements include linearity, sample-free calibration, test-free measurement, and unidimensionality [91,92].

2.11.3.1 Linearity

Quantitative observations are always ordinal (if not nominal), e.g. counting events that fulfil certain criteria. For ordinal-level measurement, order has meaning but not the relative differences between responses. Meaningful measurement, on the other hand, is based on the arithmetic properties of interval-level measurement. Interval has two meanings in this context. First, the distance in unit of measurement between response categories on the continuum is consistent within and across items. For example, in the case of the Likert item discussed above, the distance between the response category "Strongly disagree" and "Disagree" for one item is the same as the distance between "Agree" and "Strongly agree" for another item. Second, the distance between total scores is consistent across the range of the continuum represented by scale. For example, the distance between a total score of 1 and 2 is equal to the distance between a score of 3 and 4. Interval scales allow for the meaningful interpretation of total scores, calculation of mean scores, as well as changes in total scores. Figure 7 shows the difference of a change of three points between an ordinal-level and interval-level measurement system.

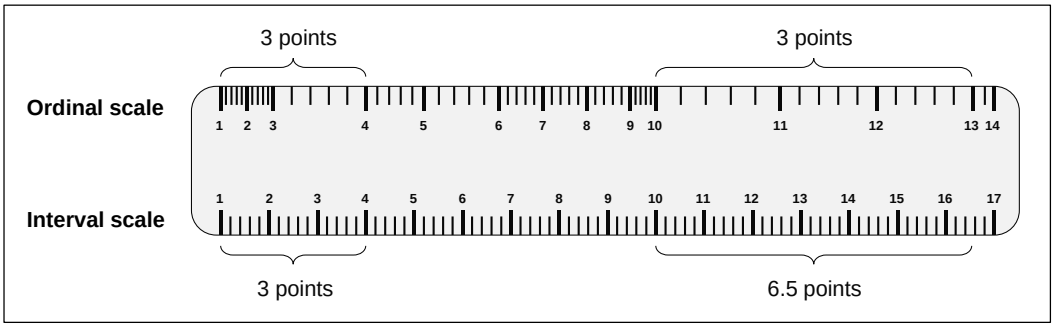


Figure 7: Example of an ordinal- and interval-level measurement system

2.11.3.2 Sample-free calibration

The rating-scale must not be affected by the sample it is measuring. That is, the continuum, and the location of items on the continuum, must be sample-independent. In short, "[t]he scale must transcend the group measured." [94]

2.11.3.3 Test-free measurement

Results for samples must be independent on the scale on which they are measured and the sample within which they are measured. Accordingly, it must be possible to measure a trait independently of the items included in the scale. For example, the measure of a person's momentary level of functional ability must not be dependent on the scale or the functional ability of others.

2.11.3.4 Unidimensionality

The measurement of any object or person describes only one attribute of the object measured. This is a universal characteristic of all measurement.

2.11.4 Traditional psychometric methods

Psychometrics may be defined as "the study of methods for measuring psychological variables" [92]. Psychometric methods determine the extent to which a quantitative conceptualization of a latent trait has been operationalized successfully. In social sciences, most rating-scales are constructed and evaluated using traditional psychometric methods, which are based on a theory known as classical test theory (CTT). CTT is underpinned by several assumptions, reviewed in detailed elsewhere [92] and summarized below. First, according to CTT, a persons' observed score (as measured by a rating-scale) is the sum of the true score (T) and an error term (E):

$$O = T + E$$

The true score is not observable, or measurable, due to measurement error. Second, CTT postulates that when a scale is administered to a person on multiple occasions, the mean of the observed scores is equal to the true score. Third, the error score (E) is uncorrelated with the true score (O), normally distributed, and has a mean of zero. Fourth, the error scores associated with two scales are uncorrelated. Lastly, the error score of one scale is uncorrelated with the true score of another scale.

The conclusions drawn from the assumptions of CTT and their mathematical proofs constitute the basis of traditional psychometrics. However, because true scores and error scores cannot be observed, the assumptions underpinning the theory cannot be tested. Therefore, CTT is regarded a weak theory [92].

In the paradigm of CTT, rating-scales have been evaluated in terms of several psychometric properties, three of the most commonly reported being reliability, validity, and targeting. Reliability describes the extent to which scale scores are free from random error [92]. It is commonly estimated using Cronbach's α [95], which is a measure of the correlation between items within an instrument, referred to as internal consistency. The idea behind the statistic is that items that measure the same trait would be expected to be correlated. Cronbach's α is dependent on the number of items in the scale, as well as the variance of the total score, and ranges between negative infinity and 1, although only positive values can be interpreted. It has been proposed that Cronbach's α should exceed 0.80 [96].

Validity refers to the extent to which a scale measures what it has been designed to measure. A range of statistical and non-statistical tests of different types of validity have been proposed and applied in the literature, including content validation, face validation, and factorial validation, but "these validation techniques are based on circular reasoning, generate circumstantial evidence, enable only limited development of construct theories, and result in only a basic understanding of what is being measured." [97].

Targeting refers to the extent to which the distribution of the latent trait in the sample matches the distribution of the latent trait measured by the rating-scale. For example, in an instrument measuring general functional ability, items covering running and climbing would be poorly targeted in a sample of non-ambulatory patients with DMD. The proportion of respondents scoring the maximum score (known as "ceiling effect") and minimum score ("floor effect") are two commonly reported measures of targeting, for which upper-limit threshold values of 15% and 20% have been suggested [98,99].

2.11.5 Modern psychometrics and Rasch analysis

Rasch measurement, or Rasch analysis, is a modern psychometric method improving the quality of social measurement by creating a model which adheres to the requirements of stable measurement (reviewed above) [92]. Specifically, given good fit of the data to the model, Rasch analysis transforms ordinal-level scales to interval-level measures. For this reason, Rasch analysis has become "the measurement standard for patient reported outcomes in general" [100].

Rasch analysis is the formal testing of a scale against a mathematical model developed by Danish mathematician Georg Rasch [101]. In brief, there are three main components to the theory of Rasch measurement. First, the response from person n to item i is governed by two factors only:

(1) Person ability, θ_n (e.g. level of functional ability); and
(2) Item difficulty, δ_i (e.g. the level of functional ability expressed by the item).

The probability that a person will affirm an item is a function of the distance between person ability and item difficulty, that is, $\theta_n - \delta_i$. Alternatively, expressed using natural logarithms:

$$0 \leq e^{(\theta_n - \delta_i)} \leq \infty$$

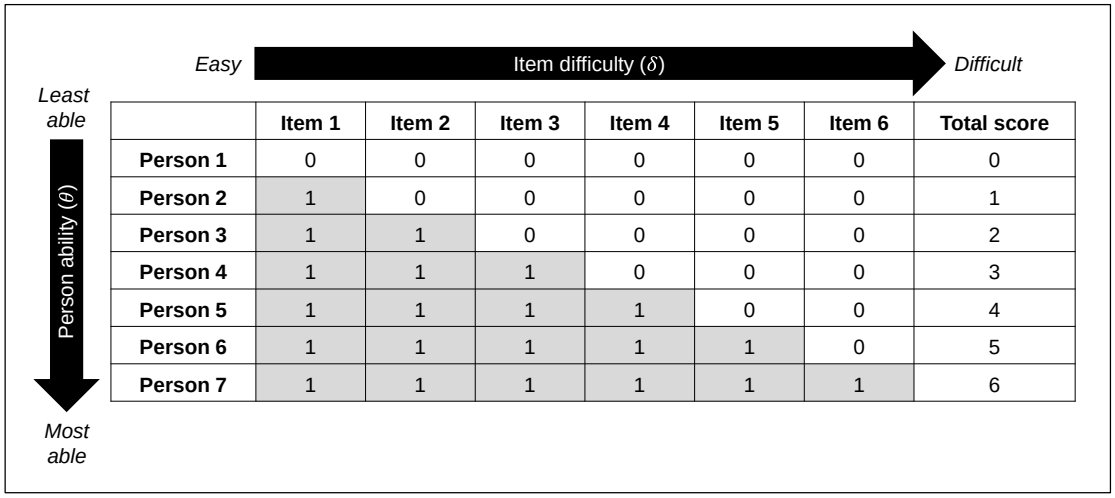
Or as an odds ratio:

$$0 \leq \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}} \leq 1$$

Accordingly, the probability that person n will affirm item i given ability θ_n and item difficulty δ_i is:

$$P_i(\text{affirm}; \theta_n) = \frac{e^{(\theta_n - \delta_i)}}{1 + e^{(\theta_n - \delta_i)}}$$

The second component of the Rasch model is a probabilistic form of the Guttman response pattern, which states that if a person affirms a task then there is a high probability that easier tasks will also be affirmed (e.g. that a person who is able to run longer distances would also be expected to be able to walk longer distances). An example of a Guttman response pattern is shown in Figure 8.



	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Total score
Person 1	0	0	0	0	0	0	0
Person 2	1	0	0	0	0	0	1
Person 3	1	1	0	0	0	0	2
Person 4	1	1	1	0	0	0	3
Person 5	1	1	1	1	0	0	4
Person 6	1	1	1	1	1	0	5
Person 7	1	1	1	1	1	1	6

Figure 8: Example of a Guttman response pattern

Note: 1=Affirmed; 0=not affirmed.

The last component to the theory of Rasch measurement is Rasch's criterion of invariance, where item locations can be estimated independently of the distribution of person locations on the continuum, which ensures that results for scales are sample independent and results for samples are scale independent [92].

The Rasch analysis output consists of an interval-level scale or metric (logit scale) to which both respondents and items are located following the three listed analysis components. In addition, Rasch analysis provides a unified approach to test several important measurement issues, including disordered thresholds (which occurs when respondents have difficulty discriminating between levels of an item given their ability) and differential item functioning (which occurs when, at the same level of ability, response to a particular item differs by a factor, e.g. gender) [100,102,103].

3 LITERATURE REVIEW OF PREVIOUS RESEARCH

This section presents results from a targeted literature review of previous research of DMD cost of illness, HRQoL in patients with DMD, caregiver HRQoL and burden in DMD, measures of disease progression in DMD, and model frameworks for economic evaluation in DMD. The review was performed in PubMed and Web of Science.

3.1 DMD COST OF ILLNESS

No previous study has estimated costs associated with DMD. However, two studies have estimated costs in mixed populations of patients with different neuromuscular disorders, in which DMD may have been included (Table 6).

Table 6: Previous research of costs associated with muscular dystrophies

Authors	Sample	Data sources	Costs	Main findings
Ouyang et al. (2008) [104]	572 US patients with congenital hereditary muscular dystrophy or hereditary progressive muscular dystrophy.	Commercial claims data.	Direct medical costs (in 2004 US dollars).	The mean per-patient annual medical expenditure was estimated at $20,467.
Larkindale et al. (2014) [105]	1,966 US patients with hereditary progressive muscular dystrophy.	Commercial and Medicare claims data and a survey.	Direct medical and non-medical costs; loss of family income (in 2010 US dollars).	The mean per-patient annual medical cost was estimated at $22,533, non-medical cost at $12,939, loss of family income at $15,481, and total cost of illness at $50,952.

Ouyang et al. [104] assessed direct medical costs (comprising outpatient visits, prescription drug claims, and hospitalizations) for a privately insured US cohort of patients (30 years of age or younger) with congenital hereditary muscular dystrophy or hereditary progressive muscular dystrophy based on claims data from 2004. They estimated the mean per-patient annual medical expenditure for men and women at $20,467 (in 2004 US dollars).

Larkindale et al. [105] estimated direct medical, non-medical, and indirect costs (measured as household income loss) for US patients with hereditary progressive muscular dystrophy based on commercial and Medicare claims data and a survey. The authors estimated the mean per-patient annual medical cost for these diseases at $22,533 (in 2010 US dollars). Non-medical cost, comprising e.g. costs associated with adaptations to the home or car and professional caregiving, were estimated at $12,939, and loss of family income at $15,481. In total, the mean per-patient annual cost of illness was estimated at $50,952.

3.2 HEALTH-RELATED QUALITY OF LIFE IN PATIENTS WITH DMD

Previous research of HRQoL in patients with DMD has usually been based on small samples (<60 cases), recruited from a single clinic or country, and contains results from a wide range of different rating-scales (Table 7). Comparison and interpretation of findings between studies may also be challenging due to differences in patient characteristics, in particular age, and inadequate stratification across stages of disease.

A total of seven studies [57,58,106-110] have measured self-assessed HRQoL in patients with DMD using the PedsQL GCS (discussed in Section 2.6.1). Results from these reports show that DMD has a negative impact on multiple domains of HRQoL, including

psychosocial aspects of life (one study [110] did not report scores). Specifically, mean self-reported PedsQL GCS scores (ranging from 0 to 100 on an ordinal scale, where a higher scores indicates higher HRQoL) have been estimated at between 33 and 68 for the physical functioning domain, 60 and 72 for emotional functioning, 60 and 77 for social functioning, 54 and 70 for school functioning, and 53 and 67 for the total score [57,58,106-109]. Corresponding reference scores for healthy children have been estimated at between 85 and 91, 73 and 84, 80 and 89, 77 and 81, and 82 and 87, respectively [57,58,108].

Evidence that DMD has a negative impact on multiple domains of HRQoL has also been shown in caregiver proxy-assessments. Bray et al. [111] measured patient HRQoL using the Child Health Questionnaire - Parent Form 50 (which includes 50 items across 14 domains) in 34 Australian boys with DMD and found scores for all domains to be significantly lower than the general paediatric population. Similar findings were reported by Baiardini et al. [112] in a study of 27 Italian patients using the same instrument, and by McDonald et al. [106] in a study of 52 ambulatory patients using the PedsQL GCS. In contrast, Kohler et al. [113] found, using the SF-36 (discussed in Section 2.6.1), that only physical function was impaired in patients with DMD, not psychosocial aspects of HRQoL. Elsenbruch et al. [114] reported similar results for adolescent and adult patients with DMD, but not young children.

A few studies have investigated the association between disease progression and HRQoL in patients with DMD. A negative association between wheelchair use, ventilation support, and HRQoL was reported by Baiardini et al. [112] In addition, Bray et al. show in two studies [107,111] that patient HRQoL is significantly, negatively associated with physical functioning, but uncorrelated with psychosocial domains. Comparable findings were reported by Kohler et al. [113], Uzark et al. [57], and Elsenbruch et al. [114] In fact, in the two latter studies, older children and adolescents were found to have better psychosocial HRQoL compared with their younger counterparts.

The level of agreement between patient self-assessments and caregiver proxy-assessments has been investigated in several studies using the PedsQL GCS and the PedsQL NMM, reporting poor to fair [57,58], poor to fair to moderate [107], moderate [110], and moderate to good agreement [109].

One previous study [115] has assessed HRQoL in DMD using the EQ-5D-3L (discussed in Section 2.6.1), estimating the mean patient utility at 0.44 (value set not reported). However, the study only included 57 adult patients with ventilation support (mean age 27 years), and the external validity of the results with respect to the general DMD population is thus limited.

Table 7: Previous research of health-related quality of life in patients with DMD

Authors	Country; sample size	Instruments	Type of assessment	Main findings
Kohler et al. (2005) [113]	Switzerland; 35	The SF-36	Self	Compared with general population reference data, patients with DMD had lower scores for physical function, but not vitality, role-emotional, social function, or mental health.
Davis et al. (2010) [58]	The US; 44	The PedsQL GCS and the PedsQL NMM	Self and proxy	Compared with reference data for healthy peers, patients with DMD had lower scores in all domains of the PedsQL GCS.
McDonald et al. (2010) [106]	The US; 52	The PedsQL GCS and the Pediatric Outcomes Data Collection Instrument	Proxy	Compared with healthy controls, patients with DMD had lower scores across all domains.
Baiardini et al. (2011) [112]	Italy; 27	The Child Health Questionnaire - Parent Form 50	Proxy	Compared with normative data, patients with DMD had lower scores in 10 of 15 instrument domains. The use of wheelchairs and ventilators was significantly associated with lower HRQoL in the physical functioning domain.
Bray et al. (2010) [107]	Australia; 35	The PedsQL GCS	Self and proxy	Compared with reference data for healthy peers, patients with DMD had lower scores in all instrument domains.
Simon et al. (2011) [116]	Brazil; 95	The Life Satisfaction Index for Adolescents	Self	HRQoL in patients with DMD was not associated with age or disease progression.
Bendixen et al. (2012) [108]	The US; 50	The PedsQL GCS and the Children's Assessment of Participation and Enjoyment	Self	Compared with healthy controls, patients with DMD had significantly lower physical, social, school-related, and total scores, but comparable emotional domain score.
Bray et al. (2011) [111]	Australia; 34	The Child Health Questionnaire - Parent Form 50	Proxy	Compared with healthy controls, patients with cancer, and patients with cerebral palsy, patients with DMD had lower scores in all instrument domains.
Pangalila et al. (2012) [115]	The Netherlands; 57	EQ-5D-3L	Self	Patient HRQoL was low (mean EQ-5D-3L utility estimated at 0.44, value set not reported).
Uzark et al. (2012) [57]	The US; 117	The PedsQL GCS and the PedsQL DMD Module Scales	Self and proxy	Compared with reference data for healthy peers, patients with DMD had lower scores in both physical and psychosocial domains.
Elsenbruch et al. (2013) [114]	Germany; 50	The SF-36, the DISABKIDS questionnaire, and the Depression Inventory for Children and Adolescents	Self	Compared with reference data for patients with other chronic illnesses, children with DMD had lower HRQoL across all instrument domains. Adolescents and adults with DMD had lower scores only for physical domains, not psychosocial.
Hu et al. (2013) [110]	China; 56	The PedsQL GCS and the PedsQL NMM	Self	The Chinese version of the PedsQL NMM has acceptable psychometric properties.
Lim et al. (2014) [109]	The US; 63	The PedsQL GCS	Self and proxy	Good to moderate agreement in instrument scores between patient self-assessment and parent proxy-assessment.

Note: Pediatric Quality of Life Inventory 4.0 Generic Core Scales (PedsQL GCS). Pediatric Quality of Life Inventory 3.0 Neuromuscular Module (PedsQL NMM). EuroQol EQ-5D-3L (EQ-5D-3L). Short Form (36) Health Survey (SF-36).

3.3 CAREGIVER BURDEN OF DMD

A total of eight previous studies of burden and related aspects, such as HRQoL, distress, and strain, in caregivers to patients with muscular dystrophies were identified in the literature review (Table 8). Four reports present estimates for caregivers to patients with DMD only, of which one study measured caregiver HRQoL using a generic HRQoL instrument mapped to utilities (discussed in Section 2.6.2). Specifically, Pangalila et al. [115] measured HRQoL in 80 Dutch parents to adult patients with DMD (mean patient age 27 years) using the EQ-5D-3L (discussed in Section 2.6.1). The authors estimated the mean caregiver utility at 0.87, comparable to that of the general population, and found no increased prevalence of anxiety and depression. In contrast, strain and stress in caregivers to patients with DMD were reported by Baiardini et al. [112] (based on the Family Strain Questionnaire) and Nereo et al.

[117] (based on the Parenting Stress Index). In addition, Abi Daoud et al. [118] found that parents to children with DMD had lower self-esteem and higher risk of depressive episodes. Elevated risks of distress and depression among caregivers have also been reported in studies of mixed cohorts of patients with muscular dystrophies [119,120].

Table 8: Previous research of the caregiver burden in DMD

Authors	Country; sample size	Instruments	Main findings
DMD			
Chen et al. (2002) [121]	Taiwan; 30	The Chronic Impact and Coping Instrument and the coping scale (excluding the subscale of cognitive reconstruction)	Caregivers to patients with DMD had elevated levels of stress. Coping strategies involved searching for information, emotional expression, and self-blame.
Nereo et al. (2003) [117]	The US; 112	The Parenting Stress Index	Mothers to boys with DMD had elevated levels of stress, primary due to problems with social interactions.
Abi Daoud et al. (2004) [118]	Canada; 42	The Depression Scale, Distress Scale, Self-Esteem Scale, and Mastery Scale (from the National Population Health Survey)	Compared with national reference data, caregivers to patients with DMD had lower self-esteem and an increased risk of going through a major depressive episode.
Samson et al. (2009) [122]	Canada; 12	Semi-structured interviews	The experience of parental hope emerges from the cognitive appraisal of DMD; hope can help absorb the initial crisis, sustain adaptation, or prepare for the fatal outcome.
Baiardini et al. (2011) [112]	Italy; 27	The Family Strain Questionnaire	Caregiving in DMD was associated with family strain, but the strain was not influenced by the severity of disease.
Pangalila et al. (2012) [115]	The Netherlands; 80	The EQ-5D-3L, the Hospital Anxiety and Depression Scale, the Caregiver Strain Index, the Self Rated Burden Scale, the Carer Quality of Life measurement, the Utrecht Coping List, and the General Self-Efficacy Scale	Caregiving in DMD was associated with a substantial burden, but was also found to be rewarding; mean caregiver EQ-5D-3L utility (estimated at 0.87) was comparable to general population reference data (value set not reported).
Thomas et al. (2014) [123]	India; 60	The Family Burden Assessment Scale, the COPE Inventory, and the Caregiver Well-Being Scale	Caregiving in DMD was associated with a moderate burden, in particular concerning financial domains of family life.
MD			
Boyer et al. (2006) [119]	France; 56	The ZBI, the SF-36, the General Health Questionnaire, and the Hospital Anxiety Depression Scale	Caregiving in MD was associated with a high burden (mean global ZBI score estimated at 23), poor social functioning, and anxiety.
Mah et al. (2008) [124]	Canada; 55	Parenting Stress Index and the Stress Index for Parents of Adolescents	Stress levels among caregivers to patients with neuromuscular disease requiring home mechanical ventilation was comparable to the general population.
Kenneson et al. (2010) [120]	The US; 1,238	The ZBI (four-item screening version), the Brief Resilient Coping Scale, the ENRICHD Social Support Instrument, the Johnson & Johnson Stress Profile, and the Kessler psychological distress scale	Caregiving in MD was associated with high levels of caregiving demands, serious psychological distress, stress, and low life satisfaction.
Magliano et al. (2013) [125]	Italy; 502	The Family Problems Questionnaire and the Social Network Questionnaire	Being a caregiver to a patient with MD was found to be a positive, rewarding experience.

Note: Muscular dystrophy (MD). EuroQol EQ-5D-3L (EQ-5D-3L). Short Form (36) Health Survey (SF-36). Zarit Caregiver Burden Interview (ZBI).

3.4 MEASURES OF DISEASE PROGRESSION IN DMD

A summary of clinical scales commonly used to measure disease progression in patients with DMD are presented in Table 9. Evident from the table, no current rating-scale is applicable to both ambulatory and non-ambulatory patients.

Historically, primary endpoints in clinical trials of DMD drugs, as well as outcomes in natural history studies, have been defined in terms of the Six-Minute Walk Test (6MWT) [126]. The 6MWT measures the distance that a person can quickly walk on a flat, hard surface in a period of six minutes. It was originally developed as a 12-minute test, but to

accommodate the impaired functional capacity of patients with e.g. respiratory disease, a shorter version comprising only six minutes was subsequently published [127]. In the context of DMD, an important limitation of the 6MWT is that it is only applicable to ambulatory patients.

Table 9: Examples of rating-scales measuring functional ability in DMD

Rating-scale	Assessment	Description
The Vignos Scale [128]	Lower extremity function	The Vignos Scale is a single-item scale comprising ten possible grades of lower extremity function, ranging from "Walks and climbs stairs without assistance" (1) to "Is confined to a bed" (10).
The Brooke Scale [129]	Upper extremity function	The Brooke Scale is a single-item scale comprising six possible grades of upper extremity function, ranging from "Starting with arms at the sides, the patient can abduct the arms in a full circle until they touch above the head" (1) to "Cannot raise hands to the mouth and has no useful function of hands" (6).
The NorthStar Ambulatory Assessment (NSAA) [130]	Lower extremity function	The NSAA includes 17 items, ranging from standing (item 1) to running (item 17), each comprising three levels ("Normal with no obvious modification of activity", "Modified method but achieves goal independent of physical assistance from another", and "Unable to achieve independently").
The Performance of the Upper Limb (PUL) [131]	Upper extremity function	The PUL includes 22 items with an entry item to define the starting functional level, and 21 items subdivided into shoulder level (four items), middle level (nine items), and distal level (eight items) (each comprising two to seven levels).

Given the inability of existing measures to map out stages across the entire progression sequence, four categories of disease, defined first in terms of current ambulatory status and second in terms of age, were proposed in the international DMD clinical care guidelines [11] to facilitate management and monitoring of DMD:

- Early ambulatory (approx. ≤7 years)
- Late ambulatory (approx. age 8-11 years)
- Early non-ambulatory (approx. age 12-15 years)
- Late non-ambulatory (approx. age ≥16 years)

Although helpful to broadly define key stages of disease, it should be emphasized that these categories lack granularity both with respect to applications of therapies in clinical practice, as well as basis for endpoints in trials.

3.5 MODEL FRAMEWORKS FOR ECONOMIC EVALUATION IN DMD

No model frameworks for the economic evaluation of treatments for DMD was identified in the literature review.

4 DATA GAPS

Based on the literature review presented in the preceding section, the following data gaps were identified concerning the health economics of DMD.

- Estimates of DMD cost of illness.
- Estimates of patient HRQoL, in particular utilities, at different stages of disease progression.
- Estimates of the caregiver burden of DMD, including measures of caregiver HRQoL, at different stages of disease progression.
- An instrument measuring progression in DMD across the entire disease trajectory linked to costs and utilities.
- Frameworks for economic evaluation, including decision-analytic models for cost-effectiveness analysis.

5 AIMS

The overall aim of this thesis was to estimate and describe the health economic context of DMD, develop a tool measuring disease progression, and construct a model framework for economic evaluation of new treatments.

The specific aims of the thesis, illustrated in Figure 9 below, were to:

- Estimate DMD cost of illness (paper I);
- Contribute to the understanding of HRQoL in patients with DMD (paper II);
- Contribute to the understanding of the caregiver burden in DMD (paper III);
- Develop a rating-scale measuring functional ability in patients with DMD to map progression in clinical practice, measure efficacy in clinical trials, and model the disease in economic evaluations (paper IV); and
- Develop a decision-analytic model framework for cost-effectiveness analysis of treatments for DMD (paper V).

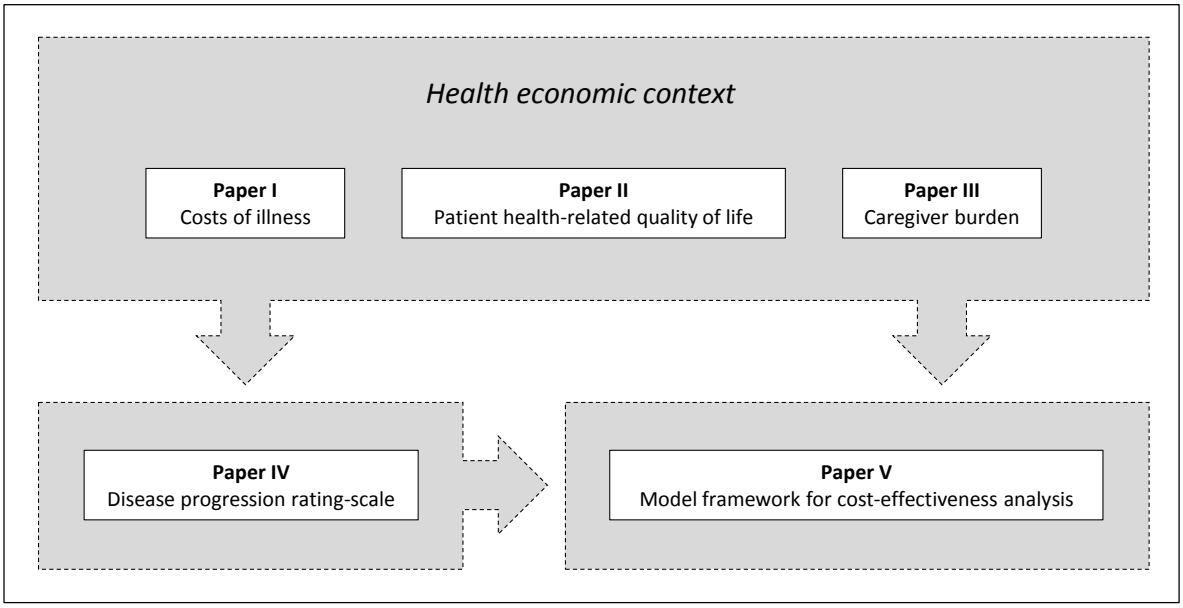


Figure 9: Illustration of the thesis aims and publications

6 MATERIALS

Paper I, II, III, and IV were based on data collected in a multi-national, cross-sectional, observational study. The sections below provide a summary description of this study. Paper V was a modelling study utilizing data derived from the preceding papers. The methods used to meet the study objectives of the different papers are presented in Section 7.

6.1 STUDY POPULATION

Patients with DMD and one of their caregivers (e.g. a parent) were identified and recruited through the TREAT-NMD network [132]. TREAT-NMD was established in 2007 as a European Union "network of excellence" to help bring advances in neuromuscular disease treatment rapidly and effectively to patients. The network has since then developed into a global infrastructure of partner organizations, e.g. patient organisations and specialist clinics, managing national patient registries. Four TREAT-NMD partners participated in the study: Ludwig-Maximilians-Universität München (Germany), Parent Project Onlus (Italy), Action Duchenne (the UK), and DuchenneConnect (the US).

To be eligible to participate in the study, patients were required to fulfil the following inclusion criteria:

- Male;
- Confirmed diagnosis of DMD; and
- Five years of age, or older.

Patients who were from Germany, Italy, the UK, or the US but resided in a different country were not eligible to participate in the study.

6.2 PROCEDURES AND COLLECTED DATA

Eligible patients were invited to complete a questionnaire administered via an online website. The questionnaire contained three parts. The first part consisted of questions regarding demographic characteristics of the patient (e.g. age, living situation, and education and work status), disease information (e.g. ambulatory status and ventilation support requirements), DMD-related healthcare resource utilization (e.g. admissions to hospitals, visits to physicians or other healthcare professionals, and medications), and patient HRQoL assessed using the HUI and the PedsQL NMM (described in Section 2.6.1). The first part of the questionnaire also contained a new rating-scale measuring functional ability in DMD, the Duchenne muscular dystrophy Functional Ability Self-Assessment Tool (DMDSAT) (paper IV). A six-month recall period was designated for all questions concerning DMD-related healthcare resource utilization, except for medication use (one month) and investments (one year).

The second part of the questionnaire consisted of questions about the caregiver (e.g. age, gender, education and work status) and the household (e.g. household composition, disposable income, co-payments, and other DMD-related expenditure). A modified version of the Work Productivity and Activity Impairment Questionnaire (WPAI) [133] was used to

assess the impact of caring for someone with DMD on work, productivity, and daily activities. Caregivers also completed the ZBI (described in Section 2.9) to assess the subjective caregiver burden. Data on caregiver HRQoL were obtained through the EQ-5D-3L (described in Section 2.6.1) and the Short Form (12) Health Survey (SF-12). The SF-12 is a short version of the SF-36 (described in Section 2.6.1) and measures functional health and well-being through a total of 12 questions, each described in three to five levels. SF-12 outcomes include two composite scores, the Physical Component Summary Score (PCS) and the Mental Health Component Summary Score (MCS), as well as eight separate scores.

In the final part of the questionnaire, patients were asked to complete age-specific self-report versions of the PedsQL NMM. We asked the caregivers to indicate if their sons completed the PedsQL NMM on their own (i.e. without any support, explanation, or suggestions), and only those who did so were included for analysis.

6.3 ETHICAL APPROVAL

Study ethical approval was granted from Ludwig-Maximilians-Universität München (Germany), Comitato Etico IRCCS E. Medea–Associazione La Nostra Famiglia (Italy), North East Research Ethics Service, NHS (the UK), and the Western Institutional Review Board (the US). Approval was also obtained from the TREAT-NMD Global Databases Oversight Committee.

7 METHODS

This section provides a summary description of the methods used to meet the study objectives of the papers included in this thesis.

7.1 DMD COST OF ILLNESS (PAPER I)

We estimated annual costs of DMD from a societal perspective including direct medical costs (e.g. hospital admissions, visits to physicians and other healthcare professionals, medical tests and assessments, medications, and emergency and respite care), direct non-medical costs (i.e. costs associated with non-medical aids, devices, and investments), non-medical community services (e.g. transportation services and home help), informal care costs (i.e. paid and unpaid informal care by the primary caregiver), indirect costs (i.e. production losses for the patient and primary caregiver, including productivity losses due to absenteeism and impaired productivity while working), and intangible costs (i.e. costs due to impaired HRQoL).

Direct medical and non-medical costs of DMD were calculated using data on resource use and national reference prices [134-142]. Country-specific costs for aids, devices, and investments were obtained through expert input from TREAT-NMD. Annual estimates were derived by assuming that a similar proportion of patients would use the same quantity of resources in any given one-month, six-month, or one-year period.

The indirect cost (i.e. costs associated with production losses) of DMD was valued according to the human capital approach at the cost of employment. Outcomes from the WPAI were used to estimate productivity losses due to absenteeism and impaired productivity while working.

We valued each hour of paid informal care according to the human capital approach. To estimate unpaid informal care, we first calculated the proportion of the caregivers' leisure time devoted to informal care, estimated using outcomes from the WPAI and data from the Organisation for Economic Co-operation and Development (OECD) on the country-specific mean daily number of hours of leisure time for an adult in the general population [143]. Each hour of leisure time was then conservatively valued at 35% of the country-specific national mean gross wage, in line with previous research and recently updated estimates of the value of travel time savings [144,145].

The intangible cost (i.e. costs due to pain, anxiety, social handicap, etc.) of DMD was estimated by assigning a monetary value to the loss in HRQoL for patients and caregivers in relation to the age- and sex-specific mean HRQoL in the general population [146,147]. The societal WTP for one year of full health, also known as one QALY, varies by method of assessment and setting and commonly cited values lie between $50,000 and $100,000 [77]. In this thesis, a WTP for a QALY of $75,000 was used in the analysis.

We related our results to the progression of DMD by classifying patients into four groups defined in terms of current ambulatory status and age (as defined in Section 3.4). A

generalized linear model (GLM) was fitted to the data to investigate if the mean per-patient annual cost of illness varied between the four ambulatory classes and to predict costs for these groups. To control for confounding effects, the GLM was adjusted for country, household income class, diagnosis of common mental and behavioural disorders, as well as a dummy variable indicating additional household member with DMD. Determinants of patient HRQoL and annual household cost burden were investigated analogously.

Results are presented in 2012 international (US) dollars calculated using Purchasing power parity (PPP) data from Eurostat.

7.2 HEALTH-RELATED QUALITY OF LIFE IN DMD (PAPER II)

We measured mean caregiver proxy-assessed HUI-derived utilities and mean caregiver proxy-assessed and patient self-assessed PedsQL NMM scores. We compared estimates across four ambulatory classes (as defined in Section 3.4). In addition, to further investigate the relationship between patient HRQoL (as perceived by the caregivers) and HRQoL as measured through the rating-scales, we also stratified our estimates by the caregivers' subjective rating of their sons' current health (five categories, ranging from "Poor" to "Excellent") and the caregivers' subjective rating of their sons' current mental status (five categories, ranging from "So unhappy that life is not worthwhile" to "Happy and interested in life").

Agreement between patients' self-assessments and caregivers' proxy-assessments of PedsQL NMM scores was investigated by estimating intraclass correlations (ICCs) from one-way random-effects models. In accordance with previous research [57,58], we considered ICCs<0.40 to indicate poor/fair agreement, 0.40 to 0.60 moderate agreement, 0.61 to 0.80 good agreement, and >0.80 excellent agreement.

We assumed the sampling distributions of the sample means to be approximately normally distributed in accordance with the central limit theorem and used Welch's t tests and analysis-of-variance models to compare estimates across strata due to heterogeneous variances.

7.3 CAREGIVER BURDEN OF DMD (PAPER III)

We assessed mean EQ-5D-3L utility scores (using the UK value set derived through the TTO method [69]), mean VAS scores, mean global ZBI scores, and mean PCS and MCS scores from the SF-12. We assessed and reported EQ-5D-3L results for anxiety and depression separately as we hypothesized that this domain would be most influenced by the caregiver role.

We compared estimates across four ambulatory classes (as defined in Section 3.4), countries, and the caregivers' subjective rating of their sons' current health and mental status with general population reference data [147,148] using Welch's t tests and analysis-of-variance models. We used logistic regression to test for differences in anxiety and depression across ambulatory classes, caregivers' ratings, and two objective measures of the caregiver burden

(i.e. annual household cost burden and number of hours of leisure time devoted to informal care) taken from paper I.

7.4 FUNCTIONAL ABILITY RATING-SCALE (PAPER IV)

We declared the underlying latent trait to be operationalized by the new tool as "functional ability" (i.e. physical and respiratory functioning), encompassing the full range within the DMD progression sequence (i.e. from the early ambulatory to late non-ambulatory disease stage). Our aim was to create a clinically and personally relevant disease-specific rating-scale that could be easily completed by the patient or a caregiver (e.g. a parent), and we therefore sought to capture the trait through manifestations in common activities of daily living (e.g. getting on and off the floor, on and off the toilet, and climbing up and down stairs). In addition, to allow further discrimination of functional ability, we also included items relating to manifestations of the lower and upper extremities, respectively, as well as current ventilation status.

Initial instrument item identification and selection was led by a group of neuromuscular experts (i.e. specialist neuromuscular physicians and physiotherapists with extensive experience in the medical management of DMD). The bank of item and level candidates was also informed by a literature review of existing measures. To capture the patient-perspective, items within a draft set were discussed with and tested on patients and caregivers as part of patients' routine clinical follow-up to assure understandability (e.g. that the items and levels made sense and were easily understood) and completeness (e.g. that all essential levels were represented, relevant, and appropriately formulated in terms of hierarchy). We also conducted a pilot study to further assure understandability and completeness.

A full Rasch analysis (described in Section 2.11.5) was conducted to assess the psychometric properties of the new rating-scale.

7.5 DECISION-ANALYTIC MODEL FRAMEWORK (PAPER V)

We synthesized our previously published health economic data in three Markov state-transition models (Figure 10). The framework of our main model, model I, was based on the DMDSAT (from paper IV). For comparison, we also developed a model based on stages of disease as specified in the international DMD clinical care guidelines (defined in Section 3.4) (model II) as well as patients' ventilation status (model III). The models were constructed in accordance with the ISPOR-SMDM Modeling Good Research Practices Task Force guidelines [149] and designed to evaluate the cost-effectiveness of a hypothetical treatment versus standard of care in a cohort of patients with DMD in a UK setting.

In each model, every cycle (i.e. every 12 months) patients had a probability to remain in the current state, progress to a more severe state, or die. In the base-case scenario, we assumed that the hypothetical intervention reduced the probability of DMD progression by 25% across model states.

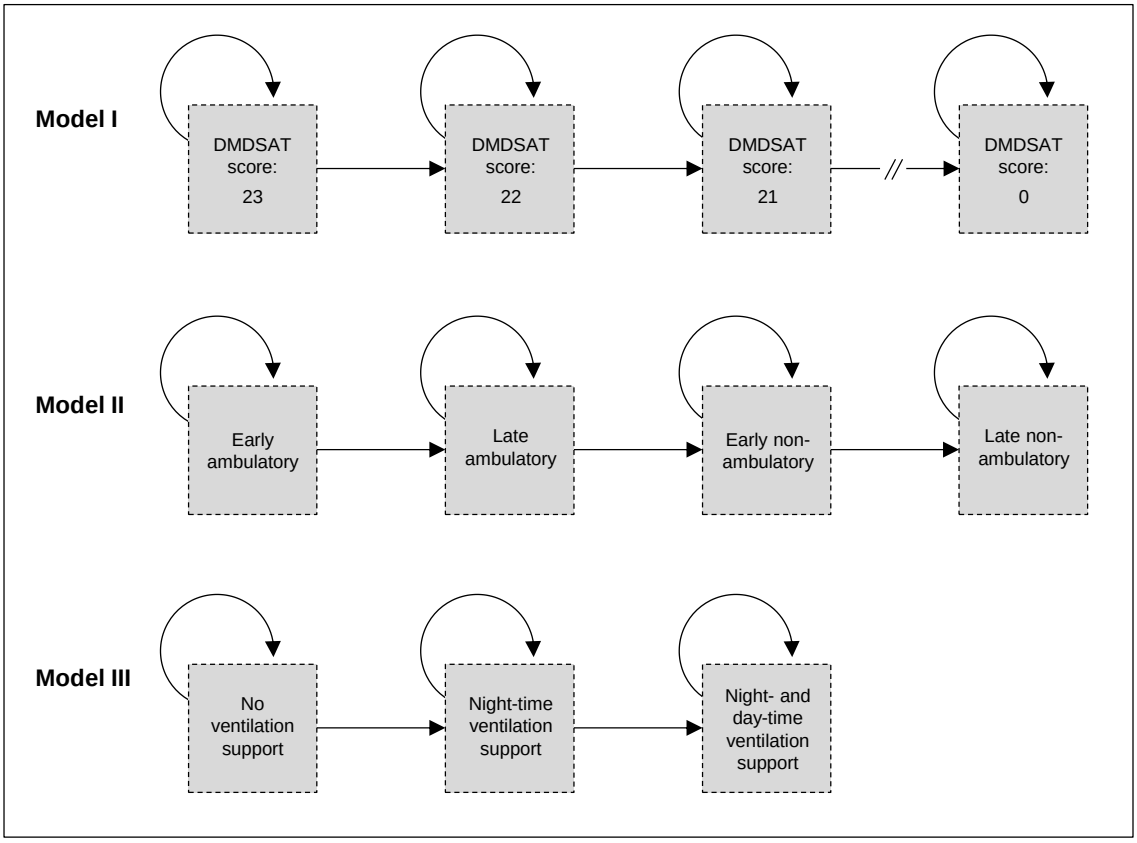


Figure 10: Schematic illustration of the DMD Markov model frameworks

Note: The absorbing health state Dead, linked to all states within each model structure, was excluded for simplicity.

Instead of assuming an annual treatment cost, we estimated and reported the maximum cost given NICE's cost-effectiveness threshold, i.e. a WTP for a QALY of £30,000 [150]. Our base-case scenario was specified in accordance with NICE's reference case [68] and included costs from the perspective of NHS and Personal Social Services (PSS) (i.e. direct medical costs) and impact of the disease on patient HRQoL. We also analysed a scenario from the societal perspective comprising all costs (i.e. direct medical and non-medical, and indirect costs) as well as the impact on patient and primary caregiver HRQoL. Main model outputs comprised the mean total lifetime discounted cost and number of QALYs, used to calculate the ICER.

We conducted deterministic one-way scenario analysis to investigate the impact of efficacy on mortality, different discount rates, and a fully curative, life-long treatment. To investigate second-order uncertainty of the input parameters, we performed PSA for a wide range of different efficacy levels (relative risk of disease progression ranging from 0.01 to 0.99, in increments of 0.01) and eight annual treatment costs (1,000x99x8=792,000 simulations in total per model) and estimated the probability of cost-effectiveness given a WTP for a QALY of £30,000.

Face validation of the appropriateness of the conceptual models (in terms of e.g. structure, health states, and comparator), model input data, and model outcomes were judged by the participating DMD experts. The validity of the computerized models was assessed through derivation of Markov traces and by comparing modelled mortality and disease progression

probabilities with the populated data. Extreme value and unit testing comprised setting model transition probabilities to 0 and 1, respectively and turning off specific cost and utility components, as well as mortality. No previous models of DMD were identified in the literature and cross validation testing was therefore not possible. However, one model of a subtype of DMD has been described as part of a NICE appraisal (discussed in Section 9.1.5).

8 RESULTS

8.1 DMD COST OF ILLNESS (PAPER I)

The mean per-patient annual total direct cost of illness was estimated at $42,360 in Germany, $23,920 in Italy, $54,160 in the UK, and $54,270 in the US (in 2012 international dollars) (Table 10). Across countries, costs associated with paid and unpaid informal care accounted for between 18% and 31% of the total cost of illness. We found the mean per-patient annual cost of illness to be significantly associated with disease progression, with an estimated adjusted mean increase of 17%, 74%, and 143% from the early ambulatory to the late ambulatory, early non-ambulatory, and late non-ambulatory class, respectively.

Table 10: Per-patient annual costs of DMD (in 2012 international $)

	Germany	Italy	The UK	The US
Hospital admissions[a]	2,080 (1,020-4,950)	1,420 (900-2,470)	2,300 (1,500-3,720)	2,220 (900-5,050)
Visits to physicians and other healthcare practitioners	3,850 (3,410-4,340)	2,590 (1,970-3,440)	8,230 (6,360-13,150)	18,210 (15,450-22,260)
Nurse	40 (10-80)	40 (10-220)	550 (300-1,160)	1,270 (650-2,530)
General practitioner	110 (80-160)	40 (30-60)	340 (220-670)	230 (180-340)
Specialist physician	330 (280-410)	170 (130-240)	3,290 (2,380-7,100)	3,730 (3,140-4,840)
Psychologist or therapist	50 (30-110)	50 (30-120)	160 (80-390)	720 (430-1,220)
Physiotherapist or occupational therapist	2,810 (2,480-3,180)	2,210 (1,610-3,020)	3,290 (2,420-5,820)	9,920 (8,220-12,030)
Other healthcare practitioner[b]	500 (360-700)	70 (50-120)	600 (370-1,400)	2,350 (1,740-3,200)
Tests and assessments	2,400 (2,180-2,680)	600 (530-690)	1,580 (1,450-1,750)	2,860 (2,660-3,070)
Medications	1,020 (770-2,000)	1,550 (890-4,650)	930 (820-1,070)	2,070 (1,720-2,710)
Non-medical community services[c]	8,920 (6,890-12,400)	2,740 (1,640-5,380)	19,250 (13,240-28,670)	7,610 (6,030-9,790)
Aids, devices, and investments[d]	5,560 (4,160-7,460)	1,850 (970-4,450)	7,520 (5,690-9,790)	7,930 (6,210-10,260)
Informal care	18,530 (16,440-20,580)	13,160 (11,270-15,280)	14,340 (13,030-15,990)	13,370 (12,060-14,930)
Total direct cost of illness	42,360 (38,640-46,880)	23,920 (20,420-28,300)	54,160 (47,310-63,510)	54,270 (48,740-62,220)
Indirect cost of illness (production losses)	20,770 (17,670-24,250)	18,220 (15,430-21,380)	18,700 (16,280-21,150)	21,550 (18,490-24,720)
Total annual cost of illness	63,140 (57,600-69,710)	42,140 (36,940-47,730)	72,870 (64,350-84,150)	75,820 (69,350-85,270)
Intangible costs	45,860 (41,630-50,160)	37,980 (32,400-43,550)	46,080 (42,360-50,050)	45,080 (41,100-48,260)
Total burden of illness	109,000 (100,390-119,510)	80,120 (71,030-89,190)	118,950 (108,280-132,710)	120,910 (111,460-130,770)

Note: Data presented as mean (95% CI), rounded to nearest ten.

[a] Including emergency and respite care.

[b] Care coordinator/care advisor, dentist, dietitian/nutritionist, and speech/language/swallowing therapist.

[c] Home help, personal assistants, nannies, and transportation services.

[d] Include investments to and reconstructions of the home (e.g. adaptations for wheelchair accessibility).

Indirect costs of DMD were substantial. Specifically, across countries, between 27% and 49% of caregivers had reduced their working hours or stopped working completely due to their sons' DMD, and for employed caregivers, the mean overall work impairment (loss in work time and productivity while working) was estimated at between 20% and 29% (corresponding to more than one day of a five-day work week). In addition, labour-force participation among patients 18 years of age or older not in full-time education was low (<4%). In total, indirect costs accounted for between 26% and 43% of total cost of illness in the studied countries.

Per patient and annum, the total societal burden of DMD, comprising direct, indirect, and intangible costs, was estimated at $109,000 in Germany, $80,120 in Italy, $118,950 in the

UK, and $120,910 in the US. The corresponding national burden of DMD was estimated at $278,058,000, $154,465,000, $200,478,000, and $1,217,373,000, respectively.

We also found affected households to be carrying a considerable cost burden. The mean total annual cost of insurance premiums, co-payments for healthcare, and out-of-pocket payments for investments (e.g. adaptations of the home for wheelchair accessibility) was estimated at $5,940 in Germany, $7,550 in Italy, $3,490 in the UK, and $14,390 in the US. The corresponding total household cost burden, including income losses, costs associated with lost leisure time, and intangible costs, was estimated at $70,190, $58,440, $63,600, and $71,900, respectively. Adjusted regression analysis results showed that the total household cost burden on average was 38%, 181%, and 191% higher for late ambulatory, early non-ambulatory, and late non-ambulatory patients compared with their counterparts in the early-ambulatory stratum.

8.2 HEALTH-RELATED QUALITY OF LIFE IN DMD (PAPER II)

The vast majority (>84%) of patients were perceived as happy and in good health by their caregivers. In contrast, mean patient HUI-derived utility, proxy-assessed by the caregivers, was estimated at 0.46, ranging between 0.75 and 0.15 across ambulatory classes, 0.62 and 0 across caregivers' rating of their sons' current health, and 0.55 and 0.04 across caregivers' rating of their sons' current mental status (Figure 11). Mean caregiver proxy-assessed PedsQL NMM total scores were significantly associated with ambulatory class (mean score ranged between 52 and 72), caregivers' rating of their sons' current health (mean score ranged between 26 and 73), and caregivers' rating of their sons' current mental status (mean score ranged between 43 and 67).

We found patients' self-assessed PedsQL NMM total scores to be consistently higher than caregivers' proxy-assessments, but that overall agreement was good to excellent across the different instrument domains (the ICC was estimated at 0.78 for the total score in the pooled sample).

8.3 CAREGIVER BURDEN OF DMD (PAPER III)

Half of all caregivers reported that they were moderately or extremely anxious or depressed. Adjusted logistic regression results showed that the prevalence of anxiety and depression was significantly associated with the caregivers' rating of patients' health and mental status, as well as annual household cost burden and hours of leisure time devoted to informal care, but not ambulatory class (Table 11).

The mean EQ-5D-3L caregiver utility was estimated at 0.81. The sex- and age-matched loss in caregiver utility in relation to the general population was estimated at between 0.09 and 0.14 across ambulatory classes, 0.06 and 0.18 across caregivers' rating of their sons' current health, and 0.09 and 0.30 across caregivers' rating of their sons' current mental status. Mean VAS scores were lower than the estimated EQ-5D-3L utilities in all strata except for caregivers to patients rated to be very unhappy.

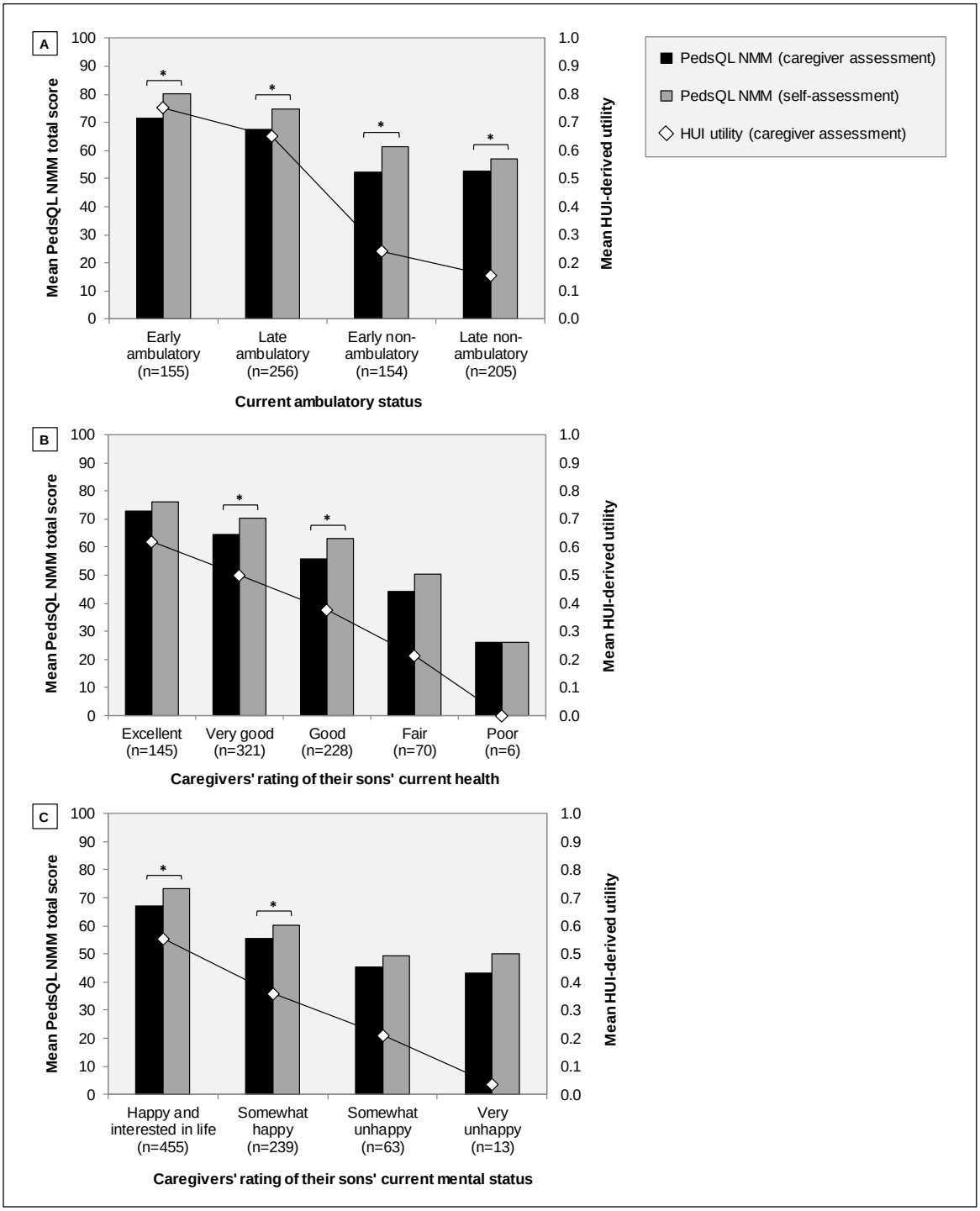


Figure 11: Self-assessed and caregiver proxy-assessed patient health-related quality of life, by ambulatory status (a), patients' current health (b), and patients current mental status (c)

*Note: *Statistically significant difference (at a 5% level) between patients' self-assessed Pediatric Quality of Life Inventory 3.0 Neuromuscular Module (PedsQL NMM) total scores and caregivers' proxy-assessed total scores. Health Utilities Index Questionnaire (HUI).*

Results from the SF-12 also indicated impaired caregiver mental health. Specifically, the mean SF-12 MCS score was estimated at 45, ranging between 44 and 46 across ambulatory classes, 48 and 37 across caregivers' rating of their sons' current heath, and 46 and 33 across caregivers' rating of their sons' current mental status.

The mean global ZBI score was estimated at 29, ranging between 25 and 32 across ambulatory classes, 23 and 38 across the caregivers' rating of their sons' current health, and 26 and 41 across the caregivers' rating of their sons' current mental status.

Table 11: Predictors of anxiety and depression in caregivers to patients with DMD

	n	Odds ratio (95% CI)[a]	P-value
Model I: Patients' ambulatory status			
Early ambulatory	155	1	
Late ambulatory	256	1.08 (0.70-1.65)	0.742
Early non-ambulatory	154	1.04 (0.64-1.70)	0.873
Late non-ambulatory	205	0.93 (0.53-1.64)	0.807
Model II: Caregivers' rating of patients' health			
Excellent	145	1	
Very good	321	1.53 (1.00-2.33)	0.049
Good	228	3.85 (2.40-6.20)	<0.001
Fair/poor	76	5.87 (3.05-11.29)	<0.001
Model III: Caregivers' rating of patients' mental status			
Happy and interested in life	455	1	
Somewhat happy	239	1.85 (1.32-2.58)	<0.001
Somewhat unhappy	63	4.67 (2.44-8.92)	<0.001
Very unhappy	13	7.22 (1.79-29.09)	0.005
Model IV: Annual household cost burden			
<$1000	380	1	
$1000-$5000	170	1.43 (0.95-2.16)	0.090
>$5000	220	1.76 (1.18-2.63)	0.006
Model V: Hours of leisure time devoted to informal care			
<25 per week	294	1	
25-50 per week	203	2.01 (1.37-2.94)	<0.001
>50 per week	273	3.35 (2.32-4.83)	<0.001

Note: Confidence interval (CI).

[a] Adjusted for country, caregiver sex, caregiver age, caregiver university degree, caregiver marital status, additional household member with DMD, household income class, patient diagnosis for depression, ADHD, ASD, and OCD, patient learning disabilities, patient glucocorticoid use, and patient-caregiver relationship (parent vs. other).

8.4 THE DMDSAT (PAPER IV)

The administered version of the new rating-scale (referred to as the DMDSAT) included a total of eight items in four domains (Figure 12). Upon re-scoring of two items that initially displayed disordered thresholds, the total DMDSAT score ranged from 0 (low functional ability) to 23 (high functional ability). Item fit to the Rasch model was good, mean item dependency low (-0.007), and agreement between model and clinical expert rankings of items difficulties in terms of functional ability excellent (Spearman's ρ=0.95). Fit of individual responses to the Rasch model was also good, with minor floor and ceiling effects (2% and 6%, respectively), and the estimated scale encompassed ability levels both lower and greater than those observed in the sample. The Person Separation Index (PSI) was estimated at 0.95 and Cronbach's α at 0.93, indicating good reliability. The adjusted mean change in per-patient annual cost of illness associated with a one-point increase in DMDSAT total score was estimated at 5.9%. The corresponding mean loss in patient utility was 9.5%. Spearman's ρ between predicted and observed total cost of illness and patient utility was estimated at 0.62 and 0.85, respectively.

The DMD Functional Ability Self-Assessment Tool (DMDSAT)

The questions below describe levels of activity for arm function, mobility, transfers, and need for ventilatory support. The activities are intended to be in order of difficulty and we would like you to **tick the circle that best applies to your current level of function.**

Question 1: Arm function	*Select one*
Can put an item such as book onto a shelf above shoulder height	◯
Can lift at least one arm above head	◯
Can lift at least one arm to shoulder height	◯
Can eat a meal without any help	◯
Needs help to cut up food but can feed and drink independently	◯
Needs help to drink or feed self	◯
Can pick objects up e.g. pen/money	◯
Can move fingers e.g. press on mobile or other electronic device	◯
Cannot move fingers	◯

Question 2: Mobility	*Select one*
Walks independently long distances outdoors (more than 1 km)	◯
Walks independently medium distances outdoors (less than 1 km)	◯
Walks independently outdoors for short distances, e.g. to car	◯
Walks outdoors with help from a person	◯
Walks indoors independently but requires wheelchair for outdoors	◯
Walks indoors with help from a person requires wheelchair outdoors	◯
Uses wheelchair indoors and outdoors	◯
Uses wheelchair but unable in some situations e.g. cold weather	◯
Unable to control wheelchair without help	◯

Question 3 to 7: Transfers	*Can do independently*	*Can do with help*	*Needs to be lifted or hoisted or cannot*
Get on and off the floor	◯	◯	◯
Get in and out of a chair	◯	◯	◯
Get in and out of bed	◯	◯	◯
Get on and off the toilet	◯	◯	◯
Go up and down stairs	◯	◯	◯

Question 8: Ventilatory support	*Not ventilated*	*Ventilated at night*	*Ventilated during day and night*
Ventilatory status	◯	◯	◯

Figure 12: The DMD Functional Ability Self-Assessment Tool (DMDSAT)

8.5 MODEL FRAMEWORK FOR ECONOMIC EVALUATION IN DMD (PAPER V)

In all models, starting the simulation at an age of five years, 50% of patients survived until an age of 25 years (in accordance with input data) and the estimated mean number of (undiscounted) life-years was 23. Results for the base-case (i.e. NICE reference case) and the societal perspective for each model are presented in Table 12.

Table 12: Cost-effectiveness of a hypothetical intervention in DMD

	Model I (DMDSAT) Annual treatment cost: £3,770			**Model II (ambulatory status) Annual treatment cost: £3,000**			**Model III (ventilation status) Annual treatment cost: £2,290**		
Cost outcomes (C)	**Trt**	**SoC**	**Δ**	**Trt**	**SoC**	**Δ**	**Trt**	**SoC**	**Δ**
Intervention cost	58,290	0	58,290	46,440	0	46,440	36,060	0	36,060
Direct medical costs	190,840	217,510	-26,670	221,250	244,120	-22,860	263,510	286,550	-23,040
Direct non-medical costs	184,330	201,290	-16,960	194,520	204,840	-10,310	204,860	207,440	-2,580
Patient indirect costs	69,000	69,000	0	69,000	69,000	0	69,000	69,000	0
Caregiver indirect costs	125,850	136,440	-10,590	139,490	145,560	-6,070	150,890	154,090	-3,200
Total, NHS/PSS perspective	249,130	217,510	31,620	267,690	244,120	23,580	299,570	286,550	13,020
Total, societal perspective	628,310	624,240	4,070	670,700	663,500	7,200	724,310	717,080	7,230
Effect outcomes (E)									
Patient QALYs	8.13	7.07	1.05	7.96	7.17	0.79	6.39	5.96	0.43
Caregiver QALYs	12.93	12.80	0.12	12.89	12.82	0.07	12.72	12.66	0.06
ICER (ΔC/ΔE) (£)									
NHS/PSS perspective			30,000			30,000			30,000
Societal perspective			3,450			8,400			14,636

Note: Cost results in 2015 Great British Pounds (£) rounded to nearest ten. Treatment (Trt). Standard of care (SoC). Incremental cost-effectiveness ratio (ICER). Quality-adjusted life-years (QALYs). Costs and effects were discounted at 3.5%.

The maximum annual treatment cost given a WTP for a QALY of £30,000 from the NHS/PSS perspective was estimated at £3,770 in model I, £3,000 in model II, and £2,290 in model III. Corresponding maximum treatment costs for the societal perspective was £5,790 in model I, £4,200 in model II, and £2,760 in model III. Results from the PSA showed that the probability of cost-effectiveness was <10% at drug costs exceeding £22,500 in model I, £17,500 in model II, and £15,000 in model III.

All three models were judged to have good validity with regards to the appropriateness of the conceptual representation of the disease, model input data, and model outcomes. Model Markov traces for the treatment arms (provided in the electronic supplementary Appendix) showed that the simulated cohorts transitioned across model states in agreement with the input data (i.e. modelled mortality and disease progression probabilities). Extreme value and unit testing revealed no errors with regards to the mathematical implementation of the models.

9 DISCUSSION

In accordance with the stated aims, this thesis provides a description of the previously unknown health economic context of DMD, including a portfolio of cost and utility data, a new tool designed to measure DMD disease severity, and a fully populated decision-analytic model framework for cost-effectiveness analysis. Importantly, as most of the long-term care in DMD is provided at home, the thesis work also comprise the health economic impact of the disease on affected family caregivers.

This section provides a summary discussion of the thesis results in relation to previous research, methodological and ethical considerations, implications for health policy, and suggestions for future research.

9.1 FINDINGS IN COMPARISON WITH PREVIOUS RESEARCH

9.1.1 DMD cost of illness (paper I)

Patients with DMD utilized a considerable variety of healthcare resources and services, associated with an annual direct medical cost of $11,240 in Germany, $7,300 in Italy, $15,940 in the UK, and $28,590 in the US. The mean per-patient annual direct cost of DMD was estimated at $42,360, $23,920, $54,160, and $54,270 for patients from Germany, Italy, the UK, and the US, approximately 10, 8, 16, and 7 times higher than the mean per-capita health expenditure in these countries, respectively. Prior to paper I, no study had estimated costs in a population comprising only patients with DMD. However, two studies, Ouyang et al. [104] and Larkindale et al. [105], had estimated costs for US patients with muscular dystrophies based on historical claims data. Although not directly comparable due to substantial differences with respect to disease and demographic characteristics, as well as study design (e.g. measured costs, included healthcare resource types, and valuation techniques), our estimate of the direct medical cost for the US sample was similar to equivalent estimates reported in these previous studies ($28,590 compared with $20,467 and $22,533, respectively).

9.1.2 Patient health-quality of life (paper II)

As a result of the severe disability and morbidity associated with DMD, we found caregiver proxy-assessed HUI-derived utility to be significantly impaired in relation to sex- and age-matched general population reference values. Specifically, across ambulatory classes, the mean utility ranged from 0.75 to 0.15, corresponding to a mean disutility of 0.20 to 0.79, well above the estimated clinically important difference threshold in HUI utility of between 0.03 and 0.05 [55]. No previous study has estimated HRQoL using the HUI in patients with DMD and only one study has assessed HRQoL in DMD using the EQ-5D-3L. Specifically, Pangalila et al. [115] estimated the mean EQ-5D-3L utility of a sample comprising 57 adult patients with ventilation support (mean age 27 years) at 0.44. Although our estimates are not directly comparable due to considerable differences between the HUI and the EQ-5D-3L (as

described in Section 2.6.1), our HUI-derived utility for patients requiring ventilation support (16%, 126 of 770) was notably lower at 0.10.

Interestingly, despite our low utility estimates, we found that the vast majority of patients were perceived as happy and in good health by their caregivers. As discussed in Section 2.6.3, a potential reason for this finding is a phenomenon known as "response shift" or "the well-being paradox", in which patients adjust their preferences and perception of health, happiness, and HRQoL over time, as the disease progresses. In addition, caregivers may adjust their perception of their sons' HRQoL based on disease history, as well as the anticipated disease trajectory (where the current health state may appear relatively good given the expected well-being in more advanced stages of DMD). Lastly, since DMD is a genetic condition, patients are not familiar with a life free from disease, and therefore do not have the same references as healthy individuals against which to compare their current situation.

Evidence of coping mechanism has been reported in previous research of patients with DMD. In two separate studies, Bray et al. [107,111] found patient HRQoL, measured using the PedsQL GCS and the Child Health Questionnaire, to be significantly negatively associated with physical functioning (measured using the Vignos scale), but that physical functioning was uncorrelated with psychosocial domains. Moreover, Uzark et al. [57] reported that self-assessed psychosocial HRQoL measured using the PedsQL GCS was higher among adolescent patients (13-18 years) than in younger boys (8-12 years) (although this pattern was not found for caregiver proxy-assessment of patient HRQoL, or for results from the PedsQL 3.0 DMD Module Scale). Similar findings were also reported in studies by Kohler et al. [113] and Elsenbruch et al. [114], in which older children and adolescents with DMD had better psychosocial HRQoL compared with their younger counterparts. However, it should be noted that most previous research of HRQoL in DMD have been conducted on small cohorts of patients with insufficient statistical power to perform stratified analyses by e.g. age or level of functional ability. In addition, a range of different instruments have been used to measure HRQoL, which also limits the possibility to compare existing research across studies and with our results.

In paper II, in addition to the HUI, we also measured patient HRQoL using the PedsQL NMM. We estimated the mean patient self-assessed PedsQL NMM total score at 67, ranging between 57 and 80 depending on patient ambulatory status. Only two previous studies [58,110] have to our knowledge assessed HRQoL in patients with DMD using the PedsQL NMM, and they estimated the mean parent-proxy score at 60 and 53, respectively. However, given that these preceding studies do not stratify their results by age or some other measure of disease stage or severity, it is not meaningful to conduct further comparisons. Furthermore, there are several methodological issues with the PedsQL NMM instrument, discussed in detail in Section 9.2.2, which limits interpretation of results across cohorts.

In our sample, patients' self-assessment of their HRQoL through the PedsQL NMM was consistently higher than the proxy-assessment made by their caregivers. Still, overall agreement between proxy- and self-assessments of patient HRQoL was found to be good to

excellent, in contrast to previous studies in DMD reporting poor to fair, poor to fair to moderate, moderate, and moderate to good agreement [57,58,107,109,110]. Potential reasons include differences between studies in, for example, patient-caregiver relationships (parents vs. more distant relatives) and extent of caregiver involvement in the daily life of the patient with DMD.

9.1.3 Caregiver burden of DMD (paper III)

We found caregivers to patients with DMD to have lower HRQoL compared with age- and sex-matched general population reference values. Specifically, the mean EQ-5D-3L disutility was estimated at 0.11, well above the minimally important difference threshold of 0.074 [151]. This finding suggests that the impact on HRQoL, as measured through the EQ-5D-3L, of being a caregiver to a patient with DMD is similar to or greater than very serious and sometimes rapidly fatal diseases, e.g. lung cancer and schizophrenia (0.11), systemic lupus erythematosus (0.08), and epilepsy (0.07) [152]. Only one study, Pangalila et al. [115], has measured HRQoL in caregivers to patients with DMD using a preference-based measure linked to utilities, and in contrast to our findings, the authors estimated the mean EQ-5D-3L utility at 0.87 (based on an undisclosed value set), similar to general population reference data. However, Pangalila et al. studied a cohort comprising 80 caregivers to adult DMD patients receiving ventilation support, and our estimates are thus not directly comparable. In our sample, the mean utility for caregivers to patients receiving ventilation support was 0.75, significantly lower than 0.88 ($p<0.001$).

Moreover, based on the results from the EQ-5D-3L, we found that half of all caregivers in our cohort were moderately or extremely anxious or depressed. Evidence of strain and stress in caregivers to patients with DMD have been reported previously, both in cohorts comprising only patients with DMD [112,117,118,121], as well as in mixed cohorts of patients with muscular dystrophies [119,120]. Interestingly, we found caregiver HRQoL, including anxiety and depression, to be predominantly associated with the well-being of the patient (as assessed by the caregiver), not disease stage. A similar observation was made by Baiardini et al. [112] in a study of 27 caregivers to patients with DMD, in which outcomes from the Family Strain Questionnaire were not associated with age, ventilator use, or wheelchair status. In addition, Nereo et al. [117] found that patient behaviour problems and verbal intelligence quotient were associated with maternal stress, but not age or wheelchair use, and the authors also reported that maternal stress related to child variables diminished over time. These findings suggest that caregivers to patients with DMD to some extent learn to cope with the increased level of dependency and requirements of assistance associated with the progression of the disease, although the tasks may still be considered burdensome.

We found caregiver anxiety and depression to be significantly associated with annual household cost burden and hours of leisure time devoted to informal care per week (estimated in paper I). In addition, based on the results from the ZBI, depending on their rating of patient well-being, between 12% and 40% of caregivers replied that they frequently or always felt that they should be doing more for their sons, 27% and 69% that they were stressed between

the demands of caring for the relative and trying to meet other responsibilities for family or work, and 17% and 62% that they did not have enough money to take care of their sons. These results emphasize that lack of time and money is an important source of distress in caregivers to patients with DMD. Similar findings have been reported by Kenneson et al. [120], who show that employment outside of the home is a predictor of stress in caregivers to boys with DMD and Becker muscular dystrophy. The stresses associated with meeting responsibilities associated with informal care at home is further underscored by the fact that a non-trivial proportion of caregivers in our sample had terminated their employment, either partially or in full (as reported in paper I). Similar observations were also reported by Baiardini et al. [112].

We estimated the mean global ZBI score at 29, ranging from 23 to 42 across investigated strata. No previous study has estimated caregiver burden using the ZBI in DMD, but Kenneson et al. [120] administered a four-item screening version of the ZBI to a sample of caregivers to patients with DMD and Becker muscular dystrophy. In their study, Kenneson et al. found that more than half of all caregivers reported a high level of caregiving demands. Comparing our estimates with findings from studies of other diseases, the burden as measured through the ZBI was comparable to or higher than estimates for e.g. neuromuscular diseases in general (23) [119], irritable bowel syndrome (22) [153], Alzheimer's disease (29) [154], OCD (29) [155], and Parkinson's disease (24) [156]. However, it should be noted that in the case of DMD, the burden is carried by affected families for several decades, in contrast to e.g. geriatric diseases such as Alzheimer's, which usually develop late in life. In addition, as further discussed in Section 9.2.3, it is challenging to interpret the outcome of the ZBI as there is no link between ZBI scores and the trait measured (i.e. subjective burden).

9.1.4 The DMDSAT (paper IV)

Prior to the development of the DMDSAT, no instrument was designed to measure severity in DMD along the entire disease trajectory. Instead, several different tools were used to assess and monitor functional ability in clinical practice and measure efficacy in trials depending on the ambulatory status of the patients (e.g. the NSAA [130] for ambulatory patients and PUL [131] for non-ambulatory patients). The DMDSAT complements these more complex tools, which involves physical tasks and/or timed tests that must be supervised by a healthcare practitioner, and is also the first rating-scale linked to costs and utilities.

Comparing results from the Rasch analysis of the DMDSAT with the psychometric properties of the NSAA and the PUL (both of which have been validated using the same methodology [131,157]), all three instruments were found to have excellent targeting and reliability (PSI: 0.95, 0.91, and 0.96, respectively), stable item locations, and good item and person fit to the Rasch model. For the DMDSAT and NSAA, item difficulty was in good agreement with clinical opinion (Spearman's ρ: 0.95 and 0.80, respectively), indicating good clinical validity (data not reported for the PUL).

9.1.5 Economic evaluation in DMD (paper V)

As noted in Section 3.5, no model or economic evaluation in DMD had been published prior to paper V. However, one model of the disease was made available in April 2016 (before paper V was submitted for publication) as part of a NICE evaluation of ataluren [158], a new treatment targeting a subgroup of patients with DMD caused by nonsense mutation in the dystrophin gene. This model, which was based on cost and utility data from this thesis, comprised a total of five states (in addition to an absorbing state for dead) defined in terms of patient ambulatory status, ventilation status, and scoliosis: (i) ambulatory, (ii) non-ambulatory, (iii) non-ambulatory with scoliosis, (iv) non-ambulatory with ventilation support, and (v) non-ambulatory with ventilation support and scoliosis. Several comments can be made with regards to the ataluren evaluation in relation to the findings presented in this thesis. First, in the ataluren model, differences between arms (i.e. ataluren versus placebo) related only to time to non-ambulation based on outcomes from the 6MWT, the primary trial endpoint, with identical transition probabilities for subsequent states. As a result, the mean costs associated with scoliosis and ventilation support could have been added to the non-ambulatory state (accounting for assumed differences between arms), and the model therefore essentially only comprised two states; ambulatory and non-ambulatory (in addition to dead). Accordingly, the ataluren model is indeed a less granular version of model II (which was based on four ambulatory classes) as specified in paper V.

Second, as noted above, the ataluren model was partially based on health economic data from this thesis. Specifically, our cost and utility estimates for late ambulatory patients were used for the ambulatory model state and our cost and utility estimates for late non-ambulatory patients were applied to the non-ambulatory state. In addition, in the ataluren model, scoliosis, but not ventilation support, was assumed to be associated with an additional cost and patient utility loss. However, no adjustments were made to the costs and utilities that were assigned to the non-scoliosis non-ventilation support ambulatory and non-ambulatory model states, despite the fact that these estimates were derived for a sample of patients of which a substantial proportion in fact had scoliosis and received ventilation support. Thus, our cost and utility data appears to have been incorrectly implemented into the structure of the ataluren model.

Third, in the ataluren evaluation, assumptions regarding mortality seem to be in poor agreement with the current body of evidence (reviewed in Section 2.1.2). Specifically, from the Markov traces it appears as if patients receiving placebo had a median survival of about 33 years, with a substantial proportion (30% of patients) surviving beyond 40 years, as well as fifty years of age (15%). Moreover, although not explained in detail the evaluation report, it appears as if ataluren was assumed to be associated with a substantial reduction in mortality, with a mean survival in the ataluren arm of about 42 years, with a non-trivial proportion of patient surviving beyond 50 years (35%) and even 60 years (20%) of age. Furthermore, in contrast to the findings in paper V, efficacy on mortality was associated with improved cost-effectiveness in the analysis of ataluren (possibly due to differences between arms in terms of costs and utilities for the most advanced model states). Specifically, in the

ataluren arm, patients resided in the non-ambulatory ventilation support state from a median age of 30 years until a median age at death of 42 years, with only a trivial proportion developing scoliosis. Patients treated with placebo, on the other hand, transitioned from the ventilation support state to the scoliosis and scoliosis and ventilation support states, with a median age of scoliosis of 17 years and a median age of scoliosis and ventilation support of 32 years. These differences in transition probabilities and costs and utilities between the ventilation support and scoliosis states may help explain our conflicting results relating to efficacy on survival.

Fourth, comparing Markov traces for the two arms in the ataluren model, it appears as if patients receiving placebo were at risk of transitioning from the ambulatory to the non-ambulatory state from 9 years of age (at a probability so that the mean age at loss of ambulation was in agreement with existing evidence, i.e. 14 years). In the ataluren arm, however, patients were not at risk of becoming non-ambulatory until they became 14 years of age, with a median age at loss of ambulation of 25 years. This inconsistency in implementation would be expected to have resulted in an overestimation of the mean time to non-ambulation for patients treated with ataluren and thus an overestimation of the drug's cost-effectiveness.

9.2 METHODOLOGICAL CONSIDERATIONS

9.2.1 Estimating costs of DMD

There are several methodological considerations concerning the methods underlying the cost of illness approach applied in this thesis work, of which five are discussed below. First, decisions regarding which costs that are to be included in a cost of illness study have non-trivial consequences for the results. For example, a cost study in DMD that only includes resource use related to physician visits would underestimate the direct medical cost of the disease, given that visits to other healthcare practitioners (e.g. physiotherapists) also are very common in this patient group. Similarly, a cost study that only covers direct medical costs of DMD is likely to underestimate the total societal cost burden of the disease, given that a substantial proportion of total care consists of informal care provided at home. In paper I, given that our aim was to estimate costs from a societal perspective, as well as from the perspective of affected households, we sought to include all resources and costs that could be attributed DMD, both directly and indirectly. However, some items may have been omitted from our analysis, e.g. resource use related to end-of-life care, and the total economic burden of DMD may as a consequence have been underestimated.

Second, as described in Section 2.5, estimating costs of an illness comprise several steps, including identification of resources consumed as a consequence of the disease, measuring the quantities of identified resources, and obtaining price data for each resource type. Thus, the price component, which comprises the opportunity cost of all identified resources, is central to the analysis. Indeed, high-quality data on healthcare resource use are worth little if multiplied with uncertain, low-quality price estimates. However, for many resources,

opportunity costs are seldom known. In cost of illness studies, it is therefore common practice to rely on national reference lists that contain prices which reflect the monetary reimbursement value, or the prices charged between different hospitals. Accordingly, these prices may or may not reflect the true opportunity cost, and differences between countries may be substantial depending on e.g. national healthcare systems, reimbursement policies, etc. It is therefore important to keep in mind that inter-country differences in cost of illness estimates may be driven both by differences in resource use and prices (which was the case with e.g. our direct medical cost estimate for Italy, as reported in in paper I).

Third, given its prevalence and magnitude in genetic childhood diseases such as DMD, methods concerning the valuation of informal care would also be expected to have a considerable impact on total cost of illness estimates. As described in Section 7.1, in paper I, we employed a method which explicitly differentiated between the time spent providing informal care instead of working (accounted as a productivity cost) and the number of hours of leisure time devoted to informal care (accounted as an informal care cost). We based our calculation of unpaid informal care costs on the country-specific mean daily number of hours of leisure time for an adult in the general population, instead of assuming that caregivers who reduced their working hours or stopped working completed due to their sons' disease spent all of their time providing informal care. We chose this approach because the true number of hours of informal care was unknown, to avoid double-counting costs of care provided via nurse visits to the home, personal assistants, etc., and to allow for an unadjusted estimation of indirect costs in accordance with the human capital approach (which is arguably a more established method compared with valuation techniques of informal are). As a consequence, we may have underestimated informal care costs for caregivers who provided informal care instead of working (hours which instead were accounted for in full as a productivity cost) and it is therefore likely that we to some degree underestimated informal care costs and overestimated productivity costs. Moreover, in accordance with previous research and recently updated estimates of the value of travel time savings, we valued each hour of leisure time at 35% of the country-specific national mean gross wage. Alternatively, we could have utilized the price for e.g. a nurse or physician visit, but we opted for a more conservative approach, primarily because it is unlikely that all hours of informal care encompass tasks that require medical training (e.g. feeding, dressing, getting in and out of bed, etc.). It is also worth pointing out the fact that we included informal care as a direct non-medical cost with the rationale that the care otherwise would have had to be provided by healthcare professionals. In fact, given the healthcare needs and assistance with activities of daily living associated with DMD, our results suggest that many patients with DMD, in the absence of informal care, would have to be institutionalized, especially in more advanced stages of the disease.

Fourth, when modelling cost data, ordinary least square regression models may not be appropriate due to e.g. skewness of the data, heteroscedasticity, censoring, and many observations with zero-costs [159]. Instead, other models have been proposed in the literature, and in case of no censoring and no zero-costs, which is an accurate description of

our DMD cost data, a vast body of literature favours using a GLM model assuming a gamma distribution and a log-link function [160], which was applied in this thesis work. Another important consideration when modelling data for discrete states in a decision-analytic model concerns the choice of explanatory variables in addition to the main covariates. For example, in a regression model of disease stage and total cost of illness, scoliosis (which is more common in more advanced disease stages) is a mediator (i.e. it carries some of the influence of disease stage on cost of illness), in contrast to e.g. OCD which is not associated with disease stage but may have an impact on cost of illness. Consequently, to obtain relevant relative coefficients for disease stage, it is appropriate to control for OCD, but not scoliosis.

Fifth, when reporting costs for a progressive disease such as DMD, it is essential to stratify estimates by some measure of disease progression. Otherwise, the mean estimate will be a direct function of the distribution of disease severity in the sample. This applies to other study outcomes as well, e.g. HRQoL.

9.2.2 Measuring health-related quality of life

There are several methodological limitations concerning rating-scales used to measure HRQoL in paediatric and adult populations. These limitations, of which four are discussed below, have important implications for the interpretation of our results. First, as described in Section 2.11.3, ordinal rating-scales fail to adhere to the basic principles of sound measurement. In the case of the PedsQL NMM, for example, a child indicating that he "always" has problem explaining his disease to other people is attributed the same loss in HRQoL as a child indicating the he "always" feels tired, although the impact on HRQoL may be very different (and the difference may also vary across samples). For this reason it is not meaningful to directly compare scores, or changes in scores, from the PedsQL NMM between studies.

Second, as noted in Section 2.6.4, value sets for measures such as the EQ-5D-3L and the HUI are based on preference data estimated many years ago from specific populations. For example, utilities for the HUI Mark 3, which describe a total of 18,000 health states, are based on valuation of only 25 health states by 256 randomly selected members of the general population in Hamilton, Ontario, Canada in 1994 [71]. Thus, when administering the HUI to an individual, the estimated health state utility represents the average predicted utility of that particular health state as valued by this sample of men and women more than two decades ago. Consequently, given that preferences for health states would be expected to be influenced by cultural and social aspects, among categories of other factors that may vary by time and geographical setting, the HUI value set may not be fully relevant for all caregiver in our study population.

Third, in this thesis, the estimated patient HUI-derived utilities were compared with age- and sex-matched general population reference data for Canada (based on the HUI Mark 3), because HUI reference utilities does not exist for Germany, Italy, or the UK, and since US data only comprises ages 18 years, and older. Thus, given that preferences for health states

may differ across countries, our estimated patient disutilities may be subject to some bias. Moreover, regarding caregiver utility as measured using the EQ-5D-3L, estimates for our sample were compared against UK reference data derived using the TTO value set (i.e. the same value set as we applied in the thesis work), but for comparison and reference, we also included alternative utility estimates based on country-specific value sets for the EQ-5D-3L. It should be noted that we chose to compare our utility estimates against a single set of reference data for the HUI and EQ-5D-3L, respectively, also because it allows for an assessment of the absolute impact of DMD on HRQoL excluding any country-specific differences. In other words, instead of estimating country-specific disutilities, which also capture differences in general population HRQoL, we derived disutility estimates from a single comparative continuum.

Fourth, regarding the HUI Mark 3, as described in Section 2.6.4, the value set (i.e. mean predicted utilities for each health state) were derived from a sample comprising individuals 16 years of age, or older. Thus, although the HUI may be appropriate for proxy-administration in paediatric populations, the accompanying value set does not comprise preferences of individuals below the age of 16 years. In this sense, the HUI should not be regarded a measure of HRQoL in children and adolescents.

9.2.3 Measuring caregiver burden

Similarly to HRQoL, caregiver burden is a multidimensional and complex construct, and although a definition has been proposed, as noted in Section 2.9, there is no consensus how to quantify the trait. Indeed, different studies focus on different aspects, for example "subjective" components (e.g. mental health and stress or strain) or "objective" factors (e.g. number of hours of informal caregiving or household cost burden). In addition, many rating-scales used to measure aspects of caregiver burden were developed using CTT and their scores are thus not easily interpreted. The ZBI, for example, map out a continuum ranging from 0="low burden" to 88="high burden", but it is not clear what "low" and "high" means in this context, or what values in between 0 and 88 represents. For these reasons, it is seldom straightforward to compare estimates of the caregiver burden across studies. This may also apply to specific aspects of the burden, e.g. prevalence of depression, due to differences in diagnostic criteria and data sources used in the analysis (i.e. self-reported versus inpatient register data).

9.2.4 Measuring progression in DMD

Developing a single scale that measures progression across the entire lifetime of disease in DMD is challenging due to considerable heterogeneity in the clinical manifestations of the disease. For example, younger patients usually experience problems with their lower limbs, whereas more pronounced upper limb weakness generally occurs in later stage of the disease. Therefore, in our development of the DMDSAT, to capture the full range of manifestations without including irrelevant questions (e.g. asking a fully ambulant boy if he can raise his arms above his head), we first created two questions with hierarchical response options

concerning arm function and mobility. To add further granularity to the measure, we then included five questions relating to transfers, capturing activities of daily living, and one question concerning ventilatory support. Evident from the results from the Rasch analysis presented in paper V, the included questions together marked an extensive range of functional ability on the estimated continuum, from "Cannot move fingers" followed by "Ventilated during day and night" to "Walks independently long distances outdoors (more than 1 km)" and "Can independently go up and down stairs". Moreover, only five of 23 item thresholds marked the same position on the continuum in terms of difficulty, and the level of dependency among questions was very low (−0.007 on average). These data, together with the high association between model and expert rankings of item thresholds, suggest that the DMDSAT successfully operationalise functional ability in DMD across the disease trajectory.

Comparing different measures of disease stage, an important strength with the interval DMDSAT scale is that it can be included in a regression model as a continuous variable instead of several dummy variables, which saves degrees of freedom, of particular importance when modelling rare disease data which often is based on small samples.

9.2.5 Modelling DMD

In paper V, we synthesized our previously published health economic data in three Markov model frameworks of varying granularity comprising states representing stages of disease, each attributed a total cost and utility estimate. An alternative approach would have been to model specific disease-related complications, e.g. scoliosis and cardiomyopathy, and include the health economic impact of these events separately. However, there are several limitations which such an approach. First, given the wide variety of disease-related complications associated with DMD across the progression of the disease, it would have been challenging to include all relevant manifestations without making models that are very complex. Second, cost and utility data, as well as risks, for many complications of DMD are unknown, in particular the combined impact and risk of several concurrent and/or subsequent manifestations. Third, in a given cycle, patients would only be at risk of a limited number of complications, and risks for minor complications may as a result compete with risks for more serious events, a phenomenon sometimes referred to as "competing risks". Fifth, many therapeutic strategies currently being explored for DMD aim to slow down the rate of disease progression (i.e. delay the decline in overall muscle strength and loss of functional ability) and it may therefore be more relevant to model stages of disease, not distinct complications.

9.2.6 Collecting DMD data

As described in Section 6, paper I to IV in this thesis were based on observational data collected using a questionnaire (and this data also served as input to paper V). Alternatively, study data could have been extracted from e.g. registers or patient charts. However, as noted in Section 2.5, in the case of DMD, it is generally not possible to utilize registers, claims databases, or other administrative sources since DMD does not have a unique ICD-9 or ICD-

10 classification code. Moreover, some of the collected data would usually not be available in registers nor patient charts, e.g. caregiver work status, information about informal care, and HRQoL data. Thus, although self-reported data are associated with some limitations (discussed below), to meet the stated thesis objectives, collecting new data was necessary.

9.2.7 General strengths and limitations

The main strengths of this thesis work include:

- A comparatively large international sample of individuals with DMD, allowing stratification of result by e.g. stage of disease;
- A detailed cost of illness analysis comprising costs associated with formal healthcare, informal care, productivity losses, as well as costs carried by affected households;
- A thorough investigation of patient and caregiver HRQoL and burden, comprising both generic and disease-specific instruments, as well as outcomes in terms of utilities suitable for economic evaluations;
- A comprehensive psychometric analysis of the DMDSAT using Rasch analysis, a method deemed superior to traditional psychometric methods; and
- A fully populated model framework for economic evaluation in DMD, as well as comprehensive sensitivity analysis to help inform HTA and future health economic programmes of treatments for DMD.

The main limitation of this thesis work concerns external validity. Patients were recruited through the TREAT-NMD network at a mean response rate of 42% (33% if only counting complete responses), and as participation in the registries is voluntary and family-initiated, we cannot rule out a degree of selection bias. Implications of this potential systematic error on our estimates depend on the characteristics of those who chose to participate compared with those who did not. For example, under-representation of severely ill patients may imply that costs were under-estimated and patient HRQoL over-estimated, with unclear impact on caregiver HRQoL and burden. Over-representation of particularly motivated patients and caregivers may instead imply that costs were over-estimated and the impact on caregiver HRQoL and burden over-estimated, with unclear impact on patient HRQoL. The latter may also be true in the case of an under-representation of financially deprived participants. Due to lack of data, we were unable to conduct a full analysis of non-response bias, but found that the distribution of age was similar among responders and non-responders, and that the collected clinical and epidemiological data were characteristic for the different ambulatory classes in our sample. In addition, although enrolment via register infrastructure may not be ideal, it should be noted that currently there are no alternative pathways available to recruit large number of patients with DMD from different countries to research. It is also worth pointing out that the response rate among those who actually received a study invitation in the thesis study would be expected to be higher as a result of e.g. lost invitations due to recent changes to email addresses and spam filters.

An additional important limitation concerns information bias, a systematic error which arises from measurement error [161]. The primary source of information bias in this thesis was misclassification due to the self-reported nature of our cross-sectional data. Specifically, although all individuals included in the study had a confirmed diagnosis of DMD, other important variables (e.g. ambulatory status) may have been misclassified as a result of e.g. recall bias or incorrect reporting. This would also be true for the study outcomes, e.g. DMD-related healthcare resource use and the prevalence of anxiety and depression. Moreover, for some variables, the misclassification may have been differential (i.e. dependent on the values of other variables), as opposed to non-differential (i.e. not dependent on the values of other variables), which would also impact comparisons of outcomes across strata, e.g. stages of disease. We tried to alleviate the problem of recall bias by specifying recall periods for resources in accordance with standard DMD care and by conducting a pilot study with detailed feedback questions to further improve the validity of the responses. To minimize incorrect reporting, we included help texts (explaining e.g. specific resource types and other terminology) for all questions in the questionnaire, as well as logical tests and skip patterns to ensure that the collected data was accurate and complete.

A third study limitation concerns confounding. A confounding variable, Z, is a variable that is: (i) associated with the outcome, Y (as a cause but not as an effect), and (ii) associated with the exposure, X (as a cause but not as an effect) [161]. Confounding results in "confusion of effects" [161], since it is not possible to obtain a meaningful measure of the association between X and Y without controlling for the cofounding variable Z. This limitation primarily concerns paper III, in which we modelled the prevalence of anxiety and depression as a function of patients' ambulatory status, patients' health and mental status as perceived by the caregivers, and two measures of the objective caregiver burden, respectively. Our main approach to manage confounding was by adjusting the regression models for potential confounders (e.g. caregiver sex, age, university degree, marital status, and household income class). However, it is worth pointing out that our findings may still be biased due to unmeasured confounding. In addition, regarding the interpretation of our fitted regression models, due to the observational nature of our data, we were unable to draw conclusions regarding causality.

A fourth limitation concerns the size of our sample. Research of rare diseases generally involve analysing small samples and is therefore often associated with low precision due to random error. This is a particular problem for rare progressive diseases, such as DMD, where outcomes must be stratified by disease stage to allow for a meaningful interpretation. For this reason, although our pooled sample comprised a total of 770 patient-caregiver pairs (a cohort referred to as "laudably large" in a recent commentary [162] on paper II), precision of some point estimates for some strata was relatively low due to the small number of patients or caregivers.

9.3 ETHICAL CONSIDERATIONS

A study of a rare childhood disease such as DMD involves several important ethical considerations. For example, if the research comprises patients along the entire disease trajectory, which was the case in the study reported in paper I to IV, many participants will be children or adolescents and may thus be too young to fully appreciate benefits and risks associated with participating in a study. In addition, a considerable proportion of patients with DMD have some degree of cognitive impairment, and many also suffer from ASD or OCD, which may further compromise their understanding of the involved study tasks. To ensure that all participants were adequately informed and to safeguard the confidentiality and integrity of the study participants, the following components were implemented into the design of the thesis study in accordance with national IRB regulations.

- Patient contact (e.g. identification and recruitment) was managed exclusively by the national TREAT-NMD partners. Each eligible patient was assigned a unique identification number which was used to identify the individual throughout the study. Each patient was also allocated a unique study website log-in.
- Patients and caregivers were asked to review Participant Information Sheets (PISs) with information about the study, benefits and risks of participating, etc. Patient PISs were formatted for different age-groups to ensure understandability. In some jurisdictions, depending on the age of the patient, the legal guardian of the patient had to certify that his/her child had understood the content of the PIS.
- Informed Consent Forms (ICFs) were completed by all participating caregivers and patients. In some jurisdictions, depending on the age of the patient, the legal guardian of the patient could provide informed consent on behalf of his/her child.
- The encrypted study website was developed and hosted by a third-party vendor specialising in research data collection infrastructure which conform to relevant regulatory standards, guidelines, and recommendations.
- Once data collection had been completed, the anonymous study data were encrypted and transferred to a secure database for analysis. Access to the study data was restricted to staff directly involved in the study.

9.4 SUMMARY OF IMPLICATIONS FOR HEALTH POLICY

The findings from the papers in this thesis may have several implications for health policy, summarized in the bulleted list below.

- Our portfolio of cost and HRQoL data from paper I, II, and III should be helpful to inform HTAs of interventions in DMD and facilitate cost-effectiveness analyses and other evaluations of new health technologies.
- Our estimates of the previously unknown household cost burden in DMD may help inform support schemes directed towards affected families, of particular importance given the association between objective measures of the caregiver burden (i.e.

household cost burden and hours of informal care) and caregiver HRQoL as identified in paper III.

- Our estimates of informal care from paper I highlight the importance of family caregiving as a complement to and substitute for formal care in a childhood disease such as DMD.
- Our estimates of the subjective caregiver burden in DMD from paper III suggest that it may be relevant to screen for depression in caregivers to patients with DMD and that patients' health and mental status may be helpful predictors of caregiver distress.
- As the first rating-scale designed to measure functional ability across the entire lifetime of disease in DMD, the DMDSAT should be a helpful complement to existing tools used to measure efficacy in clinical trials, monitor progression in clinical practice, and model the disease in economic evaluations.
- Our model frameworks from paper V should be helpful to inform HTAs of interventions in DMD. The results from our model simulations also discern the impact of different model structures, analysis perspectives, and drug efficacy and cost data profiles, amongst other factors, on estimated cost-effectiveness and may thereby help inform the design of health economic programmes of DMD drugs currently in development.

9.5 SUGGESTIONS FOR FUTURE RESEARCH

Based on this thesis work, the following topics for future research was identified.

- Experience-based utilities in DMD to inform HTAs in jurisdictions where these type of data are preferred (in e.g. Sweden).
- Costs associated with end-of-life care in DMD, which may not have been fully accounted for in this thesis work.
- Further validation of the DMDSAT to assess properties such as responsiveness, test-retest reliability, and minimally important difference thresholds.
- Patient preferences for disease complications and treatment benefits and side-effects (through e.g. a conjoint analysis) to help inform therapeutic decision-making and drug development.
- Survival in DMD for patients who are receiving care in accordance with current clinical care guidelines.
- The relationship between arm and leg function in DMD to further inform validation of clinical rating-scales measuring functional ability and disease severity.
- Modelling the cost-effectiveness of future treatments for DMD.

10 CONCLUSIONS

In paper I, we show that DMD is associated with a substantial economic burden to society and affected households. Across countries, the mean per-patient annual direct cost of illness was found to be 7 to 16 times higher than the mean per-capita health expenditure. Indirect and informal care costs were substantial, each component constituting between 18% and 43% of total costs.

In paper II, we show that HRQoL in DMD, measured through public preferences, is substantially impaired in relation to general population reference values, strongly negatively associated with disease progression, and in good agreement with caregivers' rating of patients' current health and mental status. Still, most children and young adults with DMD were perceived as happy and in good health by their caregivers, indicating that some domains of HRQoL remain intact through the progression of DMD.

In paper III, we show that caring for a person with DMD can be associated with a substantial burden and impaired HRQoL. Our findings also suggest that caregivers to patients with DMD should be screened for depression and emphasize the need for a holistic approach to family mental health in the context of chronic childhood disease.

In paper IV, we present a new rating-scale, the DMDSAT, developed to measure functional ability in patients with DMD. Results from the psychometric analysis show that the DMDSAT is an instrument fit for purpose to measure progression across the entire disease trajectory.

In paper V, we found that model structure and perspective of analysis have a substantial impact on assessments of cost-effectiveness of treatments for DMD. Our results show that the DMDSAT represents a sensitive and clinically relevant option for modelling DMD across the entire trajectory of disease in economic evaluations compared with frameworks based on conventional staging of disease progression.

11 ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Associate Professor Peter Lindgren, Professor Katharine Bushby, and Professor Lars Alfredsson. It has been an honour and a great learning experience to work with you. I am very grateful.

I also want to thank Professor Hanns Lochmüller, Professor Volker Straub, Dr. Michela Guglieri, Dr. Anna Mayhew, and Dr. Michelle Eagle for their contributions to the thesis work and sharing of their vast experience and clinical knowledge. I hope we can continue our collaborations both in the near and distant future.

I am also grateful for the study funding from GlaxoSmithKline.

I want to thank Dr. Mattias Ekman for being my mentor during my thesis studies and for proof-reading an earlier version of the thesis summary chapter.

I want to thank my colleagues (both past and present) at i3 Innovus/Optum/Mapi. A special thanks to Oskar Ström and Mia Gannedahl Franco.

I want to express my deepest gratitude to the patients with DMD and their caregivers who participated in this research.

Finally, I want to thank my fiancée Josefin for her endless encouragement and support during the past four years. Home is wherever I am with you.

12 REFERENCES

1 Aymé S, Rodwell C. 2014 Report on the State of the Art of Rare Disease Activities in Europe. The European Union Committee of Experts on Rare Diseases. Available at: http://www.eucerd.eu (accessed January 15, 2016).

2 National Institutes of Health. The Genetic and Rare Diseases (GARD) Information Center. Available at: https://rarediseases.info.nih.gov (accessed January 15, 2016).

3 Schieppati A, Henter JI, Daina E, et al. Why rare diseases are an important medical and social issue. *Lancet* 2008; 371(9629): 2039–2041.

4 Kumar Kakkar A, Dahiya N. The evolving drug development landscape: from blockbusters to niche busters in the orphan drug space. *Drug Dev Res* 2014; 75(4): 231–234.

5 Franco P. Orphan drugs: the regulatory environment. *Drug Discov Today* 2013; 18(3-4): 163–172.

6 Gammie T, Lu CY, Babar ZU. Access to Orphan Drugs: A Comprehensive Review of Legislations, Regulations and Policies in 35 Countries. *PLoS One* 2015; 10(10): e0140002.

7 Drummond MF, Wilson DA, Kanavos P, et al. Assessing the economic challenges posed by orphan drugs. *Int J Technol Assess Health Care* 2007; 23: 36–42.

8 Simoens S, Picavet E, Dooms M, et al. Cost-Effectiveness Assessment of Orphan Drugs. *Appl Health Econ Health Policy* 2013; 11: 1–3.

9 Hughes-Wilson W, Palma A, Schuurman A, Simoens S. Paying for the Orphan Drug System: break or bend? Is it time for a new evaluation system for payers in Europe to take account of new rare disease treatments? *Orphanet J Rare Dis* 2012; 7: 74.

10 Drummond M, Towse A. Orphan drugs policies: a suitable case for treatment. *Eur J Health Econ* 2014; 15(4): 335–340.

11 Bushby K, Finkel R, Birnkrant DJ, et al. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and pharmacological and psychosocial management. *Lancet Neurol* 2010; 9(1): 77–93.

12 Eagle M, Bourke J, Bullock R, et al. Managing Duchenne muscular dystrophy–the additive effect of spinal surgery and home nocturnal ventilation in improving survival. *Neuromuscul Disord* 2007; 17: 470–475.

13 Emery AE. The muscular dystrophies. *Lancet* 2002; 359: 687–695.

14 Tyler KL. Origins and early descriptions of "Duchenne muscular dystrophy". *Muscle Nerve* 2003; 28(4): 402–422.

15 Mercuri E, Muntoni F. Muscular dystrophies. *Lancet* 2013; 381: 845–860.

16 Koenig M, Beggs AH, Moyer M, et al. The molecular basis for Duchenne versus Becker muscular dystrophy: correlation of severity with type of deletion. *Am J Hum Genet* 1989; 45(4): 498–506.

17 Den Dunnen JT, Grootscholten PM, Bakker E, et al. Topography of the Duchenne muscular dystrophy (DMD) gene: FIGE and cDNA analysis of 194 cases reveals 115 deletions and 13 duplications. *Am J Hum Genet* 1989; 45(6): 835–847.

18 Mendell JR, Shilling C, Leslie ND, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. *Ann Neurol* 2012; 71: 304–313.

19 Moat SJ, Bradley DM, Salmon R, Clarke A, Hartley L. Newborn bloodspot screening for Duchenne muscular dystrophy: 21 years experience in Wales (UK). *Eur J Hum Genet* 2013; 21(10): 1049–1053.

20 Mah JK, Korngut L, Dykeman J, Day L, Pringsheim T, Jette N. A systematic review and meta-analysis on the epidemiology of Duchenne and Becker muscular dystrophy. *Neuromuscul Disord* 2014; 24(6): 482–491.

21 Norwood FL, Harling C, Chinnery PF, Eagle M, Bushby K, Straub V. Prevalence of genetic muscle disease in northern England: in-depth analysis of a muscle clinic population. *Brain* 2009; 132: 3175–3186.

22 Romitti PA, Zhu Y, Puzhankara S, et al. Prevalence of Duchenne and Becker muscular dystrophies in the United States. *Pediatrics* 2015; 135(3): 513–521.

23 Ballo R, Viljoen D, Beighton P. Duchenne and Becker muscular dystrophy prevalence in South Africa and molecular findings in 128 persons affected. *S Afr Med J* 1994; 84: 494–497.

24 Nakagawa M, Nakahara K, Yoshidome H, et al. Epidemiology of progressive muscular dystrophy in Okinawa, Japan. Classification with molecular biological techniques. *Neuroepidemiology* 1991; 10: 185–191.

25 Darin N, Tulinius M. Neuromuscular disorders in childhood: a descriptive epidemiological study from western Sweden. *Neuromuscul Disord* 2000; 10: 1–9.

26 Chung B, Wong V, Ip P. Prevalence of neuromuscular diseases in Chinese children: a study in southern China. *J Child Neurol* 2003; 18: 217–219.

27 Talkop UA, Kahre T, Napa A, et al. A descriptive epidemiological study of Duchenne muscular dystrophy in childhood in Estonia. *Eur J Paediatr Neurol* 2003; 7: 221–226.

28 Toussaint M, Steens M, Wasteels G, Soudon P. Diurnal ventilation via mouthpiece: survival in end-stage Duchenne patients. *Eur Respir J* 2006; 28(3): 549–455.

29 Yasuma F, Sakai M, Matsuoka Y. Effects of noninvasive ventilation on survival in patients with Duchenne's muscular dystrophy. *Chest* 1996; 109(2): 590.

30 Bushby K, Finkel R, Birnkrant DJ, et al. Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care. *Lancet Neurol* 2010; 9(2): 177–189.

31 Malik V, Rodino-Klapac LR, Mendell JR. Emerging drugs for Duchenne muscular dystrophy. *Expert Opin Emerg Drugs* 2012; 17(2): 261–277.

32 Kole R, Krieg AM. Exon skipping therapy for Duchenne muscular dystrophy. *Adv Drug Deliv Rev* 2015; 87: 104–107.

33 Landfeldt E, Lindgren P, Bell C, et al. Compliance to care guidelines for Duchenne muscular dystrophy. *J Neuromuscul Dis* 2015; 2: 63–72.

34 Drummond M, Sculpher M, Torrance G, et al. Methods for economic evaluation of health care programmes (3rd Edition). Oxford: Oxford University Press; 2005.

35 Kobelt G. Health Economics: An Introduction to Economic Evaluation (2nd Edition). London: Office of Health Economics; 2002.

36 Folland S, Goodman A, Stano M. Economics of Health and Health Care (5th Edition). New Jersey: Prentice Hall; 2006.

37 Sorenson C, Drummond M, Kanavos P. Ensuring value for money in health care: the role of health technology assessment in the European Union. London: European Observatory for Health Systems and Policies; 2008.

38 World Health Organization. Health technology assessment of medical devices. WHO Medical device technical series. Geneva: WHO Press; 2011.

39 Battista R, Hodge MJ. The evolving paradigm of health technology assessment: reflections for the Millennium. *CMAJ* 1999; 160(10): 1464-1467.

40 Jonsson E, Banta HD. Management of health technologies: an international view. *BMJ* 1999; 319: 1293.

41 Garau M, Mestre-Ferrandiz J. Access mechanisms for orphan drugs: a comparative study of selected European countries 2009. Available at: http://www.raredisease.org.uk/documents/OHEBriefingOrphanDrugs.pdf (accessed September 15, 2015).

42 Picavet E, Morel T, Cassiman D, Simoens S. Shining a light in the black box of orphan drug pricing. *Orphanet J Rare Dis* 2014; 9: 62.

43 The Dental and Pharmaceutical Benefits Agency. Health economics. Available at: http://www.tlv.se/In-English/medicines-new/health-economics (accessed September 18, 2015).

44 Koopmanschap M, Rutten F, Vanineveld B, Vanroijen L. The friction cost method for measuring indirect costs of disease. *J Health Econ* 1995; 14: 171–189.

45 Koopmanschap MA, van Exel JN, van den Berg B, Brouwer WB. An overview of methods and applications to value informal care in economic evaluations of healthcare. *Pharmacoeconomics* 2008; 26: 269–280.

46 van den Berg B, Brouwer WB, Koopmanschap MA. Economic valuation of informal care. An overview of methods and applications. *Eur J Health Econ* 2004; 5: 36–45.

47 World Health Organization. WHO Definition of Palliative Care. Available at: http://www.who.int/cancer/palliative/definition/en (accessed October 4, 2015).

48 The World Health Organization. Quality of Life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med* 1995; 41: 1403–1409.

49 Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess* 2001; 5(4).

50 Spieth LE, Harris CV. Assessment of health-related quality of life in children and adolescents: an integrative review. *J Pediatr Psychol* 1996; 21: 175–93.

51 Ravens-Sieberer U, Erhart M, Wille N, et al. Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *Pharmacoeconomics* 2006; 24: 1199–220.

52 World Health Organization. World Health Organization Constitution. Geneva: World Health Organization; 1947.

53 The EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16(3): 199–208.

54 Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992; 30: 473–83.

55 Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health Qual Life Outcomes* 2003; 1: 54.

56 Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care* 2001; 39(8): 800–812.

57 Uzark K, King E, Cripe L, et al. Health-related quality of life in children and adolescents with Duchenne muscular dystrophy. *Pediatrics* 2012; 130: e1559-e1566.

58 Davis SE, Hynan LS, Limbers CA, et al. The PedsQL in pediatric patients with Duchenne muscular dystrophy: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Neuromuscular Module and Generic Core Scales. *J Clin Neuromuscul Dis* 2010; 11: 97–109.

59 Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health* 2000; 21: 587–611.

60 von Neumann J, Morgenstern O. Theory of Games and Economic Behavior (3rd editon). New York: John Wiley; 1953.

61 Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care* 1989; 5: 559–575.

62 Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. *J Epidemiol* 2002; 12: 160–178.

63 Aronsson M, Husberg M, Kalkan A, et al. Differences between hypothetical and experience-based value sets for EQ-5D used in Sweden: Implications for decision makers. *Scand J Public Health* 2015; 43(8): 848–854.

64 Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Qual Life Res* 2003; 12(6): 599–607.

65 Wilson TD, Wheatley T, Meyers JM, et al. Focalism: a source of durability bias in affective forecasting. *J Pers Soc Psychol* 2000; 78(5): 821–836.

66 Hadorn DC. The role of public values in setting health care priorities. *Soc Sci Med* 1991; 32(7): 773–781.

67 The Dental and Pharmaceutical Benefits Agency. General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2). Available at: http://www.tlv.se/Upload/English/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf (accessed January 5, 2016).

68 The National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. Available at: http://www.nice.org.uk/article/pmg9 (accessed January 5, 2015).

69 Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997; 35: 1095–1108.

70 Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002; 21(2): 271–292.

71 Feeny DH, Furlong WJ, Torrance GW, et al. Multi-Attribute and Single-Attribute Utility Functions for the Health Utilities Index Mark 3 System. *Medical Care* 2002; 40: 113–128.

72 Landgraf JM, Abetz LN. Functional status and well-being of children representing three cultural groups: Initial selfreports using the CHQ-CF87. *Psychol Health* 1997; 12: 839–854.

73 Ravens-Sieberer U, Gosch A, Rajmil L, et al. KIDSCREEN-52 quality-of-life measure for children and adolescents. *Expert Rev Pharmacoecon Outcomes Res* 2005; 5(3): 353–364.

74 Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res* 2010; 19(6): 875–886.

75 Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life?: an analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes* 2007; 5: 1.

76 Jardine J, Glinianaia SV, McConachie H, et al. Self-reported quality of life of young children with conditions from early infancy: a systematic review. *Pediatrics* 2014; 134(4): e1129-48.

77 Ryen L, Svensson M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. *Health Econ* 2014: 24(10): 1289–1301.

78 Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. *JAMA* 2014; 311: 1052–1060.

79 Hanaoka C, Norton EC. Informal and formal care for elderly persons: how adult children's characteristics affect the use of formal care in Japan. *Soc Sci Med* 2008; 67(6): 1002–1008.

80 Triantafillou J, Naiditch M, Repkova K et al. Informal care in the long-term care system. European Overview Paper. European Centre for Social Welfare Policy and Research 2010. Available from: http://www.euro.centre.org (accessed February 8, 2015).

81 Australian Institute of Health and Welfare. Informal care. Available from: http://www.aihw.gov.au/informal-care-ageing (accessed February 9, 2015).

82 Schulz R, Beach SR. Caregiving as a risk factor for mortality: the Caregiver Health Effects Study. *JAMA* 1999; 282: 2215–2219.

83 Gouin JP, Hantsoo L, Kiecolt-Glaser JK. Immune dysregulation and chronic stress among older adults: a review. *Neuroimmunomodulation* 2008; 15: 251–259.

84 Lee S, Colditz GA, Berkman LF, Kawachi I. Caregiving and risk of coronary heart disease in U.S. women: a prospective study. *Am J Prev Med* 2003; 24: 113–119.

85 Perrin JM, Bloom SR, Gortmaker SL. The increase of childhood chronic conditions in the United States. *JAMA* 2007; 297: 2755–2759.

86 Wise PH. The future pediatrician: the challenge of chronic illness. *J Pediatr* 2007; 151: S6–S10.

87 Zarit SH, Todd PA, Zarit JM. Subjective burden of husbands and wives as caregivers: a longitudinal study. *Gerontologist* 1986; 26(3): 260–266.

88 Mosquera I, Vergara I, Larrañaga I, et al. Measuring the impact of informal elderly caregiving: a systematic review of tools. *Qual Life Res* 2016; 25(5): 1059–1092.

89 Briggs A, Sculpher M, Claxton K. Decision Modelling for Health Economic Evaluation (1st Edition). Oxford: Oxford University Press; 2006.

90 Briggs AH, Weinstein MC, Fenwick EA, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health* 2012; 15(6): 835–842.

91 Wright B. A history of social science and measurement. *Educ Meas* 1997; 52: 33–52.

92 Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess* 2009; 13.

93 Likert RA. A technique for the measurement of attitudes. *Arch Psychol* 1932; 140: 5–55.

94 Thurstone LL. Attitudes can be measured. *Am J Sociol* 1928; 23: 529–554.

95 Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297–334.

96 Nunnally JC, Bernstein IH. Psychometric theory (3rd Edition). New York: McGraw-Hill; 1994.

97 Hobart JC, Cano SJ, Zajicek JP, et al. Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 2007; 6: 1094–1105.

98 McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res* 1995; 4: 293–307.

99 Holmes WC, Shea JA. Performance of a new, HIV/AIDS-targeted quality of life (HAT-QoL) instrument in asymptomatic seropositive individuals. *Qual Life Res* 1997; 6(6): 561–571.

100 da Rocha NS, Chachamovich E, de Almeida Fleck MP, et al. An introduction to Rasch analysis for Psychiatric practice and research. *J Psychiatr Res* 2013; 47(2): 141–148.

101 Rasch G. Probabilistic models for some intelligence and attainment tests (1st Edition). Copenhagen: Danish Institute for Education Research; 1960.

102 Lundgren Nilsson Å, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM™) revisited. *J Rehabil Med* 2011; 43(10): 884–891.

103 Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007; 57(8): 1358–1362.

104 Ouyang L, Grosse SD, Kenneson A. Health care utilization and expenditures for children and young adults with muscular dystrophy in a privately insured population. *J Child Neurol* 2008; 23: 883–888.

105 Larkindale J, Yang W, Hogan PF, et al. Cost of illness for neuromuscular diseases in the United States. *Muscle Nerve* 2014; 49: 431–438.

106 McDonald CM, McDonald DA, Bagley A, et al. Relationship between clinical outcome measures and parent proxy reports of health-related quality of life in ambulatory children with Duchenne muscular dystrophy. *J Child Neurol* 2010; 25: 1130–1144.

107 Bray P, Bundy AC, Ryan MM, North KN, Everett A. Health-related quality of life in boys with Duchenne muscular dystrophy: agreement between parents and their sons. *J Child Neurol* 2010; 25: 1188–1194.

108 Bendixen RM, Senesac C, Lott DJ, Vandenborne K. Participation and quality of life in children with Duchenne muscular dystrophy using the International Classification of Functioning, Disability, and Health. *Health Qual Life Outcomes* 2012; 10: 43.

109 Lim Y, Velozo C, Bendixen RM. The level of agreement between child self-reports and parent proxy-reports of health-related quality of life in boys with Duchenne muscular dystrophy. *Qual Life Res* 2014; 23(7): 1945–1952.

110 Hu J, Jiang L, Hong S, et al. Reliability and validity of the Chinese version of the Pediatric Quality Of Life InventoryTM (PedsQLTM) 3.0 neuromuscular module in children with Duchenne muscular dystrophy. *Health Qual Life Outcomes* 2013; 11: 47.

111 Bray P, Bundy AC, Ryan MM, et al. Health status of boys with Duchenne muscular dystrophy: a parent's perspective. *J Paediatr Child Health* 2011; 47: 557–562.

112 Baiardini I, Minetti C, Bonifacino S, et al. Quality of life in Duchenne muscular dystrophy: the subjective impact on children and parents. *J Child Neurol* 2011; 26: 707–713.

113 Kohler M, Clarenbach CF, Boni L, et al. Quality of life, physical disability, and respiratory impairment in Duchenne muscular dystrophy. *Am J Respir Crit Care Med* 2005; 172: 1032–1036.

114 Elsenbruch S, Schmid J, Lutz S, et al. Self-Reported Quality of Life and Depressive Symptoms in Children, Adolescents, and Adults with Duchenne Muscular Dystrophy: A Cross-Sectional Survey Study. *Neuropediatrics* 2013; 44: 257–264.

115 Pangalila RF, van den Bos GA, Stam HJ, et al. Subjective caregiver burden of parents of adults with Duchenne muscular dystrophy. *Disabil Rehabil* 2012; 34: 988–996.

116 Simon VA, Resende MB, Simon MA, et al. Duchenne muscular dystrophy: quality of life among 95 patients evaluated using the Life Satisfaction Index for Adolescents. *Arq Neuropsiquiatr* 2011; 69(1): 19–22.

117 Nereo NE, Fee RJ, Hinton VJ. Parental stress in mothers of boys with duchenne muscular dystrophy. *J Pediatr Psychol* 2003; 28: 473–484.

118 Abi Daoud MS, Dooley JM, Gordon KE. Depression in parents of children with Duchenne muscular dystrophy. *Pediatr Neurol* 2004; 31: 16–19.

119 Boyer F, Drame M, Morrone I, Novella JL. Factors relating to carer burden for families of persons with muscular dystrophy. *J Rehabil Med* 2006; 38: 309–315.

120 Kenneson A, Bobo JK. The effect of caregiving on women in families with Duchenne/Becker muscular dystrophy. *Health Soc Care Community* 2010; 18: 520–528.

121 Chen JY, Chen SS, Jong YJ, et al. A comparison of the stress and coping strategies between the parents of children with Duchenne muscular dystrophy and children with a fever. *J Pediatr Nurs* 2002; 17: 369–379.

122 Samson A, Tomiak E, Dimillo J, et al. The lived experience of hope among parents of a child with Duchenne muscular dystrophy: perceiving the human being beyond the illness. *Chronic Illn* 2009; 5: 103–114.

123 Thomas PT, Rajaram P, Nalini A. Psychosocial challenges in family caregiving with children suffering from Duchenne muscular dystrophy. *Health Soc Work* 2014; 39: 144–52.

124 Mah JK, Thannhauser JE, Kolski H, et al. Parental stress and quality of life in children with neuromuscular disease. *Pediatr Neurol* 2008; 39: 102–107.

125 Magliano L, Patalano M, Sagliocchi A, et al. "I have got something positive out of this situation": psychological benefits of caregiving in relatives of young people with muscular dystrophy. *J Neurol* 2014; 261: 188–195.

126 McDonald CM, Henricson EK, Abresch RT, et al. The 6-minute walk test and other endpoints in Duchenne muscular dystrophy: longitudinal natural history observations over 48 weeks from a multicenter study. *Muscle Nerve* 2013; 48(3): 343–356.

127 Butland RJA, Pang J, Gross ER, et al. Two-, six-, and 12-minute walking tests in respiratory disease. *BMJ* 1982; 284: 1607–1608.

128 Vignos PJ, Spencer CE, Archibald KC. Management of progressive muscular dystrophy in childhood. *JAMA* 1963; 184: 389–396.

129 Brooke MH, Griggs RC, Mendell JR, et al. Clinical trial in Duchenne dystrophy. I. The design of the protocol. *Muscle Nerve* 1981; 4(3): 186–197.

130 Scott E, Eagle M, Mayhew A, et al. Development of a functional assessment scale for ambulatory boys with Duchenne muscular dystrophy. *Physiother Res Int* 2012; 17: 101–109.

131 Mayhew A, Mazzone ES, Eagle M, et al. Development of the performance of the upper limb module for Duchenne muscular dystrophy. *Dev Med Child Neurol* 2013; 55: 1038–1045.

132 Bladen CL, Rafferty K, Straub V, et al. The TREAT-NMD Duchenne muscular dystrophy registries: conception, design and utilisation by industry and academia. *Hum Mutat* 2013; 34: 1449–1457.

133 Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics* 1993; 4: 353–365.

134 British National Formulary (BNF). British Medical Association and the Royal Pharmaceutical Society. Available at: http://www.bnf.org (accessed October 15, 2012).

135 Department of Health. National schedule of reference costs 2010–11 for NHS trusts. Available at: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_131145.xls (accessed November 19, 2012).

136 Personal Social Services Research Unit (PSSRU). Unit costs of health & social care 2011. Available at: http://www.pssru.ac.uk/archive/pdf/uc/uc2011/uc2011.pdf (accessed October 15, 2012).

137 Die Kassenärztliche Bundesvereinigung (KBV). Einheitlicher Bewertungsmaßstab (EBM). Stand: 4. Quartal 2012. Available at: http://www.kbv.de (accessed November 4, 2012).

138 Rote Liste Service GmbH. Rote Liste 2012: Arzneimittelinformationen fur Deutschland. Frankfurt am Main: Rote Liste Service GmbH; 2012.

139 Mag Mutual Healthcare Solutions, Inc. Physicians' Fee & Coding Guide. Atlanta: Mag Mutual Healthcare Solutions; 2012.

140 OptumInsight, Inc. DrugReimbursement.com. Available at: http://www.drugreimbursement.com (accessed October 17, 2012).

141 Prontuario Farmaceutico Italiano. Available at: http://www.prontuariofarmaceutico.it (accessed March 16, 2013).

142 Regione Emilia Romagna. Nomenclatore tariffario regionale (in vigore dall'1 marzo 2013). Available at: http://www.saluter.it/documentazione/nomenclatore-tariffario-rer/nomenclatore_tariffario_rer_marzo2013.pdf/view (accessed April 23, 2013).

143 The Organisation for Economic Co-operation and Development (OECD). Work-life balance. Available at: http://www.oecdbetterlifeindex.org/topics/work-life-balance (accessed August 28, 2012).

144 Johannesson M, Borgquist L, Jonsson B, Rastam L. The costs of treating hypertension: an analysis of different cutoff points. *Health Policy* 1991; 18: 141–150.

145 United States Department of Transportation. The value of travel time savings: departmental guidance for conducting economic evaluations, revision 2 (2011). Available at: http://www.dot.gov/sites/dot.dev/files/docs/vot_guidance_092811c.pdf (accessed March 25, 2013).

146 Statistics Canada. National Population Health Survey Overview. Available at: http://www5.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=82-567-X (accessed 15 February 2015).

147 Kind P, Hardman G, Macran S. UK Population Norms for EQ-5D (1999). Discussion Paper 172. Centre for health Economics. The University of York.

148 Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* 1998; 316: 736–741.

149 Caro JJ, Briggs AH, Siebert U, et al. Modeling good research practices - overview: A report of the ISPOR-SMDM modeling good research practices task force-1. *Value Health* 2012; 15: 796–803.

150 McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008; 26(9): 733–744.

151 Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005; 14: 1523–1532.

152 Sullivan PW, Slejko JF, Sculpher MJ, Ghushchyan V. Catalogue of EQ-5D scores for the United Kingdom. *Med Decis Making* 2011; 31: 800–804.

153 Wong RK, Drossman DA, Weinland SR, et al. Partner burden in irritable bowel syndrome. *Clin Gastroenterol Hepatol* 2013; 11: 151–155.

154 Naglie G, Hogan DB, Krahn M, et al. Predictors of family caregiver ratings of patient quality of life in Alzheimer disease: cross-sectional results from the Canadian Alzheimer's Disease Quality of Life Study. *Am J Geriatr Psychiatry* 2011; 19: 891–901.

155 Torres AR, Hoff NT, Padovani CR, Ramos-Cerqueira AT. Dimensional analysis of burden in family caregivers of patients with obsessive-compulsive disorder. *Psychiatry Clin Neurosci* 2012; 66: 432–441.

156 Leroi I, Harbishettar V, Andrews M, et al. Carer burden in apathy and impulse control disorders in Parkinson's disease. *Int J Geriatr Psychiatry* 2012; 27: 160–166.

157 Mayhew A, Cano S, Scott E, et al. Moving towards meaningful measurement: Rasch analysis of the North Star Ambulatory Assessment in Duchenne muscular dystrophy. *Dev Med Child Neurol* 2011; 53: 535–542.

158 The National Institute for Health and Care Excellence. NICE guidance for ataluren. Available at: https://www.nice.org.uk/guidance (accessed April 15, 2016).

159 Gregori D, Petrinco M, Bo S, et al. Regression models for analyzing costs and their determinants in health care: an introductory review. *Int J Qual Health Care* 2011; 23(3): 331-341.

160 Basu A, Manning WG, Mullahy J. Comparing alternative models: log vs Cox proportional hazard? *Health Econ* 2004; 13(8): 749–765.

161 Rothman KJ. Epidemiology: An introduction (2nd Edition). New York: Oxford University Press; 2012.

162 Pangalila R. Quality of life in Duchenne muscular dystrophy: the disability paradox. *Dev Med Child Neurol* 2016; 58(5): 435–436.