Analysing and enhancing the MD Anderson Dysphagia Inventory: a mixed methods analysis of the head and neck cancer swallowing-related quality of life patient reported outcome tool

Katherine Toft

Student ID **2636551**

Submitted to the

University of Stirling

For the degree of

Doctor of Professional Health Studies

March 2023

ABSTRACT

Background

Dysphagia, or difficulty swallowing, is a highly prevalent symptom of Head and Neck Cancer (HNC) and treatment, and has a marked impact on patients' quality of life (QoL). Measurement of dysphagia-related QoL is therefore a key part of HNC clinical and research practice. Currently the only HNC-specific tool that caters to this need is the MD Anderson Dysphagia Inventory (MDADI). However, the MDADI is now more than 20 years old and has never undergone validation in the United Kingdom.

Aim

The aim of this study was to explore the psychometric properties and clinical utility of the MDADI.

Methods

This study followed a pragmatic, mixed methods approach to evaluate the MDADI, using the COSMIN methodology as a framework. Qualitative data from UK Speech & Language Therapists were gathered via an online survey, focussing on content validity and clinical utility of the MDADI, and analysed using a reflexive Thematic Analysis approach. Quantitative MDADI data from patients with HNC treated in NHS Lothian were used to analyse the MDADI's structural validity and internal consistency, using Item Response Theory (IRT). Data generated by both arms of the study were then combined to explore the presence of Differential Item Functioning (DIF) in specific tool items, and to formulate suggestions for shortened versions of the tool.

Results

Analysis of the survey data uncovered issues with both the content validity and clinical utility of the MDADI. IRT analysis of tool structural validity and internal consistency showed these properties to be acceptable, however DIF analysis for the variables of age, sex and socioeconomic status indicated these variables all affected patient responses to specific MDADI items. Pre- and post-treatment versions of a shortened 5-item MDADI were generated, in addition to suggestions for future development of the MDADI tool.

DECLARATION

I declare that this thesis is my own work except where otherwise stated.

Signed

Katherine Toft

CONFIRMATION OF ETHICAL APPROVAL

The study presented in this thesis has undergone ethical review and received approval from the following bodies:

- The University of Stirling NHS, Invasive or Clinical Research ethics panel: reference no. NICR 19/20 – 093
- IRAS via the Solihull Research Ethics committee: IRAS no. 279483; REC no. 20/WM/0319
- The NHS Lothian Research & Development Team: project number 2021/0039
- The NHS Lothian Caldicott Guardian granted permission to access patient data from patient notes: Application no. CRD20098

ACKNOWLEDGEMENTS

Writing this thesis part time whilst also working for the NHS and trying to maintain some semblance of a life during a global pandemic has been quite an undertaking, and I am exceedingly grateful to a number of people who have helped me along this journey.

First of all thank you to my supervisors Professor Jayne Donaldson and Dr Cath Best for their support, feedback and encouragement, and tolerance of my trademark overthinking at every stage of this process.

I am very thankful to NHS Lothian and the University of Stirling for jointly funding my place on the Stirling Clinical Doctorate programme: I would not have been able to achieve this otherwise.

Special thanks goes to people who have shown particular kindness during this process, who have given much needed guidance and advice, or displayed appreciated sympathy towards various predicaments: Professor Jo Patterson, Dr Kath Stoddart, Dr Federico Andreis, Andy Peters, my NHS Lothian SLT managers, Rachel Beaton, and Alan Harrow.

Thank you to my SLT colleagues and peers around the UK who took the time to be part of this study, and to my PPI group members for their valued insights and contributions.

To Rob, for always making me laugh, talking me down, and feeding me like a Queen; to my parents for absolutely always supporting me in everything I do; to Em, Kirsty and Annabelle for tolerating my nerdiness and providing much-appreciated decompression and distractions.

And finally, my most heartfelt thanks to all the people with head and neck cancer who I have worked with over the years who have contributed to this study both directly and indirectly. You have been the greatest teachers and it has been my honour and privilege to be involved in your care.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

1.1 Clinical context

Back in 2015 my NHS Lothian head & neck cancer Speech & Language Therapy (SLT) team started providing pre-treatment information sessions to patients about to undergo surgical or oncological treatment for head & neck cancer (HNC). Scottish SIGN guidelines mandated that these sessions take place but did not specify content. One of the things we wanted to achieve in these sessions was to accurately counsel patients about the physiological, functional and quality of life impact of the treatment they were about to receive on their eating, drinking and swallowing in order to help them prepare, and to manage their post-treatment expectations in the longer term.

We realised we had a wealth of clinical experience but no data on how patients were functioning before and after their treatment in terms of formal outcome measurement. So, we investigated the available tools, and wanting to capture data from several perspectives (physiological, functional and quality of life impact) we decided on a suite of HNC appropriate swallow outcome assessment tools.

Several years on we have become very familiar with these tools and their advantages and disadvantages, both in terms of data generated and clinical useability. One such tool is the MD Anderson Dysphagia Inventory (MDADI). This is an important and unique tool as it is the only tool specifically designed to assess dysphagia-related quality of life for patients with HNC. As such, the MDADI is also used widely in multi-centre clinical trials of HNC treatments, often as a primary outcome. The more we used it in clinical practice however, at both pre-treatment and post-treatment timepoints, the more frustrated with it we became. We felt the results it gave often did not match with clinician-rated swallowing outcome measures of swallowing or instrumental swallow assessments, or what patients were telling us during consultations. Also, the long length and written format of the tool meant that it was challenging to use for many of our patients.

The MDADI is 22 years old and was devised by a group of American clinicians, with minimal information available about the tool's development process. As I progressed through my Clinical Doctorate, I thought more and more about the importance of this tool in the published HNC literature and the mismatch with my clinical experience of its use in practice. I started to develop an idea for studying the tool in detail from several perspectives, to analyse its strengths and weaknesses. This would be with a view to laying the groundwork for future enhancement and strengthening of the tool, which could ultimately have a positive impact on patient care, clinical practice and research in this field.

1.2 Research aim, questions and outcome

The aim of this study is to explore the psychometric properties and clinical utility of the MDADI, and to provide direction for potential future development and strengthening of this tool which is in many ways central to HNC dysphagia practice and research. This study's research aim and questions are as follows:

1.2.1 Research Aim

To analyse and evaluate the MD Anderson Dysphagia Inventory on a UK-based population.

1.2.2 Research Questions

1. What is the content validity and clinical utility of the MD Anderson Dysphagia Inventory from a UK SLT perspective?
2. What are the results of an exploration of psychometric properties of the MDADI on a UK population using Item Response Theory?
3. What are the potential factors which might result in differential item functioning?

1.3 Thesis overview

Chapter One of this Clinical Doctorate thesis has provided the clinical background and context to the formulation of the research aim and questions, which have also been presented. The structure of the rest of this thesis is as follows.

Chapter Two sets the scene for the thesis with an introduction to HNC associated dysphagia, and the current 'state of the art' in terms of patient-reported outcome measures (PROMs) used to measure dysphagia-related quality of life. This chapter also sets forth a novel summary of the concept of clinical utility, an aspect of tool properties which is distinct from but complementary to more familiar psychometric properties of PROMs. Also presented is the current knowledge of the properties of the MDADI tool with regards to all aspects of its use. The chapter closes with a summary argument for the necessity for an in-depth analysis of the MDADI's properties.

Chapter Three presents the methodological decision-making and action process around designing and carrying out this study. The benefits of a pragmatic, mixed methods approach are presented and the investigations made in the study are outlined. Ethical and research governance considerations are discussed, and the structure is formed for the presentation of data analysis and results.

Chapters Four, Five and Six contain the results for both the qualitative and quantitative strands of the study, followed by a synthesis of data provided by these two strands. A

complex picture forms of the content validity, clinical utility, structural validity and internal consistency of the MDADI. The presence of differential item functioning within the tool is confirmed; in addition, suggestions for item reduction are presented.

Chapter Seven contains a discussion of the study results set in the context of the published literature. The fresh insights into properties of the MDADI not previously investigated are considered. Finally, Chapter Eight presents the implications of the thesis results for clinical practice, policy and research practice and draws final conclusions.

CHAPTER 2: LITERATURE REVIEW

2.1 Background

2.1.1 Head and neck cancer

Head and neck cancer (HNC) is the 8th most common cancer in the UK (Cancer Research UK, 2018). The term 'head and neck' refers to more than 30 subsites in the upper aerodigestive tract (NICE, 2004), covering tumours arising from the nose, mouth, pharynx and larynx.

Curative treatment modalities for HNC include primary surgery, often taking the form of significant tissue resection and reconstruction, or primary oncological treatment with radiotherapy, often with added chemotherapy either given as neo-adjuvant treatment or concurrently (Chow, 2020). (Chemo)radiotherapy is often also added as an adjuvant treatment following primary surgery. Both surgical and oncological treatments, with the primary aim of cancer cure, have the potential to cause significant side effects due to local treatment toxicity or structural and physiological alteration in function.

HNC and its treatment can impact on many basic and key life functions given the primary tumour's location in the upper aerodigestive tract. Breathing, speech, swallowing, body image and aesthetic appearance may be affected, which all influence patients' daily function and quality of life. These side effects can be severe and long-lasting, and require significant levels of specialist support and rehabilitation from a multidisciplinary team of healthcare professionals, including surgeons, oncologists, dental practitioners, clinical nurse specialists, speech & language therapists and dietitians (Schache et al., 2021).

Traditionally associated with an older population and linked to alcohol and tobacco consumption, over recent years the demographic characteristics of the HNC patient population have been changing (Klussmann, 2017). There has been a marked rise in HNC associated with the Human Papilloma Virus (HPV). Patients diagnosed with HPV-related HNC tend to be younger, and have better survival outcomes, therefore this cohort of patients are faced with living longer with the side-effects of treatment. There is an increasing pressure for a better understanding of post-treatment and long-term functional and psychosocial outcomes of treatment, to facilitate the decision-making process pre-treatment and to develop and improve options for post-treatment rehabilitation (Hutcheson et al., 2012, Starmer et al., 2014).

When preparing patients to make informed decisions about their cancer treatment and counselling them about expected outcomes, it is vital that information on potential impact of treatment on functional aspects including swallowing is provided, as stated in The European

Cancer Patients' Bill of Rights (Lawler et al., 2014). Patients, carers and families want healthcare professionals to provide accessible, realistic information on functional impact so that they can understand what this may be and how it may affect their life during and after treatment (Bozec et al., 2016, Brockbank et al., 2015, D'Souza et al., 2017).

2.1.2 Head and neck cancer associated dysphagia

Crucial swallowing related structures lie within the treatment fields for the majority of HNCs, and damage to these structures by tumour or treatment can result in oropharyngeal dysphagia, which has been described as the 'key functional outcome' post-treatment for HNC (Wilson et al., 2011). Whilst diagnosis and survival statistics for HNC are improving (Klussmann, 2017, Public Health Scotland, 2020), oropharyngeal dysphagia, or difficulty swallowing, continues to be a highly prevalent treatment-associated morbidity (Hutcheson et al., 2018).

People with HNC may present with dysphagia prior to treatment, due to tumour location and impact (Ursino et al., 2022), or develop dysphagia during and/or after treatment due to the side–effects of treatment. Risk of developing dysphagia varies with treatment regime (e.g., radiotherapy alone vs. chemoradiotherapy) and cancer site and stage (Mortensen et al., 2013, Manikantan et al., 2009). Prevalence of dysphagia in the HNC patient group is difficult to quantify as it is dependent on tumour site, stage, and treatment modality, but is estimated at around 45% (Hutcheson et al., 2018, Francis et al., 2010). Dysphagia following HNC and its treatment involves not only physical impairment of swallow biomechanics, causing increased medical risk of malnutrition and aspiration pneumonia, but also a significant emotional and social impact detrimentally affecting quality of life as 'normal' patterns of eating and drinking are disrupted (McQuestion et al., 2011, Hutcheson et al., 2012, Baijens et al., 2021).

In the acute and post-acute stages following treatment, there is a complex interplay between potential swallow physiology dysfunction, and other treatment-associated toxicities such as dysgeusia (altered taste), nausea, odynophagia (pain on swallowing), and xerostomia (reduced saliva) (Manikantan et al., 2009). People with HNC also often have to have multiple dental extractions as part of their workup for treatment and this, coupled with reported prevalence of 10-50% for treatment associated trismus (restricted jaw opening) can also impact on eating and drinking, although they are not technically aspects of dysphagia per se (Abboud et al., 2020, Clough et al., 2018). HNC dysphagia can persist and further deteriorate years into cancer survivorship (Hutcheson et al., 2012).

Difficulties with eating, drinking and swallowing are commonly reported by patients as one of the most significant functional outcomes of their HNC and treatment (Mendez et al., 2020),

and people with HNC-related dysphagia describe its complex interaction with other social, emotional and physical aspects of their lives (Dawson et al., 2019).

2.1.3 Assessment of head and neck cancer associated dysphagia

Speech & Language Therapists (SLTs) are the healthcare professionals with the responsibility for the assessment and management of oropharyngeal dysphagia (RCSLT, 2006), and are involved with the evaluation and management of dysphagia associated with HNC at pre-, mid- and post-treatment stages (Krisciunas et al., 2012).

Dysphagia assessment is important in both HNC research and clinical practice, and swallowing outcome measurement is stipulated by the Royal College of Speech & Language Therapists (RCSLT), the Scottish Government, and the British Association of Head & Neck Oncologists (RCSLT, 2006, The Scottish Government, 2010, Schache et al., 2021). To understand the full spectrum of physiological, functional and quality of life impact of HNC-related swallowing difficulties, formal dysphagia assessment is required. Dysphagia-specific outcome assessment tools generate data that demonstrate baseline pre-treatment function, and impact of treatment and rehabilitation interventions on patients' swallow function (Rieger et al., 2010) and facilitate pre-treatment counselling regarding post-treatment functional outcomes (Rogers et al., 2015b). The latest guidance from the European Head and Neck Society focussing on HNC survivorship care emphasises the potential psychosocial impact of dysphagia and the need for this to be assessed, monitored and managed (Verdonck-de Leeuw et al., 2022).

High demands are placed on outcome measurement tools: they must be psychometrically robust, clinically meaningful and holistic in scope, capturing reliable and valid data that is useful to clinicians, researchers, service users and service commissioners (Speyer et al., 2022). Tools also need to be clinically practical and useable if they are to become embedded in everyday practice and be practical and feasible for use in research.

Assessment of swallowing can take many different forms. Instrumental evaluation is the gold standard for assessing swallow anatomy and physiology, with the most used tools being Modified Barium Swallow x-ray assessment (also known as videofluoroscopy) and Fibreoptic Endoscopic Evaluation of Swallowing (FEES). Other clinician led tools involve measures of swallow efficiency such as the 100ml Water Swallow Test (WST) (Patterson et al., 2011), ratings of patients' ability to eat different food textures such as the Performance Status Scale- Head and Neck (PSS-HN) (List et al., 1990) and ratings that give an indication of patients' reliance on enteral feeding such as the Functional Oral Intake Scale (FOIS) (Crary et al., 2005). In addition to these measures, there are patient-reported outcome measures of

dysphagia, such as the MD Anderson Dysphagia Inventory (MDADI) (Chen et al., 2001) and the SWAL-QOL tool (McHorney et al., 2000).

Manikantan et al. (2009) argue that instrumental assessment is essential for rehabilitation planning and evaluation. However, Nund et al. (2014a) suggest that given the far-reaching impact of dysphagia on patents' lives, the focus of intervention needs to be more than impairment based; if this is the case, outcome tools used in clinical and research practice need to capture this. Much existing research suggests that clinicians' assessment of swallow physiology is poorly aligned with patients' report of their experience of swallowing function and its impact on their quality of life, therefore it is key to ensure that both clinician and patient ratings are incorporated in a comprehensive swallowing evaluation (Pedersen et al., 2016, Baijens et al., 2021, Kirsh et al., 2019). This may be because, as suggested by Rogus-Pulia et al. (2014) "patients sense a general difficulty with swallowing but have less awareness of specific symptoms of dysphagia" (p.223). Conversely, a recent study by Wishart et al. (2022) has shown a higher degree of agreement between instrumental measures of pharyngeal swallow physiology and patient-reported dysphagia associated QoL. The debate in the literature regarding correlation of clinician-led and patient-led outcomes tools underlines the need for capture of multiple aspects of patients' eating, drinking and swallowing in HNC.

There is a large body of published literature concerning instrumental assessment of dysphagia in HNC, but significantly less concerning patient reported outcome measures (PROMs) of HNC-related swallowing difficulties. However, the value that PROMs can add when included in comprehensive HNC assessment has become increasingly clear over recent years. PROMs provide data essential to informing patients about expected health-related quality of life (HRQoL) outcome post-treatment, improve communication between the patient and the MDT and improve screening for and provision of other interventions that patients may require (Rogers and Barber, 2017).

To obtain a comprehensive and nuanced overview of a person with HNC's swallowing, eating and drinking presentation it would therefore seem that a battery of outcome measures and assessments should be used with a broad scope, to capture these different elements of the complex entity that is dysphagia.

2.2 PROMs in HNC

PROMs are tools used to measure a patient's "health, quality of life, or functional status associated with health care or treatment" (Weldring and Smith, 2013 p.61). PROMs enable the measurement of the impact of a condition or treatment on patients' day to day lives and are a common way of assessing HRQoL. They provide information on the impact of an

illness or treatment on a patient's personal and social context, and are an important adjunct to clinical, objective assessments (Ferrans, 2007). PROMs are also a way of conceptualising and quantifying patient experience, and can also be used as a way of assessing quality of services (Gibbons, 2016). Consideration of patients' quality of life as part of their healthcare provision in HNC is mandated by national guidelines (Rogers et al., 2016). Silveira et al. (2018) describe PRO assessment in HNC as an 'essential' part of clinical practice in this field. Several PROM tools exist that can be used by patients with HNC to assess their eating, drinking and swallowing, and these will now be discussed.

2.2.1 PROMs for dysphagia: functional health status vs. health-related quality of life

Speyer et al. (2014) suggest that existing tools for assessing the impact of dysphagia on patients' QoL can be divided into two groups: those considering the impact of dysphagia on health-related QoL (HRQoL), that is the patient's perception of their swallowing difficulties and their impact across physiological, functional and quality of life domains, and those looking at functional health status (FHS), which is the influence of dysphagia on specific aspects of physical function (Smith et al., 1999). It could be argued that HRQoL tools give a broader spectrum of information than FHS tools as they cover more domains than just physical function.

Speyer et al. (2014) and Timmerman et al. (2014) performed systematic reviews of the literature to identify all extant dysphagia assessment tools in these two domains. Speyer's group used the COSMIN tool (Mokkink et al., 2010b) to assess the psychometric properties of all papers identified, and in both papers, reviewers used group discussion to come to consensus whether tools assessed HRQoL or FHS aspects of dysphagia. Table 1 shows the tools considered by both papers and which domains they assess.

Tools assessing dysphagia related HRQoL	Tools assessing dysphagia related FHS
MD Anderson Dysphagia Inventory (MDADI)	Eating Assessment Tool (EAT-10)
SWAL-QOL	Swallowing Outcome After Laryngectomy (SOAL)
European Dysphagia Group Questionnaire (EDGQ)	Self-report Symptom Inventory
European organisation for the research and treatment of cancer quality of life questionnaire – gastric cancer module (EORTC QLQ-ST022)	Sydney Swallow Questionnaire (SSQ)
Deglutition Handicap Index (DHI)	Deglutition Handicap Index (DHI)
Dysphagia Handicap Index (DHI)	Dysphagia Handicap Index (DHI)
European organisation for the research and treatment of cancer quality of life questionnaire – esophageal, esophagogastric junction, or gastric cancer module (EORTC QLQ-OG25)	Dysphagia Short Questionnaire (DSQ)
European organisation for the research and treatment of cancer quality of life questionnaire – head and neck cancer module (EORTC QLQ-H&N35)	Dysphagia in Multiple Sclerosis questionnaire (DYMUS)
	Mayo Dysphagia Questionnaire-30 (MDQ-30)
	Swallowing Disturbance Questionnaire (SDQ)

Table 1: Tools for measuring dysphagia related QoL and FHS

Of these tools, the EORTC QLQ-H&N35 and the MDADI are the only tools which consider HRQoL rather than FHS specifically for patients with HNC. Of these two, the MDADI is the only measure that has a specific focus on HNC-related dysphagia; the EORTC QLQ-H&N35 includes some dysphagia related items but these are few and generic with the tool having a wider focus than just eating, drinking and swallowing. As this is the case, this thesis focuses on an analysis of the MDADI in its unique position of being the only tool of its kind: that is one assessing patient reported impact of HNC dysphagia on HRQoL.

2.2.2 The MD Anderson Dysphagia Inventory

The MD Anderson Dysphagia Inventory (MDADI) PROM is a self-administered written questionnaire which quantifies dysphagia-related quality of life and is specifically designed for use in HNC. It was developed and validated on a cross-sectional sample of 100 English-speaking adult patients with HNC and dysphagia in the USA in the late 1990s (Chen et al., 2001). The MDADI is one of the most frequently used dysphagia outcome assessment tools in HNC research practice internationally (Ojo et al., 2012) and is often used as a main outcome tool in multicentre trials (Hutcheson et al., 2016, Castellano and Sharma, 2019). In recent years it has been the primary or secondary endpoint in large multicentre trials such as De-ESCALaTE, PATHOS, CompARE, DARS and PRO-ACTIVE (Owadally et al., 2015, Petkar et al., 2016, Martino et al., 2021, Mehanna et al., 2017, Mehanna et al., 2019). The MDADI is also often used as a 'gold standard' in the validation of other dysphagia assessment tools for use with people with HNC, such as the DIGEST (Hutcheson et al., 2017) and the Sydney Swallow Questionnaire (Dwivedi et al., 2010).The MDADI was originally developed for research use (K. Hutcheson, personal communication, November

24, 2020) but in reality it is used both as a research outcomes tool and also in clinical practice (Lin et al., 2022).

The tool consists of 20 items each rated on a 5-point Likert scale. Scoring the tool produces a global score (MDADI – G), scored from the first item ("my swallowing impacts my day-to-day life") and a composite score (MDADI –C) of the remaining 19 items. These 19 items are grouped into 3 subscales: Physical, Emotional, and Functional, allowing quantification of patients' perception of the impact of their swallowing ability in these areas. MDADI-G and MDADI-C scores range from 20 (low functioning) to 100 (high functioning). The MDADI item response format is a Likert scale (strongly disagree/disagree/no opinion/agree/strongly agree). To convert responses into scores, items are scored from 1 – 5 points ('strongly agree' scoring 1, 'strongly disagree' scoring 5) except for the 'reverse worded' items 5 and 15, where the scoring is reversed (i.e., 'strongly agree' scores 5, 'strongly disagree' scores 1). The MDADI in full can be viewed in Appendix A.

The authors of the MDADI do not provide specific instructions on how it should be employed in practice, and there are no other published guidelines on its use. In my own clinical practice (NHS Lothian regional HNC service) information gained from the MDADI is used in several ways. Over time with repeated use of the tool with many patients, an impression is formed of impact of dysphagia on patients' HRQoL by different variables such as tumour site and treatment modality. This informs patient counselling pre-treatment as to the potential impact of treatment and the expected recovery trajectory. Also, at a within-patient level, use of the tool both pre- and post-treatment allows individual response to treatment or swallow interventions to be gauged, and feedback provided comparing the two data points. Use of the MDADI as part of a battery of dysphagia assessments also allows assessment of if and how patients' perceptions of their dysphagia differ from clinical perception, perhaps highlighting the need for further discussion, counselling around reframing of expectations of recovery post-treatment, or the need for referral on to other services such as psychology or dietetics.

The next sections of this literature review will explore PROM tool quality assessment and clinical utility to provide background context prior to focussed review and analysis of the current state of knowledge concerning the MDADI tool's properties.

2.3 PROM tool quality assessment

It is imperative that HRQoL PROMs are psychometrically sound so that their results can be used with confidence in their validity and reliability (Timmerman et al., 2014). However, current UK HNC guidelines such as the BAHNO standards (Schache et al., 2021) do not

provide any guidance on how to assess PROM psychometric quality when selecting tools for use in research or clinical practice.

PROM/HRQoL data need to be captured in a reliable and valid way, so that they provide quality data to complement other measures, research outcomes and clinical practice (Reeve et al., 2013). However as discussed by Rosenkoetter and Tate (2018) in their review of instruments designed to assess the psychometric properties of tools: the field of psychometrics is vast, and "there remains complexity with regard to definitions and standards of assessment of psychometric properties" (p. 114). Nevertheless, setting standards for this assessment is a valid and important task, and attempts have been made to address this by several international bodies, such as the International Society for Quality-of-Life Research (ISOQOL), the Patient Reported Outcomes Measurement Information System (PROMIS) Group and the COSMIN initiative. Lorente et al. (2020) performed a meta-review providing the widest overview to date of existing tools that can be used to assess HRQOL PROM psychometric properties. Their review found that the COSMIN tool is the most widely used tool in the PROM literature, and that it is also the most comprehensive tool currently available.

The COSMIN initiative has attempted to address the lack of available standardised PROM assessments with the formulation of the COSMIN methodological quality checklist (Mokkink et al., 2010b). This checklist focuses on evaluating both the methodological strengths and weaknesses of studies considering psychometric properties of PROMs, and the psychometric properties of the PROMs themselves. The tool was originally formulated through a Delphi process involving an international panel of 57 experts with backgrounds in epidemiology, psychology, clinical practice and statistics. The most recent iteration of the tool was published in 2019 (Mokkink et al., 2019, Terwee et al., 2018) and this includes a stronger focus on content validity assessment of PROMs as the authors consider this to be the most important psychometric property. The updated content validity COSMIN subsection went through a robust Delphi process of its own in its development, involving 159 experts from 21 countries, with the final version being subject to pilot testing and further amendment.

Table 2 below details the content of the psychometric domains covered by COSMIN as per the COSMIN manual (Mokkink et al., 2018).

COSMIN Domain	Measurement property	Aspect of a measurement property	Definition
Reliability			The degree to which measurement is free from error
	Internal consistency		The degree of item interrelatedness
	Reliability		How 'true' differences between patients contribute to the total variance in measurement
	Measurement error		Concerns the error of an individual's score that do not relate to changes in the construct being measured
Validity			The degree to which a PROM measures the construct(s) it purports to measure
	Content validity		Are all items relevant to the construct being measured? Are all key concepts included? Are the items comprehensible to the population of interest? (Terwee et al., 2018)
		Face validity	The degree to which items in a PROM looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured
		Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured
		Hypothesis testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM
	Criterion validity		The degree to which the scores of a PROM are an adequate reflection of a 'gold standard'
Responsiveness			The ability of a PROM to detect change over time in the construct to be measured.
	Responsiveness		Idem responsiveness
Interpretability			The degree to which one can assign qualitative meaning to a PROM's quantitative scores or change in scores.

Table 2: COSMIN domains, measurement properties and aspects of measurement properties (Mokkink et al., 2018)

Lorente's meta-review of HRQoL PROM assessment tools does not however go into specific detail about the different statistical ways of assessing psychometric properties. Therefore, there is no consensus for a 'gold standard' of PROM psychometric assessment by either statistical or other means. However, as the COSMIN tool currently has the strongest evidence base and most comprehensive properties, this would appear to be the tool of

choice in PROM psychometric assessment until an alternative internationally agreed tool comes to the fore.

2.3.1 Classical Test Theory vs. Item Response Theory assessment of PROMs

Classical Test Theory (CTT) has traditionally been the mathematical standard for PROM tool development and validation; however, limitations with this approach are increasingly being highlighted in the literature. In recent years other statistical models such as Rasch Measurement Theory (RMT) and Item Response Theory (IRT) have been used to develop and assess tools, showing what these models can add to the traditional CTT approach.

IRT statistical models were originally developed for assessing educational testing tools, but are appropriate for use in healthcare tool analysis as these tools also assess latent constructs (e.g. depression or quality of life) (Kean et al., 2018). In 2019, the Journal of Patient Reported Outcomes published an issue which included a series of articles comparing different statistical approaches to PROM psychometric evaluation. In the issue commentary, Patrick (2019) describes these different approaches to PROM analysis as 'many ways to skin a cat' and suggests that CTT and IRT both have strengths and weaknesses in what they can bring to tool property assessment. Conversely, Cordier et al. (2018) suggest that the statistical advantages offered by IRT outweigh the advantages of simpler and more familiar CTT analysis.

Wilmskoetter et al. (2019) suggest that CTT's assumption that an instrument is "invariant, objective and stable regarding its measurements" (p. 2) may lead to inaccurate assessment of a tool's quality. IRT does not make these assumptions. Stover et al. (2019) suggest that IRT models are helpful as they are flexible and provide information at both item and scale level, meaning that they can be used both for tool design and for existing tool refinement.

Frost et al. (2007) consider what constitutes sufficient reliability and validity of a PROM. They suggest that a mix of qualitative and quantitative approaches to tool assessment gives a more comprehensive picture, and that producing more evidence about any specific tool provides greater confidence in the data it generates. From a quantitative perspective, they consider CTT and IRT as legitimate ways of providing estimates of a PROM's reliability and validity.

IRT has been used to assess the psychometric properties of dysphagia outcomes tools in clinical areas other than HNC. Kean et al. (2018) describe their use of the method to analyse the Eating Assessment Tool (EAT-10). These authors suggest the use of IRT (versus CTT) to develop, validate and refine tools has the capacity to suggest ways to ultimately improve tool measurement precision. This has implications for tool use in both clinical and research

practice. For example, reduced error allows for more accurate measurement of clinical change in response to intervention, and in trial research would allow for reduced sample size. IRT analysis has demonstrated issues with the psychometric properties of commonly used dysphagia assessment tools (the EAT-10 and the SWAL-QOL) that were not previously identified with CTT assessment, such as issues with structural validity (Wilmskoetter et al., 2019, Cordier et al., 2017, Cordier et al., 2018).

As the most comprehensive PROM assessment tool available (Lorente et al., 2020), the most recent iteration of the COSMIN checklist includes guidance on both CTT and IRT aspects of PROM tool assessment (Prinsen et al., 2018). Specifically, IRT analysis is considered around structural validity and differential item functioning (DIF) in cross-cultural validity and measurement invariance. DIF concerns whether scores on a tool are comparable across different subgroups of respondents when the underlying variable, such as dysphagia-related QoL, has been controlled for (Teresi and Fleishman, 2007).

The potential of an IRT approach to add value to PROM tool analysis in the specific area of HNC has been highlighted by Silveira et al. (2011), who found that DIF was present in HRQoL tools commonly used in HNC practice. In their study, IRT DIF analysis showed that patient age and gender influenced patient HRQoL scores. In their European White Paper on oropharyngeal dysphagia in HNC, Baijens et al. (2021) recommend that IRT be used in conjunction with CTT when assessing the psychometric qualities of outcomes tools, and name the COSMIN tool as a suitable choice for assessing the psychometric properties of dysphagia assessments.

In summary then, both CTT and IRT provide means to assess the psychometric quality of PROM tools, and there are strengths and weaknesses to both approaches. However, IRT has not to date been used to assess the psychometric qualities of HNC dysphagia-related PROMs.

In addition to the psychometric evaluation of PROMs, COSMIN suggests that whether a tool is ‘useable’ in practice is also an important consideration, but this is difficult and complex to assess, and may be more suitable to a qualitative than a quantitative methodological approach. What the literature says about this, and how to evaluate PROM usability, will be considered in the next section of this literature review.

2.3.2 Assessment of PROM clinical utility and practicality

2.3.2.1 PROM Clinical utility

Patrick (2019), whilst summarising different approaches to psychometric assessment of PROMs, emphasises that many salient properties of a tool are not covered by traditional psychometric analyses, and that these more qualitative aspects, such as the use of a tool in its context, require 'additional consideration', although he does not suggest what form this might take. Whilst the COSMIN tool (Mokkink et al., 2019) focuses primarily on the psychometric properties of PROM tools, how 'useable' the tool is in clinical practice, or a tool's 'pragmatic characteristics', are also of paramount importance (Kroenke et al., 2015). The COSMIN authors acknowledge this as an area that is currently lacking in terms of COSMIN content, even though currently it is deemed to be the most comprehensive tool of its kind.

In their comprehensive metareview of tools designed to assess psychometric quality of PROMs, Lorente et al. (2020) highlight that although tool 'usability' is a key factor, this is not yet consistently incorporated into PROM tool assessments. COSMIN domains of interpretability (can qualitative meaning be assigned to quantitative scores) and validity (is the tool relevant to the specific caseload) have relevance here. In practice there are more criteria that are relevant when assessing or considering use of a PROM tool, specifically around whether a tool is practical to use. It could be argued a tool's 'burden' is more relevant in clinical rather than research practice (Reeve et al., 2013), when tool use has to be incorporated into busy clinics and is not supported by research staff and infrastructure. However, tools also still require to be feasible to use when employing them to collect data for research use.

In the literature these aspects of tool usability are generally referred to as 'clinical utility' (Lam et al., 2020). The PROM literature was therefore reviewed to come to an understanding and definition of 'clinical utility', and how this can be assessed, so that this could be applied to the current investigation of the MDADI. An analytical literature review aiming to explore the definition and understanding of clinical utility of PROMs follows.

2.3.2.2 Analytical review of the literature concerning PROM clinical utility

Medline, CINAHL, PubMed and EMBASE databases were accessed to perform a scoping review of the literature concerning clinical utility assessment of PROMs. This was performed to assess the quality and detail of the current evidence base and to allow formulation of assessment criteria for clinical utility that could be used in the current study. MEsH terms such as "Patient Reported Outcomes" and "Outcomes Research" were used in combination with keywords such as 'clinical utility'. The full database search strategy is provided in

Appendix B. Abstracts were reviewed and the inclusion and exclusion criteria as listed in Table 3 were applied.

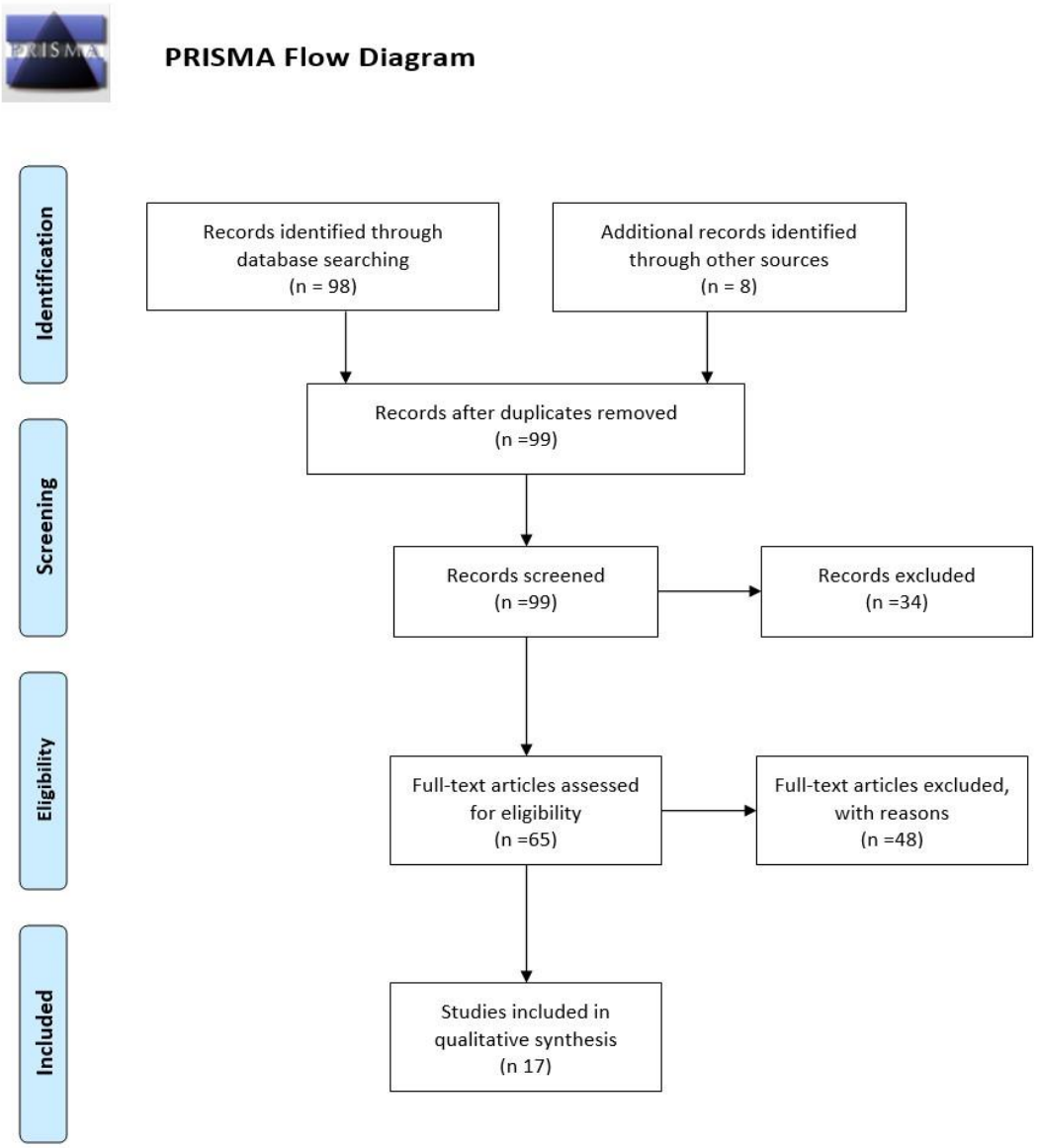
Inclusion criteria
Focus on patient reported outcome tools only
Provision of assessment criteria for clinical utility/practicality of a PROM tool
Papers focussing on clinician perspective
Papers focussing on both patient and clinician perspectives
Papers focussing on tools designed for use with adults
Exclusion criteria
Papers focussing on patient perspective only
Papers focussing on PROM implementation rather than assessment of the tool itself
Grey literature
Unpublished literature
Papers focussing on tools designed for use with children
Papers focussing solely on development of apps or e-tools

Table 3: PROM clinical utility literature search inclusion/exclusion criteria

Many papers in the initial cohort focussed only on general barriers and facilitators to PROM implementation in clinical services, rather than actual tool utility assessment criteria, and were therefore excluded. Studies considering patient perspectives only were also excluded given the focus on clinician data in the current study. Papers focussing on development of apps or e-tools were excluded as review of abstracts indicated these papers focussed on utility issues specific to the novel electronic format which are not applicable to the MDADI, the tool which is the focus of this thesis. At the time of writing, an electronic or app version of the MDADI has not been published.

Figure 1 below shows a PRISMA diagram (Moher et al., 2009) illustrating the search process. Although the PRISMA reporting system was originally intended for reporting on the process of carrying out a Systematic Review (SR) of the literature, I have used it to illustrate my literature search process for this review as although not an SR, I wanted to be thorough and systematic in my approach to the literature search. There is precedent for using PRISMA in non-SR literature reviews (McGowan et al., 2020). This method conveys the numbers of articles this kind of search reveals versus the numbers that meet inclusion criteria and provides an excellent framework for a systematic and logical way of approaching and representing the search.

Figure 1: Clinical Utility literature search PRISMA diagram



The 17 papers found to be relevant to the inclusion criteria will now be discussed and synthesised with a view to establishing an overview of what 'clinical utility' of a PROM tool constitutes. This knowledge will then guide assessment of MDADI clinical utility in this study's design.

As acknowledged by de Klerk et al. (2018), "the complexity of clinical utility makes its evaluation a challenge" (p.81). With the exception of Kroenke et al. (2015) (who used the term 'pragmatic characteristics'), all other studies used the term 'clinical utility' to signify aspects of PROM tools not covered by traditional psychometric domains of validity and reliability: aspects that cannot be statistically assessed and rather require qualitative assessment, such as 'ease of use' or 'patient burden' (Brasil et al., 2018). However, on examination of the selected papers, it becomes clear that there is no universally accepted, clearly delineated definition of the term 'clinical utility'. This is perhaps partly due to the fact

that this is an area of investigation that has only recently gained momentum, indeed many papers reviewed commented that their consideration of clinical utility was the first in their field (Jones et al., 2020, Galantino et al., 2015). Authors of reviewed papers have also come to their 'clinical utility' criteria through different methods. Four papers in this review developed their criteria for clinical utility from qualitative research involving patients, clinicians and managers (Montgomery et al., 2020, Aiyegbusi et al., 2020, Turner et al., 2020, Nic Giolla Easpaig et al., 2020), ten papers were reviews or systematic reviews of the PROM literature in their field synthesising findings from multiple papers, and three papers were opinion pieces synthesising clinical experience (Higginson and Carr, 2001, Vickers and Chen, 2017, Kroenke et al., 2015).

Terms such as 'feasibility' and 'acceptability' are used across several papers to describe aspects of clinical utility, however the definitions of these terms are not consistent. In their protocol for a systematic review looking at outcome measures used with patients receiving enteral feeding, Simpelaere et al. (2016) define 'feasibility' as comprising tool administration time, mode of administration, and whether facilitation is required to help patients complete the tool; however they provide no evidence for how these domains were selected. In Brasil et al. (2018)'s systematic review protocol for PROMs used with patients with acute coronary syndrome, 'feasibility' is characterised as amounting to 'ease of use' (although this is not defined), form of administration, duration and 'patient burden' (again not defined). Furthermore Nic Giolla Easpaig et al. (2020) contend that consideration of feasibility include whether the tool is easy for patients to navigate, whereas Thompson et al. (2016) name feasibility as a relevant aspect of PROM clinical utility but do not provide a definition of what this means.

Similarly with the term 'acceptability': to Nic Giolla Easpaig et al. (2020) this means whether there are clinical benefits to using the tool, whereas de Klerk et al. (2018) define acceptability as relating to the client, carers and their family, but doesn't expand on what this might constitute.

Across all 17 papers reviewed, no two papers defined clinical utility in the same way. As this is the case, definitions were summarised and analysed for patterns to see if overarching themes could be identified across all papers.

When each definition of clinical utility was studied, it became apparent that there were two major themes that could encapsulate discrete aspects of clinical utility discussed in each paper. To facilitate investigation of clinical utility of the MDADI in this study, I have classified these two themes as '*relevance*' and '*usability*' and I have defined these aspects of clinical utility in the following way:

‘Relevance’- refers to the relevance of the tool to both patients and clinicians, including aspects such as whether there are clinical benefits to using the tool, whether all items within the tool are pertinent to the clinical area, whether the tool covers all patients in the target group irrespective of level of disability, and whether qualitative meaning can be assigned to quantitative scores (e.g., through knowledge around clinically meaningful differences in scores).

‘Usability’ - refers to more practical aspects of PROM use such as number of items, time taken to complete and score, readability, available translations and whether there are cost or legal implications such as copyright involved in using the tool.

Table 4 summarises each paper’s definition of clinical utility; those aspects referring to ‘relevance’ and those referring to ‘usability’ have been illustrated in coloured highlighting as per the key.

Article	Clinical Utility factors

Simpelaere et al. (2016) – Systematic review protocol	• Interpretability
	  ○ Mean
	  ○ Standard Deviation
	  ○ Minimal important change
	• Feasibility
	  ○ Time to administer
	  ○ Mode of administration (interview vs. questionnaire)
	  ○ Facilitation required to complete

de Klerk et al. (2018) – Systematic review	• Respondent burden and presentation
	  ○ Time taken to complete
	  ○ Literacy level
	  ○ Availability to the public
	• Appropriateness
	  ○ Importance of the measure to clinical decision making
	  ○ Does the measure impact on treatment processes?
	• Acceptability
	  ○ Acceptability of the measure to client/carers/family

Vickers and Chen (2017) – Opinion piece based on clinical experience	• Minimize patient burden ≤15-20 items • Use simplified language with no jargon • Are items clinically appropriate • Are there potentially issues with tool use with patient subgroups
Galantino et al. (2015) – Systematic review	• Equipment needed (e.g., pen) • Cost • 'Ease of use' • 'Ease of scoring' • Normative data available

Thompson et al. (2016) – Literature review	• Patient and clinician burden
	• Feasibility and acceptance
	• Resources entailed in use
	  ○ Format (paper vs. electronic)
	  ○ Time
	  ○ Frequency
	  ○ Cost
	  ○ Effort to administer/score/store/analyze
	• Intellectual property and copyright issues
	• Clinical significance of scores

Brasil et al. (2018) – Systematic review protocol	• Interpretability
	  ○ Can qualitative meaning be assigned to quantitative scores
	  ▪ Is data available on standard error of measurement (SEM) and minimal important change (MIC)
	• Feasibility
	  ○ Ease of use
	  ○ Form of administration (interview vs. questionnaire)
	  ○ Duration
	  ○ Patient burden

Jones et al. (2020) – Systematic review	• Brief time of completion (≤3min)
	• Free cost
	• Coverage of four minimum QoL domains according to WHO classification
	  ○ Physical health
	  ○ Psychological
	  ○ Social relations
	  ○ Level of independence

Stewart et al. (2018) – Systematic review	• Number of items/subscales
	  ○ Item scaling
	  ○ Scores
	• Target population and purpose
	  ○ Number of participants in development/refinement
	  ○ Source of participant recruitment appropriate to intended clinical use

Lam et al. (2020) – Literature review	• Acceptability
	  ○ No of items
	  ○ Time to complete
	  ○ Readability

	○ 'Comfort level concerns' – are questions likely to be distressing for pts • Feasibility ○ Ease of use (training/supervision required?) ○ Role of clinician (questions for clinician/recall period) ○ Time to score ○ Costs associated with use • Appropriateness ○ Intended patient populations ○ Global purpose
Aiyegbusi et al. (2020) – Qualitative research	• Administrative issues ○ Paper vs. online ○ Frequency of use per patient • Patient-related issues ○ Use of proxy for patients with special needs • Use of computers to ease completion • Scoring burden • Using Computer Adaptive Testing to reduce length of tool • Patient literacy levels • Translations into other languages
Howell et al. (2015) – Scoping literature review	• Length and complexity of scale • Availability of translated and culturally meaningful versions • Degree of disability of user (i.e., too ill to use, too well to bother) • Acceptability to patients – would they be willing to reuse the scale in the future • Potential for PROM use to be intrusive • Availability of longitudinal data on what signifies a clinically important difference • Relevance of items • Patient ability to use response format (e.g., computer)
Nic Giolla Easpaig et al. (2020) – Qualitative research	• Feasibility ○ Is the tool easy to navigate? ○ Do the patients have the necessary 'knowledge and skills' and computer access required • Workflow ○ Whose responsibility is it to use the tool ○ Time requirement ○ Potential for positive 'focusing' impact of tool use • Acceptability ○ Is the tool relevant ○ Are there benefits to using the tool (e.g., improving care and experience for patients)
Turner et al. (2020)– Qualitative research	• Time taken to carry out tool • Knowledge around tool ○ Reliability of data and potential for patients' responses to be influenced ○ Clinical value vs. management-led 'tick box exercise' • Integration into clinical systems (e.g., into computer notes/systems)
Montgomery et al. (2020)– Qualitative research	• Conceptual characteristics ○ Number of items ○ Type of scale ○ Recall timeframe • Scoring ○ Domain, item and/or global scores • Time to complete the measure • Use of plain language • Available translations • Licence fees for use
Higginson and Carr (2001) – Opinion piece based on clinical experience	• Who completes the measure – patients/family/professional – and will they complete it? • How long does the measure take to complete? • Do staff and patients find it easy to use • Who will need to be trained and informed about the measure? • Can results from the measure be interpreted clinically and are they relevant?

Patel et al. (2017) – Systematic review	• Scoring and interpretation
	  ○ Is there documentation on how to score?
	  ○ Is there a plan for managing/interpreting missing responses?
	  ○ Is information provided on how to interpret scores (e.g., what high and low scores represent)
	• Burden and presentation
	  ○ Is time to complete/number of questions reasonable
	  ○ Is there a description of literacy level?
	  ○ Is the entire PRO publicly available

Kroenke et al. (2015)- Opinion piece based on clinical experience	• Actionability
	  ○ Do scores drive clinical decision making?
	• Appropriateness to the relevant clinical setting
	• Universality (i.e., for severity assessment)
	• Self-administration
	• Item features
	  ○ Number of items
	  ○ Bundling of issues
	• Response options
	  ○ Option number and dimensions
	  ○ Timeframe
	  ○ Uniform vs. varying options
	  ○ Intervals between options
	• Scoring
	  ○ Simplicity
	  ○ Interpretability
	• Accessibility
	  ○ Nonproprietary
	  ○ Downloadable?
	  ○ Available in different languages/for vulnerable groups
	  ○ Incorporated into electronic health records

Key: ….. = RELEVANCE ….. = USABILITY

Table 4: Summary of analysis of reviewed papers

The above summary table illustrates the complexity of the concept of clinical utility, and the lack of consistency within the literature as to what it comprises. In addition, this table facilitates synthesis of the information from the literature and provides evidence for the applicability of the 'relevance' and 'usability' elements that are the two key themes.

On synthesising the above information, it is clear that for this thesis, when considering clinical utility of the MDADI, the key elements of 'relevance' and 'usability' should be considered. Specifically, I propose that these two domains comprise the elements presented in Table 5, which have been condensed from the preceding analytical review of the literature and will form the basis of consideration when analysing the clinical utility of the MDADI in this study.

RELEVANCE	USABILITY
R1 Does the tool impact on clinical decision making? **R2** Are there potentially issues with tool use with patient subgroups? **R3** Is normative data available? **R4** Is data available on standard error of measurement (SEM) and minimal important change (MIC)? **R5** Are questions likely to be distressing for patients? **R6** Is it possible to use a proxy for patients with special needs? **R7** Is the tool acceptable to patients – would they be willing to reuse the scale in the future? **R8** Are the items relevant? **R9** Is the recall timeframe appropriate?	**U1** Does the tool have an appropriate literacy level? **U2** Are there any intellectual property issues? **U3** Is there a cost involved with using the tool? **U4** Is the time taken to administer acceptable? **U5** Is the scoring process straightforward? **U6** Is the time taken to score acceptable? **U7** Are translations into other languages available? **U8** Is there a plan for managing/interpreting missing responses?

Table 5: Summary of clinical utility factors

How these data will be used to explore the clinical utility of the MDADI will be described in Chapter Three of this thesis.

Thus far this review has discussed the literature on the psychometric assessment of PROMs and factors encompassing their clinical utility. The next section of this literature review will focus on an exploration of the current 'state of the art' of published knowledge about the MDADI, in terms of both its psychometric properties and clinical utility.

2.4 Analytical review of the literature concerning the MDADI

Medline, CINAHL, PubMed and EMBASE databases were accessed to perform an analytical review of the literature concerning the MDADI and to ascertain the current evidence base surrounding the MDADI's psychometric properties and clinical utility.

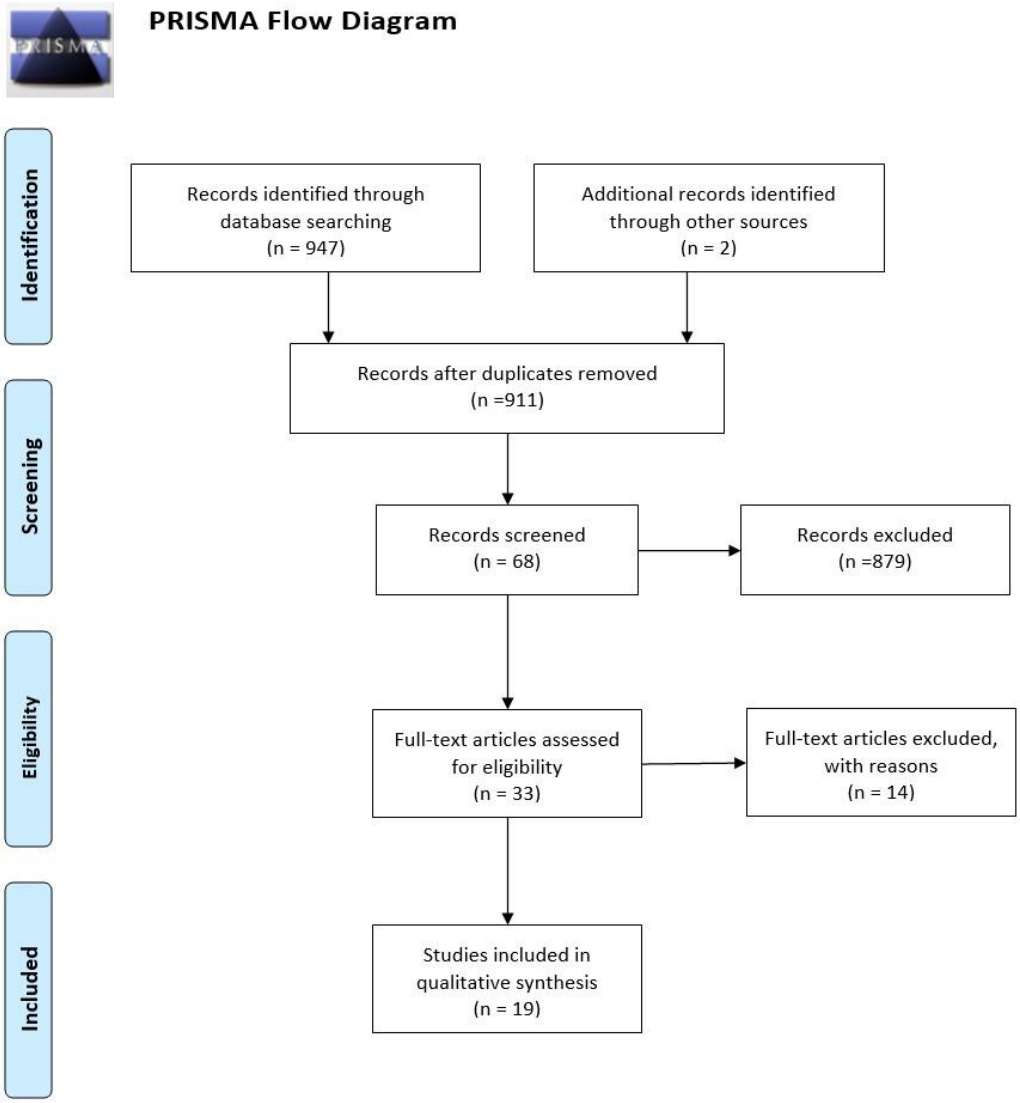
The full database search strategy is provided in Appendix C. Abstracts were reviewed and the inclusion and exclusion criteria as listed in Table 6 were applied.

Inclusion criteria
Focus on psychometric properties of the English language versions of the MDADI in use with people with HNC Focus on MDADI development Focus on clinical utility of the MDADI Focus on cross-cultural adaptation of the MDADI Papers in English language Papers in peer-reviewed journals
Exclusion criteria
Papers where MDADI is used as an outcome tool in clinical research rather than focussing on the MDADI itself Grey literature Unpublished literature

Table 6: Inclusion and exclusion criteria for MDADI literature search

Figure 2 shows a PRISMA diagram (Moher et al., 2009) illustrating the search process.

Figure 2: MDADI literature search PRISMA diagram



The search ultimately yielded 19 articles meeting the inclusion criteria, which, apart from Chen's MDADI origin paper, can be grouped into three themes:

1. Papers exploring psychometric properties of the MDADI
2. Papers documenting cross-cultural adaptations of the tool
3. Papers exploring other aspects of the MDADI

Table 7 summarises the theme and specific focus of each paper.

Theme	Papers	Specific focus
Origin paper	Chen et al. (2001)	MDADI origin paper
Psychometric properties of the MDADI	Timmerman et al. (2014)	A systematic review of several swallow-related PROMs including the MDADI
	Pedersen et al. (2016)	Concurrent validity of MDADI cf PSS NoD, PAS and WST
	Patel et al. (2017)	A systematic review of several swallow-related PROMs including the MDADI
	Khan et al. (2015)	Concurrent validity of MDADI cf PSS NoD using correlation
	Ojo et al. (2012)	A systematic review of several swallow-related PROMs including the MDADI
Cross-cultural adaptations	Matsuda et al. (2018)	Japanese version of MDADI
	Speyer et al. (2011)	Dutch version of MDADI
	Montes-Jovellar et al. (2019)	Spanish version of MDADI
	Zhang et al. (2017)	Chinese version of MDADI
	Guedes et al. (2013)	Portuguese version of MDADI
	Hajdú et al. (2017)	Danish version of MDADI
	Lechien et al. (2020)	French version of MDADI
	Yee et al. (2020)	Chinese version of MDADI
	Carlsson et al. (2012)	Swedish version of MDADI
	Kwon et al. (2013)	Korean version of the MDADI
Other properties of the MDADI	Hutcheson et al. (2016)	What constitutes a clinically relevant difference in MDADI scores?
	Lin et al. (2022)	Can the MDADI be shortened?
	Zraick et al. (2012)	What is the MDADI's readability?

Table 7: Summary of MDADI literature Key: PSS NoD = Performance Status Scale – Normalcy of Diet; WST = Water Swallow Test; PAS = Penetration-Aspiration Scale

2.4.1 Current knowledge base on the psychometric properties of the MDADI

Chang et al. (2019) emphasise the importance of continued evaluation of PROMs to ensure their validity. The literature search performed produced five studies that have explored the psychometric qualities of the MDADI (Ojo et al., 2012, Timmerman et al., 2014, Patel et al., 2017, Khan et al., 2015, Pedersen et al., 2016).Three of these papers considered multiple psychometric domains and two focussed on the concurrent validity of the MDADI.

Concurrent validity

Two papers specifically explore the concurrent validity of the MDADI: Khan et al. (2015) and Pedersen et al. (2016). Results are summarised in Table 8.

Paper	Tested against	Comments
Khan et al. (2015)	PSS Normalcy of Diet	MDADI -G mean Spearman rho = 0.45 over 3 timepoints MDADI – C Pre-treatment = 0.428 3m = 0.454 6m = 0.551 12m = 0.680
Pedersen et al. (2016)	PSS Normalcy of Diet Penetration-Aspiration Scale Water Swallow Test	Spearman rho = 0.68 (MDADI – C) Spearman rho = 0.34 (MDADI – C) Spearman rho = 0.45 (MDADI – C)

Table 8: Papers exploring MDADI concurrent validity

Khan et al. (2015) compared data from the MDADI with the Performance Status Scale (PSS), a clinician rated tool scoring a patient on the textures of diet they can swallow (List et al., 1996) for 114 patients. As per Table 8 above, MDADI-G scores had lower correlations than MDADI-C scores with PSS scores at pre-treatment, 3-month, 6-month and 12-month post treatment time points. This is similar to the findings of Chen et al. (2001) in their original piece of work on developing the MDADI, who also explored the concurrent validity of the tool against the PSS when developing the tool, finding a mean Spearman rho value of 0.53 for the MDADI-G. However, in Khan's study, although multiple time points were considered, there was a significant variation in amount of data analysed for each point with data attrition over time (pre-treatment: 86 MDADIs, at 3 months: 43 MDADIs, at 6 months: 39 MDADIs and 49 MDADIs at 12 months) meaning the power of individual time point calculations is reduced.

Further data on MDADI concurrent validity were published by Pedersen et al. (2016). Their data considered one time-point only (3 months post treatment) but compared MDADI-C results with the Penetration-Aspiration scale (Rosenbek et al., 1996), an instrumental swallow rating of airway penetration; the Water Swallow Test (WST - a test of swallow efficiency on drinking 100mls of water) and the PSS Normalcy of Diet subscale. On a sample of 173 patients, they found Spearman's correlation coefficients between the MDADI and PAS, WST and PSS to be 0.34, 0.45 and 0.68, respectively. Although the sample size was larger than the Chen study, the correlation coefficients produced show a strong correlation between the MDADI and diet texture restriction, but only a weak or moderate correlation with clinical measures of swallowing (PAS and WST). A potential weakness of this study is that data analysed for assessment were taken at only one time point post treatment (3 months) meaning the analysis was not as comprehensive as that of the Khan paper; however, their

patient numbers at this timepoint were much higher (173 for Pedersen compared with 43 for Khan), suggesting Pedersen's analysis may be more powerful.

The findings of these studies suggest that the concurrent validity of the MDADI differs at different time points in a patient's treatment journey. It also suggests that the concurrent validity for the MDADI is at best moderate, with it not strongly predicting performance on other swallow-related scales. This mismatch in MDADI scores versus other dysphagia assessment scores is corroborated by work by Kendall et al. (2014) who found that MDADI-G and MDADI-physical subscale scores did not strongly correlate with swallow pathophysiology as assessed on modified barium swallow imaging. This is perhaps not surprising given that as a PROM it is looking at swallowing from a very different perspective than clinician-rated or instrumental assessment-based assessments, which can take the signs and symptoms of dysphagia out of the context of a patient's day-to-day life.

2.4.1.1 Global evaluations of MDADI's psychometric properties

Three papers present a more comprehensive analysis of the psychometric properties of the MDADI: Ojo et al. (2012), Patel et al. (2017) and Timmerman et al. (2014). However, these papers do not focus solely on the MDADI tool, but rather include an analysis of it alongside other dysphagia outcome measures. Ojo et al. (2012) used the SAC-MOT guidelines devised by Aaronson et al. (2002) to systematically review HNC QoL instruments including the MDADI. Patel et al. (2017) carried out a systematic review of PROMs in dysphagia using the criteria devised by Francis et al. (2016) to assess instrument development and validation. Timmerman et al. (2014) used a rating scale devised by Terwee et al. (2007) and Timmerman et al. (2007) in their review of the psychometric characteristics of HRQoL questionnaires in oropharyngeal dysphagia.

Ojo and Timmerman both used only two raters, whilst Patel used three people to rate the properties of the tools in question. These results were not checked with a large group potentially reducing the validity of the findings given the small number of people involved in the process.

	Ojo et al. (2012)		Timmerman et al. (2014)		Patel et al. (2017)	
Psychometric domain	**Tested**	**Score**	**Tested**	**Score**	**Tested**	**Score**
Content validity	√	Blank	√	Indeterminate	√	√
Criterion validity	√	√	√	X		
Construct validity	√	√	√	√	√	Unclear
Test-retest reliability	√	√			√	√
Reproducibility – agreement			√	√		
Reproducibility - Reliability			√	No information available		
Description of format	√	Blank				
Interpretability	√	√	√	Indeterminate	√	√ but no plan for missing data
Structure analysis	√	x				
Responsiveness	√	√	√	N/A		
Cross-cultural adaptation	√	√				
Internal consistency			√	Indeterminate		
Floor/ceiling effects			√	X		
Conceptual model					√	√
Burden & presentation					√	√ but no literacy level assessment

Table 9: Psychometric assessment of the MDADI

Table 9 assesses the profile of each paper's psychometric analysis against the SAC-MOT guidelines used by Ojo etal. As illustrated above, these three reviews all assess slightly different profiles of MDADI properties. It is also evident that their assessments of the MDADI do not agree in some domains. For example, in the domain of content validity, not only does each paper define this parameter slightly differently, but all three papers rate the criterion differently despite all making their assessment based on the same single MDADI origin paper. These discrepancies are illustrated below in Table 10:

Paper	Ojo et al. (2012)	Timmerman et al. (2014)	Patel et al. (2017)
Content validity definition	"how appropriate the items are in measuring the construct of interest" p.4	"The concepts of interest are comprehensively represented by the items in the questionnaire". p.186	"Evidence that a PRO measures domain(s)... appropriate for its intended use" p.4
Content validity rating criteria	No criteria provided	"Clear description provided of: -measurement aim -target population -measured concepts -item selection and reduction AND target population involved in item selection" p.186	"Is there evidence that members of the intended respondent population were involved in the PRO measures' development? Is there evidence that content experts were involved in the PRO measure's development? Is there a description of the methodology by which items/questions were determined (e.g., focus groups, interviews)?" p4
Content validity rating of MDADI	No rating given	"indeterminate … information about the process of item selection and reduction was absent" p.193	Scored as present for - patient devised items - content experts involved - description of item development

Table 10: Assessment of MDADI content validity

As demonstrated in the above table, Ojo et al. (2012) do not report details of content validity, Timmerman et al. (2014) report analysis results as 'indeterminate' whereas Patel et al. (2017) consider this criterion to be fulfilled.

Ojo et al. (2012) do not consider clinical utility of the tool such as administrative and respondent burden. Likewise Timmerman et al. (2014) did not consider utility data but suggest this should be combined with psychometric data in future work to strengthen the case for tool selection. These authors also found data to be lacking, absent or indeterminate for the MDADI across the domains of content validity, criterion validity, interpretability, responsiveness, internal consistency, reproducibility/reliability and floor/ceiling effects.

This analysis indicates that there are many facets of the MDADI that have not yet been comprehensively analysed or considered in terms of validity, reliability and clinical utility, and given the tool's high profile and use as a primary endpoint in large multicentre trials, there is a pressing need to amend this.

Cross-cultural adaptations

As demonstrated in Table 9, only Ojo etal considered cross-cultural adaptations of the MDADI, by assessing whether translated versions of the tool are available. However, the only detail the paper provides is that translations exist, not how many or into which languages. The literature search performed for this review on the contrary shows that, as

demonstrated in Table 7, at the time of this literature search ten papers have been published exploring the properties of translated versions of the MDADI in a variety of languages, meaning that the tool is available to use in many countries worldwide. As the current study is focussing on the English-language version of the MDADI, papers considering alternative language versions of the tool have not been included in the analysis for this review.

2.4.1.2 Research exploring other properties of the MDADI

Following the initial publication of the MDADI in 2001 by Chen's team at the MD Anderson Cancer Centre (Chen et al., 2001), papers by Hutcheson, Lin and Zraick have explored different aspects of the tool following its adoption into mainstream HNC research and practice.

Hutcheson et al. (2016) focus on defining what a 'Clinically Relevant Difference' in composite MDADI scores constitutes with a retrospective cross-sectional study of 1136 HNC patients. MDADI scores were compared with 'clinical anchors' of feeding tube status, diet level and aspiration status as determined by Modified Barium Swallow instrumental assessment. Statistical analysis showed that an average difference of 10 points in MDADI-C scores differentiated between patients who were or were not feeding tube dependent and aspirators, and between distinct diet levels as measured on the PSS-HN diet scale. Although data from a statistically powerful number of patients was analysed, the analysis did not consider longitudinal within-patient score changes and therefore what constitutes a meaningful difference in MDADI-C scores for individual patients, which would be very relevant and useful to clinical practice; the authors suggest this may be a difference of 20 points but do not provide data to support this. In addition, to date the results of this study have not been triangulated with qualitative data from clinicians or patients to confirm its clinical relevance. Therefore, at this time further evidence is required to increase the validity of the findings of this study.

Lin et al. (2022) explore whether a multivariate factor analysis could reduce item redundancy thereby generating a shortened version of the 20-item MDADI tool. Their analysis used data from 196 questionnaires in total, completed by patients treated at two hospitals in the UK, at pre-treatment, 3-month post-treatment and 12-month post-treatment time points. The authors present preliminary findings that the tool could be reduced to a 5-item 'MiniDADI'. They propose this would have improved utility in clinical settings as the tool would be shorter and quicker to use with items found to be statistically redundant having been eliminated. The authors do however emphasize that further validation of the MiniDADI would be required prior to its adoption in clinical practice, due to the geographically limited nature of their patient data and incomplete missing data analysis, and that test-retest reliability and

concurrent validity of the tool have yet to be assessed. Work assessing its clinical utility has also yet to be undertaken.

Zraick et al. (2012) assess the readability of the MDADI and relate this to average reading levels of English-speaking adults living in the United States. The authors used a software-based analysis, using three readability formulae, used to analyse an electronic version of each PROM in question, with an emphasis on the Flesch Reading Ease (FRE) formula as the most widely used. The FRE formula generates a score from 0 to 100, where higher scores indicate easier reading material; the authors found the MDADI score to be 46, which equates to 'college level', i.e., a high level of literacy. This means that there is potential for the MDADI to be too linguistically complex for patients with lower literacy levels, thereby affecting these patients' ability to complete the tool and the validity and reliability of the data it generates. This has relevance given that evidence exists in the literature that HNC patients have complex health literacy needs (Beitler et al., 2010, Jabbour et al., 2017). Of note is that Zraick's study analysed four different dysphagia PROMs, and the MDADI was considered the most difficult PROM to read across all the calculations performed. The authors do however acknowledge that software-based literacy calculations do not provide a complete view of reading comprehension, and that a mathematical formula cannot take into account contextual factors such as individual motivation and interest, thereby potentially reducing the validity of the results. No other published study has addressed this to date with reference to the MDADI and there has been no analysis of the reading level of the MDADI with respect to UK patients.

2.4.2 Synthesis of current MDADI knowledge

The above assessment of the existing literature on the MDADI shows that whilst many aspects of its psychometric properties have been quantitatively investigated, there remain significant gaps, such as in analysis of its content validity, floor/ceiling effects, and also inconsistencies in terms of approach used to assess the MDADI's psychometric properties. In addition, there has also yet to be any qualitative research undertaken exploring the experience of use of the tool in practice or considering its clinical utility or content validity.

Gaps in knowledge of MDADI's psychometric properties

Although it has been used to assess other cancer related PROMs (Peng et al., 2020), the COSMIN tool has not yet been used as a basis for the psychometric evaluation of the MDADI. This is despite the fact that COSMIN is the most widely used and comprehensive PROM assessment tool available at this time (Lorente et al., 2020) notwithstanding Baijens et al. (2021)'s suggestion that this tool would be appropriate for this purpose. As this is the case, this thesis will be the first to carry out an evaluation of the MDADI using the COSMIN

tool. It is also of note that little detail exists of the content validity of the MDADI; this has not been re-evaluated or revalidated since the tool's inception.

This literature review also highlights that to date the psychometric properties of the MDADI have only been statistically assessed with the approach of CTT. However as discussed earlier in this chapter, the field of IRT has much to contribute to PROM tool psychometric assessment and indeed other dysphagia assessment tools have been analysed using IRT which has provided useful data on their strengths and weaknesses. Kean et al. (2018) suggest that IRT methods are particularly suited to 'multi-item survey questionnaires' as these tools use several items to explore different aspects of complex constructs. The MDADI as a 20-item tool exploring the complex construct of dysphagia-related QoL therefore meets this criterion and IRT is appropriate to use in assessment of the tool's psychometric properties. Data on IRT analysis of the MDADI is missing from the evidence base to date. Therefore, this thesis will use IRT to fill gaps in the existing psychometric profile of MDADI – rather than carrying out a fully psychometric revalidation of the tool, for many aspects of which CTT data currently exist.

People with HNC form a heterogeneous group. Patients vary in terms of cancer site, cancer stage, treatment modality, time post treatment, requirement for prophylactic feeding tubes, dental extractions, age and previous medical history. These variables also have an impact on expected functional outcomes and patient reported HRQoL (Rogers et al., 2015a, Hunter et al., 2013, Chang et al., 2019), and could all influence relationships between swallowing outcome measures and their validity for use with certain patient subgroups. There may also be potential for responses to be impacted by non-swallow factors such as age and gender (Goepfert et al., 2017, Hunter et al., 2013). This effect is known as Differential Item Functioning (DIF) (Teresi and Fleishman, 2007). Nguyen et al. (2014) define DIF as "measurement bias that occurs when individuals from different groups respond differently to an item even after controlling for the underlying trait level" (p.25). DIF has previously been shown to be an issue in other dysphagia assessment tools such as the EAT-10 (Cordier et al., 2017) however to date there has been no analysis of whether these factors may influence MDADI scores.

Gaps in knowledge of MDADI's clinical utility

As previously discussed in this literature review, the clinical utility of a tool is a complex concept measured through aspects such as accessibility, appropriateness, acceptability and practicability (Smart, 2006). It encompasses properties not assessable with statistical analysis, such as how easy the items are to understand, and the relevance of the content (Mehanna and Morton, 2006).

In the papers reviewed here considering aspects of the clinical utility of the MDADI as part of their overall evaluations of the tool, Timmerman et al. (2014) suggest wording and question order should be evaluated, as well as the relative importance of each of the domains. Patel et al. (2017) highlight a lack of literacy level assessment, plan for missing data, and longitudinal validity, as well as identifying a lack of patient involvement in tool development and a lack of justified dimensionality. Patel's is the only group to date to consider practical aspects of tool burden and presentation, that is time completed to score and tool accessibility. Ojo et al. (2012) stated they chose not to assess 'administrative and respondent burden' as this was too complex a task.

This literature review demonstrates that there has not been any published qualitative research to date on use of the MDADI tool from either a clinician or patient perspective, with a focus on its clinical utility.

Using the 'relevance' and 'usability' domains synthesised from the literature earlier in this review (see Table 5), Table 11 below shows the areas of MDADI clinical utility that have been investigated to date, and what this current study has the potential to contribute.

RELEVANCE domain	Comment	What this study can add
R1 Does the tool impact on clinical decision making?	Tuomi et al. (2020) suggest that a cutoff score <60 is useful to identify patients in need of swallow intervention	Qualitative data from clinicians on their use of the tool in practice
R2 Are there potentially issues with tool use with patient subgroups?	No data to date	Qualitative data from clinicians on their use of the tool in practice
R3 Is normative data available?	Not relevant as tool not designed for non-dysphagic population	n/a
R4 Is data available on standard error of measurement (SEM) and minimal important change (MIC)?	Hutcheson et al. (2016) propose a difference of 10 points on the MDADI-C to be an MIC.	n/a
R5 Are questions likely to be distressing for patients?	No data to date	Qualitative data from clinicians on their use of the tool in practice
R6 Is it possible to use a proxy for patients with special needs?	No data to date	Qualitative data from clinicians on their use of the tool in practice
R7 Is the tool acceptable to patients – would they be willing to reuse the scale in the future?	No data to date	Qualitative data from clinicians on their use of the tool in practice
R8 Are the items relevant?	Explored statistically by Lin et al. (2022) with factor analysis but with no qualitative component	Alternative statistical approach to potential tool shortening (IRT/DIF) Qualitative exploration of item relevance with clinicians Results of statistical analysis in this study can be compared with Lin etal's findings
R9 Is the recall timeframe appropriate?	No data to date	Qualitative data from clinicians on their use of the tool in practice
USABILITY domain	**Comment**	**What this study will add**
U1 Does the tool have an appropriate literacy level?	Only published work is looking at US English (Zraick et al., 2012)	Clinicians will have the opportunity to comment
U2 Are there any intellectual property issues?	No copyright limitations stated in origin paper	n/a
U3 Is there a cost involved with using the tool?	No cost	n/a
U4 Is the time taken to administer acceptable?	No data to date	Qualitative data from clinicians on their use of the tool in practice
U5 Is the scoring process straightforward?	No data to date	Qualitative data from clinicians on their use of the tool in practice
U6 Is the time taken to score acceptable?	No data to date	Qualitative data from clinicians on their use of the tool in practice
U7 Are translations into other languages available?	See literature review section on cross-cultural validity	Results of MDADI literature review performed for this study

U8 Is there a plan for managing/interpreting missing responses?	No plan given in origin paper	n/a

Table 11: Mapping existing MDADI knowledge to clinical utility domains

Given its length and item complexity, use of the MDADI may be challenging in terms of time to complete, wording of items for patient comprehension, time taken to score (i.e., its clinical utility). Aspects of Relevance and Usability will be explored in this study using a methodology which will be discussed in the next chapter.

2.5 Summary of literature review findings

The European Society for Swallowing Disorders (ESSD) recently published a white paper comprising recommendations for future research and clinical practice in 'screening and non-instrumental assessment for dysphagia in adults' (Speyer et al., 2022). A key recommendation of that paper was that clinicians and researchers should:

> "discontinue the use of measures with insufficient or poor psychometric properties. Instead, use measures that demonstrate robust psychometric properties that meet psychometric quality and feasibility criteria." (p. 12).

The following table summarises the key findings of the literature review and therefore the gaps in knowledge with regards to PROMs quality assessment and the MDADI.

Subject	Current knowledge	Current gaps in knowledge
PROMs	PROM use is a valuable part of clinical assessment in HNC	No guidance currently exists on which PROM tools should be used in HNC practice or specifically in dysphagia assessment
	Psychometric quality of PROM tools should be assessed and considered; different statistical approaches exist	No definitive guidance exists on which approach to use
	The clinical utility of PROM tools is an essential factor to be considered when looking at overall quality	No uniform definition exists of what clinical utility constitutes or how to assess it
MDADI	The psychometric properties of the MDADI have been investigated using a CTT approach	No published work exists to date on IRT analysis of the MDADI
		No published work exists to date on analysis of the content validity or clinical utility of the MDADI

Table 12: Summary of literature review findings

The analysis in this literature review demonstrates a lack in the quality and comprehensiveness of MDADI psychometric and utility assessment to date. This analysis, together with my clinical experience of MDADI use, has provided the evidence and impetus to carry out this study. The MDADI is an important and useful tool: results of major clinical trials that use the MDADI as a primary or secondary endpoint have the potential to make a

significant difference to the standard of care for HNC oncological and surgical treatment (PATHOS, CompARE, DARS) (Petkar et al., 2016, Mehanna et al., 2017, Owadally et al., 2015) as well as dysphagia therapy intervention (PRO-ACTIVE) (Martino et al., 2021). For the results of these trials to be valid and reliable, it is essential that the outcomes tools they use are also valid and reliable – and this includes the MDADI. In addition, as the MDADI is often used to facilitate the appraisal of psychometric properties of other dysphagia assessment tools, it is again imperative that the MDADI's own psychometric profile is strong.

However, the MDADI has not yet been validated on a UK population, its psychometric profile has not been analysed using IRT or the COSMIN tool, complexities of tool use in terms of DIF with this heterogeneous patient group have yet to be explored, and its clinical utility is yet to be comprehensively assessed. Further research is required to investigate and correct these omissions. In conclusion, the research evidence supports the need for re-exploring the psychometric properties of the MDADI and potential for DIF using an IRT statistical approach, as well as assessing its clinical utility.

Kroenke et al. (2015) describe how 'second generation questions' around PROM tools properties can arise following their development and a period of use in research and clinical practice. This literature review has endeavoured to highlight what some of these questions in respect to the MDADI might be, and the current study will attempt to answer these questions in addition to potentially identifying new, as yet unidentified issues with the tool with a view to providing direction for strengthening and enhancing it for future clinical and research use. Assessment of the psychometric and pragmatic properties of an existing PROM tool can inform its use in clinical and research practice and suggest ways a tool might evolve to better fulfil its role. The assessment of the MDADI presented in this thesis will contribute to the evidence base regarding both its psychometric properties and feasibility, in keeping with the ESSD's stated goals.

Snyder et al. (2007) suggest that "modifications may occur as part of the natural evolution of instruments" (p. 581). This study aims to provide suggestions as to what the 'natural evolution' of the MDADI might incorporate. In the next chapter of this thesis, research methods to generate this data will be described.

CHAPTER 3: METHODS

3.1 Introduction

This chapter will describe the methodology and research design of this study. Research questions and study aims will be stated and the structure and process of the study designed to answer these questions and achieve these aims will be described.

3.2 Research aim, questions and outcome

The aim of this study is to explore the psychometric properties and clinical utility of the MD Anderson Dysphagia Inventory (MDADI) PROM. As shown in the preceding literature review, to date there has been no qualitative exploration of the content validity or clinical utility of the tool, and psychometric analysis of the tool using the statistical approach of Item Response Theory (IRT) has not yet been undertaken. This study's research aim and questions are therefore as follows:

3.2.1 Research Aim

To analyse and evaluate the MD Anderson Dysphagia Inventory on a UK-based population.

3.2.2 Research Questions

1. What is the content validity and clinical utility of the MDADI from a UK SLT perspective?
2. What are the results of an exploration of psychometric properties of the MDADI on a UK population using Item Response Theory?
3. What are the potential factors which might result in differential item functioning?

3.2.3 Outcome

Qualitative and quantitative insight will make clear the strengths and weaknesses of the MDADI tool, both in terms of psychometric analysis and the utility of the tool, that is the experience of using the tool in practice. These data will be useful to inform future tool development and use in both clinical and research practice providing guidance on how the validity of a future version of the tool could be improved, and how its use in practice could be enhanced.

3.3 Methodological considerations

Research paradigms are systems of shared belief that influence both research questions asked and research methods (Moseholm and Fetters, 2017). It is important to consider and define the paradigm of research as it acts as the context for enquiry, providing a framework for how a researcher views the world and their work (Cooper and Meadows, 2015). Being cognisant of their preferred paradigm allows a researcher to acknowledge the specific lens

through which their work is filtered and the impact this may have on all aspects of their research.

The research questions in this thesis have been approached with a pragmatic philosophical worldview to allow for a more practical and problem focussed approach to research design and method choice than would a focus on specific positivist or interpretivist epistemological/ontological research paradigms.

Taking a solely positivist position would lend itself to a purely quantitative analysis of the MDADI focussed on statistical assessment of its properties. Conversely, taking a purely constructivist/interpretivist approach would focus the analysis on qualitative aspects focussing on building up a rich and dimensional picture of use of the tool (Leung, 2015).

A pragmatic philosophical approach however provides the opportunity to combine the best of both worlds and gives a more complete picture of the topic of investigation, transcending the "forced dichotomy of quantitative and qualitative methods and data" (Feilzer, 2009 p.9). Pragmatism is a problem centred approach which is suited to mixed methods studies as it gives freedom of choice in methods and techniques: selection is guided by the practicalities of addressing the research question (Creswell, 2014).

As an in-depth analysis, and given the research questions being asked, purely qualitative or quantitative data would have been insufficient to cover all aspects of assessing both psychometric and clinical utility strengths and weaknesses of the MDADI. Quantitative methods could not explore the content validity of the tool, or the clinical experience of using the tool and its utility in practice, whereas qualitative methods could not produce data on differential item functioning or the means to explore psychometric properties of the tool. A review of the PROM evaluation literature shows that a mixed methods approach is consistently applied for these reasons.

3.3.1 Mixed Methods Approach

Cappelleri et al. (2014) suggest that both qualitative and quantitative data are necessary when evaluating PROM tool validity. Magasi et al. (2012) considered the two types of data to be complementary and that both types of evidence are necessary to avoid a 'false dichotomisation' of methods when analysing a tool. These combined data encompass both mathematical analysis of tool strengths and weaknesses, and information on experience of tool use in practice. Therefore, to analyse and potentially enhance the MDADI, a pragmatic approach to the research questions, using mixed methods methodology, best fits this dual requirement (Bryman, 2016). This study therefore has analysed the tool both qualitatively, exploring what clinicians think about the validity and utility of the tool, and quantitatively,

reviewing the psychometric properties of the MDADI. Information on both aspects enables a comprehensive analysis of the strengths and weaknesses of the tool, to guide and inform its future use and development.

3.4 Study design

3.4.1 The SARS COV-2 Pandemic

The onset of the global COVID-19 pandemic in Spring 2020 and its effect on the UK healthcare system had a significant impact on the design of this study. Initially in addition to gathering qualitative data from clinicians about the content validity and clinical utility of the MDADI, people with HNC were to be recruited to generate data from the patient perspective. Face-to-face focus groups were originally planned to have taken place on a hospital site within NHS Lothian with groups of patients who had had a diagnosis of HNC. The focus groups would have involved group tasks using MDADI materials to study and discuss the tool in depth.

The pandemic had a significant impact on HNC service delivery within NHS Lothian. The weekly multidisciplinary HNC clinic, which was the planned recruitment location to access patients who would be approached to participate in the focus groups, was cancelled with immediate effect. In addition, MDT members who would have acted as recruiters were not available due to redeployment or clinical pressures. An alternative recruitment pathway that would have maintained ethical integrity was not available.

In addition to the loss of the planned recruitment pathway, holding face-to-face focus groups was no longer possible due to restrictions on group gatherings on NHS Lothian sites, and the health risks to participants from attending such events (Jarvis et al., 2020).

Although over the course of 2020 the infrastructure and technology were developed for NHS Lothian staff to carry out internet-based video consultations with patients, such an approach continued to be challenging with the HNC patient group, with SLT clinical experience during this time indicating many patients did not have access to compatible technology, or the computer literacy or confidence with which to engage with the technology. In addition to this, the planned focus group activities involving physical materials and group activities with these materials would not have been possible via the online format.

Given the uncertain trajectory of the pandemic, lack of ability to project when circumstances might return to normal, and the inherent time constraints on an academic study being carried out as part of a degree, the pragmatic decision was made to change the study design and amend the research questions to include clinicians only for the qualitative strand of the study.

3.4.2 Final study design

The COSMIN framework of psychometric domains and methodology for their assessment was used to inform and guide the quantitative and qualitative data analysis in this thesis. Table 13 below demonstrates which aspects of the COSMIN domains were explored in this thesis.

COSMIN Domain	Measurement property	Aspect of a measurement property	Being explored in this thesis?
Reliability			
	Internal consistency		Yes – quantitative strand
	Reliability		No
	Measurement error		Yes – quantitative strand
Validity			
	Content validity		Yes – qualitative strand
		Face validity	Yes – qualitative strand (part of content validity)
	Construct validity		
		Structural validity	Yes – quantitative strand
		Hypothesis testing	Yes – quantitative strand (part of structural validity)
		Cross-cultural validity	No
	Criterion validity		No
Responsiveness			
	Responsiveness		No
Interpretability			Yes – qualitative strand (includes aspects of content validity and clinical utility)

Table 13: COSMIN measurement properties being explored in this thesis

The methods for qualitative (Strand A) and quantitative (Strand B) elements of this project will now be outlined.

3.4.2.1 Research Strand A

Content validity and clinical utility of the MDADI tool are the two areas of focus of this strand of the study (Research Question 1). SLTs were surveyed on the content and administrative aspects of the MDADI tool. These SLTs were registered with HCPC, had experience of working with patients with HNC, and were practising within the UK.

Using an online survey, SLTs were asked about their opinions and experiences of using the MDADI with patients with HNC. The survey explored content validity and clinical utility of the MDADI tool, where clinical utility is defined as previously stated in Chapter Two: comprising the two aspects of ‘relevance’ and ‘usability’. As patients were excluded from the research design, the questionnaire also sought specific data from clinicians around comments patients have made to them on use of the tool.

Content validity is one of the aspects assessed by the COSMIN tool and is given increased weight in the latest COSMIN iteration (Terwee et al., 2018). Content validity is important to consider when assessing PROM tools, as it concerns whether the tool covers all the important and relevant aspects of the subject under investigation (Connell et al., 2018). It is crucial that tools are appropriate and suitable to the relevant client groups if they are to be used in clinical practice (Higginson and Carr, 2001). COSMIN suggests content validity of PROMs can be established by assessing the relevance, comprehensiveness and comprehensibility of the items (Terwee et al., 2018); however as established in Chapter Two of this thesis, no study to date has explicitly analysed content validity of the MDADI.

The COSMIN content validity criteria of relevance, comprehensiveness and comprehensibility are used as a framework in the discussion of the MDADI content validity in Chapter Seven of this thesis.

Clinical utility is not singled out as a specific measurement property by COSMIN; therefore the taxonomy developed and presented in Chapter 2 of this thesis will form the framework for clinical utility analysis and discussion.

In summary, Strand A used an online questionnaire with SLTs to examine the content validity and clinical utility of the MDADI.

3.4.2.2 Research Strand B

Strand B, the quantitative phase of the study, was designed to answer Research Questions 2 and 3. To enhance previously carried out CTT measures of validity and reliability available in the literature, complementary and augmentative analysis of the MDADI was carried out on NHS Lothian patient data using IRT guided by the COSMIN framework. This strand was also informed by data collection and analysis performed in Strand A. Section 3.8.2 of this Chapter includes detail of the quantitative analysis and the COSMIN parameters which were investigated in this study.

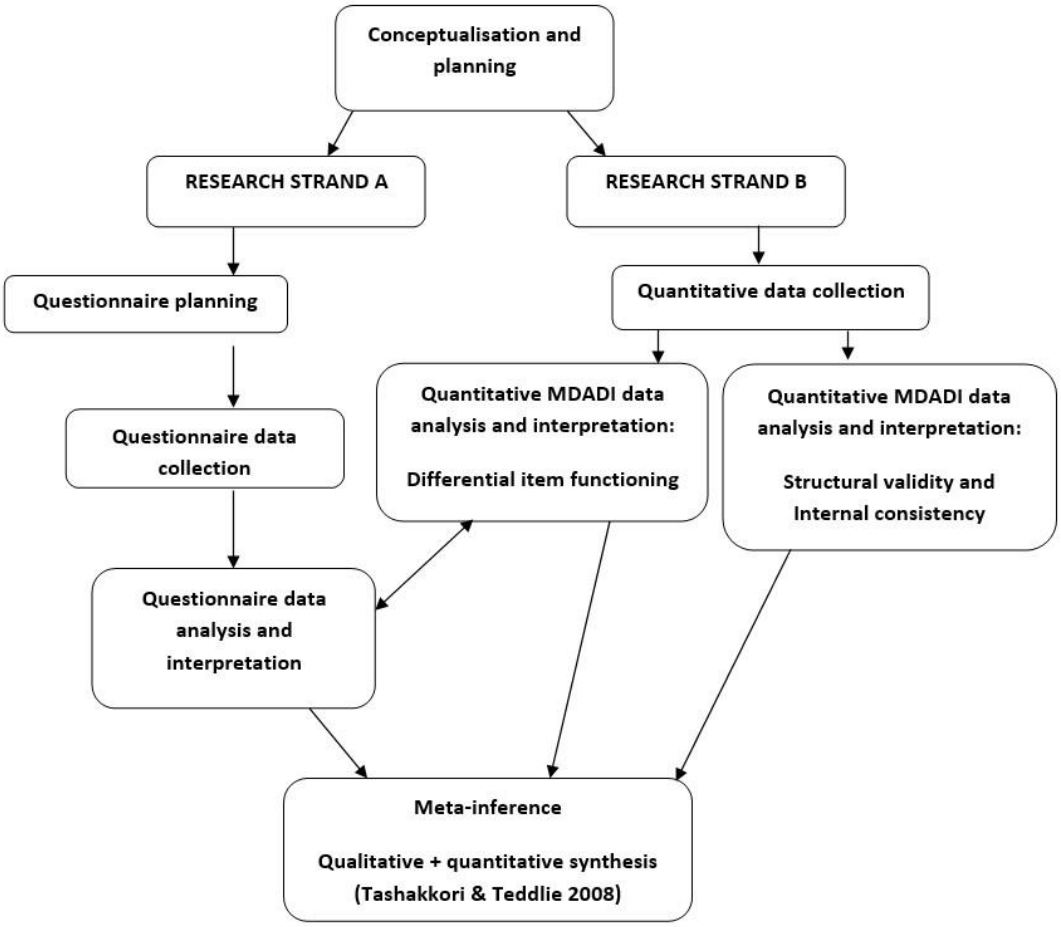
Research Strand B involved a quantitative statistical analysis of certain psychometric properties of the MDADI tool. The quantitative methodology of this study was driven predominantly by item response theory (IRT) rather than classical test theory (CTT). IRT is a theoretical basis for measurement (Embretson and Reise, 2013); a set of statistical models that allow description of the relationships between a person's 'trait' (i.e. swallow related QoL on the MDADI) and how they respond on the tool (Nguyen et al., 2014). Given the nature of the MDADI's use of Likert rating scales for item responses, IRT models for polytomous responses were used. IRT analysis provided item level data which complements the information and data from CTT validation (Jean-Pierre et al., 2014).

IRT attempts to explain the relationship between an underlying construct (e.g., dysphagia-related QoL) and individual item responses on a scale, and generates data visualisations of these relationships in the form of graphs. In the results of this thesis, various graphs are presented to illustrate aspects of the MDADI IRT analysis. Item Information Function (IIF) and Test Information Function (TIF) graphs illustrate the results of internal consistency and item specific properties of the tool. Generation of Item Characteristic Curves (ICCs) and the associated data allows exploration of whether different groups of patients perform differently on specific items in the MDADI (i.e. whether they display DIF), and whether therefore there is systematic bias in the MDADI tool in terms of how it performs (Smith et al., 2016). DIF represents a serious threat to test validity therefore is important to consider (Zanon et al., 2016). The DIF analysis will be driven by hypotheses generated from qualitative data generated by the questionnaire. This process of IRT/DIF analysis may also enable tool 'scaling' (Magasi et al., 2012), by identifying items to be targeted for removal or revision, potentially simplifying and strengthening the MDADI without adversely affecting tool content validity.

Rationale for mixed methods approach

In summary the final study design for this project was that of a mixed methods multistrand design that was sequential in nature, as the qualitative data collection and analysis preceded the quantitative data analysis (Teddlie and Tashakkori, 2009). This is illustrated in Figure 3.

Figure 3: Study design diagram



3.5 Data Collection

3.5.1 Research Strand A

The literature on PROMs and the MDADI, the COSMIN tool, and the experience of use of the MDADI tool in the principal researcher's clinical team drove questionnaire content described in Research Strand A below. The goal of the questionnaire was to gather data on clinicians' opinion of MDADI content (i.e., content validity) and their experience of using the tool in practice (i.e., clinical utility) at a whole-tool and an item-by-item level. In the absence of any patient generated data an additional goal was to include clinicians' insights into how patients found using the tool.

Questionnaire tool development

An online survey design tool (www.onlinesurveys.ac.uk) was used to design and host the survey as this is easily accessible, GDPR compliant, password protected and provides a link to the questionnaire which can be disseminated via email. An internet-based survey method allowed access to a wide range of SLTs to gather their opinion on the MDADI using open-ended questions (Gilbert and Stoneman, 2016) generating more data than interviews or focus groups as participant numbers were anticipated to be higher.

The number of SLTs working in the field of HNC dysphagia in the UK is relatively small, although exact numbers are not known, and they are geographically dispersed: another reason why interviews or focus groups would be logistically challenging. With an internet-based survey however clinicians can be accessed via professional networks regardless of their working location. A literature search shows that to date no questionnaire-based studies concerning clinicians' opinions of dysphagia outcome tools have been published, although this method has been used with patients (Mehanna and Morton, 2006). However, there is precedent for use of surveys to elicit data on general clinical practice in dysphagia (Krisciunas et al., 2012, Roe et al., 2012, Moloney and Walshe, 2019).

A questionnaire with a mix of fixed-choice and open questions with free-text response format was chosen as a suitable method to collect these data from as large a group of UK HNC SLT clinicians as possible with the aim of providing qualitative data on use of the tool in clinical practice. Fixed choice items provided basic demographic information and open questions on validity and utility allowed participants to give their opinions in free text.

Narrative, qualitative data from the free text items gave a picture of the content validity of the MDADI tool in everyday clinical practice. These data then helped to inform quantitative analysis, for example, providing examples of variables worth exploring for potential DIF (Patrick et al., 2011), as data included detail on patient variables outside of dysphagia itself that may impact on MDADI scores. The qualitative phase generated data that was valuable to inform the quantitative analysis; the nature of this data could not be predicted a priori. If the quantitative strand of the research were to occur first and following this the qualitative data provided new and unexpected insights, the quantitative data analysis would have required to be amended post hoc which is not in keeping with a scientific approach.

Greenhalgh (2014) suggests the goal of questionnaire design is clarity and ease of completion, therefore, length, layout, question wording and content were considered (Boynton, 2004, Kelley et al., 2003, Jones et al., 2013) as were criteria for designing and judging appropriateness of survey questions (Bowden et al., 2002, Sullivan and Artino, 2017).

Questionnaire format and item content

Questions on the MDADI were open-ended in nature to allow collection of qualitative data, with the response taking the format of a free-text comment box. The questionnaire was designed to collect data on individual MDADI items as well as the tool as a whole. Clinicians' experiences with use of the tool were elicited, considering overall validity, as well as clinical utility as defined in the literature review. An item asking for detail on comments made by patients to clinicians concerning the tool was also included.

Demographic data gathered on questionnaire participants was limited to ensure anonymity. Only information on UK country of work, number of years of dysphagia experience and number of years' experience in HNC was requested. These data allowed assessment of range of representativeness of clinicians accessed. The final questionnaire format is presented in Appendix D.

When developing a new questionnaire tool, Rickards et al. (2012) suggest initially conducting a literature review to establish whether a relevant survey with validity evidence is already extant. In this case, a review of the literature shows there is no other study published to date exploring clinicians' opinions of the MDADI, therefore questionnaire content was guided by my own experience of using the tool, and the published literature investigating PROM clinical utility as discussed in the Literature Review. The dual clinical utility aspects of 'relevance' and 'usability' as delineated in the Literature Review and Table 14 below were considered to establish which aspects could be feasibly assessed via survey format, rather than purely from published literature about the MDADI. In Table 14 items in bold are those deemed possible to investigate through survey questions, whereas items in italics are questions that could potentially be answered by the existing MDADI literature.

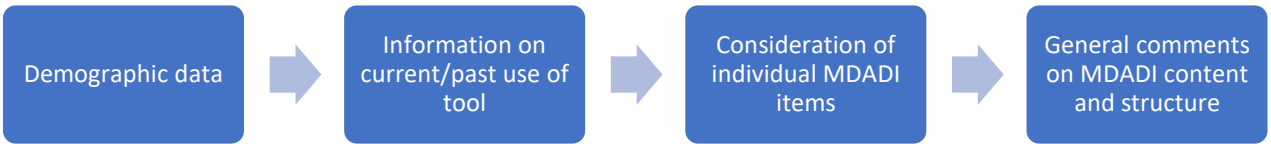
RELEVANCE	USABILITY
R1: Does the tool impact on clinical decision making? **R2: Are there potentially issues with tool use with patient subgroups?** *R3: Is normative data available?* *R4: Is data available on standard error of measurement (SEM) and minimal important change (MIC)?* **R5: Are questions likely to be distressing for patients?** **R6: Is it possible to use a proxy for patients with special needs?** **R7: Is the tool acceptable to patients – would they be willing to reuse the scale in the future?** **R8: Are the items relevant?** **R9: Is the recall timeframe appropriate?**	**U1: Does the tool have an appropriate literacy level?** *U2: Are there any intellectual property issues?* *U3: Is there a cost involved with using the tool?* **U4: Is the time taken to administer acceptable?** **U5: Is the scoring process straightforward?** **U6: Is the time taken to score acceptable?** *U7: Are translations into other languages available?* *U8: Is there a plan for managing/interpreting missing responses?*

Table 14: Aspects of clinical utility

Ultimately questions R1, R2 and R5-9, and U1, and U4-6 above were included in the questionnaire to gather data about clinician's perceptions of the clinical utility of the MDADI. Responses to R2 also fed into Research Strand B quantitative data analysis (Section 5.3) where potential for DIF was explored for the MDADI including issues raised by clinicians in the qualitative data collection phase.

The order of items in the questionnaire is summarised in Figure 4 below:

Figure 4: Order of questionnaire content



Following formulation of the survey content, it was piloted as outlined in Chapter Four of this thesis.

3.5.2 Research Strand B

NHS Lothian tertiary cancer centre SLT service swallow outcomes data were accessed to provide MDADI scores for the quantitative analysis. These data were collected between 2016-2021 at two points in patients' episodes of care (pre-and 6 months post-treatment). Existing MDADI data were sourced from the NHS Lothian SLT clinical notes and inputted into a Microsoft Excel™ spreadsheet. MDADI global score, composite score, subscale scores and individual item scores were included in the database to facilitate in-depth quantitative analysis. Patient demographic data: gender, SIMD via postcode, diagnosis and treatment, and age at time of assessment were also accessed via TRAK, the NHS Lothian patient information password protected IT system, and linked to the MDADI data.

3.6 Ethical and regulatory considerations

Research governance encompasses a range of standards and principles that ensure ethical, transparent and high quality research (NICE, 2018). Principles of research governance were adhered to in the planning and execution of this study as detailed below.

3.6.1 Regulatory review & compliance

To ensure the study design complied with ethical standards, the study protocol underwent a process of review by regulatory bodies prior to initiation of the research. The University of Stirling NHS, Invasive or Clinical Research ethics panel formally approved the protocol (Reference no. NICR 19/20 – 093) and the NHS Lothian Caldicott Guardian granted permission to access patient data from patient notes (Application no. CRD20098). The project was submitted via the NHS Integrated Research Application System (IRAS) for ethical approval and was assessed as being suitable for Proportionate Review. Ethical approval was granted by the Solihull Research Ethics committee on 11th December 2020 (IRAS no. 279483; REC no. 20/WM/0319). However, data collection did not commence immediately as there was a moratorium on non-COVID research within NHS Lothian at that time due to the ongoing pandemic and pressure on NHS infrastructure. This situation continued until April 2021 when the moratorium was lifted; permission to commence data collection was subsequently granted by NHS Lothian Research & Development in October 2021 (NHS Lothian R&D project number 2021/0039).

The University of Stirling acted as the Sponsor for this study in concordance with the 'Research Governance Framework for Health & Social Care Research' (HRA, 2020).

3.6.2 Patient & Public Involvement

Patient involvement in overseeing this study was actively sought, with the aim of providing a different perspective on the study and improving the overall quality and relevance of the research. Patients who had experienced HNC and treatment who had expressed an interest in opportunities to be involved in research were approached to form a study 'project advisory group'. Their expertise and knowledge was sought to help guide study design, process, analysis and dissemination in line with the National Standards for Public Involvement (NIHR, 2018) and the NIHR INVOLVE guidelines (INVOLVE, 2012). Specifically, the project advisory group was comprised of people with a previous diagnosis of HNC and experience of HNC-related dysphagia who were involved at the initial protocol development stage to inform study design and protocol. Correspondence with PPI group members confirmed the validity of the study; group members reported that further development of the MDADI would be of high relevance to patients. Whilst the PPI group were disappointed that the patient voice ultimately had to be excluded from the study protocol, they did appreciate the unavoidable nature of this change. The emotional impact of completing the MDADI was highlighted by PPI group members in their comments on the study protocol. In light of this, and the fact that patient data was not directly sought in the final study design, PPI group members were particularly keen for a question in the survey to be included around the potential emotional impact of the MDADI on patients completing the tool. A question specifically exploring this (question 35 in the questionnaire, see Appendix D) was therefore included in the online survey. Of note is also that the PPI group corroborated aspects of the clinical utility assessment taxonomy, confirming that they felt this was a worthwhile subject of analysis and reporting that they found the MDADI overly long and confusing.

3.6.3 Data protection and participant confidentiality

In accordance with the General Data Protection Regulation (GDPR) and the Data Protection Act (2018) the three main data safeguards were followed during this study. The minimum amount of data necessary was collected and stored, and all identifiable data were anonymised, for example by using pseudonyms. Individual patient data was not included in the study writeup or report and only aggregated statistical data were included. All data collected for this project was stored securely in keeping with Stirling University's data security policies. Electronic data files were stored on the University of Stirling 'X drive' which is a secure networked drive where the data is only accessible by the researcher, and permanently deleted on completion of the study. Prior to closure of the questionnaire, participants had the right to both withdraw their participation and to ask for their data not to

be quoted in the final study. This was feasible through correspondence with the chief investigator to identify their individual response number.

3.7 Study setting, sample and recruitment

3.7.1 Population and sample

Population

Two different populations were involved in this study: research strand A involved UK-based SLTs with experience of working in the field of HNC dysphagia. Research strand B involved accessing a secondary data source i.e., NHS notes of patients who had been treated, between 2016-2021, for HNC in the South-East Scotland Cancer Network region and had had swallow outcome measures taken by NHS Lothian SLTs.

Sample and eligibility

Convenience sampling was chosen for both strands of this study as no other options were available. Response rates with a probabilistic design for strand A would likely have been too low, and for Strand B it would not have been possible to collect prospective data on sufficient new patients within the time constraints of the study. In addition, non-probability sampling for Strand A was chosen as the aim was to capture as many HNC experienced UK SLTs as possible, rather than a smaller random statistically representative sample. As by necessity all samples were convenience in nature, data were checked for representativeness of the target populations by analysing the spread of respondents in terms of geographic location and clinical experience.

Research Strand A

A sample of SLTs working in the UK with a clinical caseload of HNC patients was accessed by email via professional networks and invited to complete the online questionnaire as described in Section 3.7.2. Calculation of response rate was not possible, as there is no national database of SLTs working in the field of HNC, so the exact sampling frame is not known.

Research Strand B

MDADI data collected by SLTs in NHS Lothian with HNC patients as part of routine clinical practice were accessed to allow exploration of the psychometric properties of the MDADI based on data from a UK population of HNC patients. Checking accuracy and completeness of data is part of data quality assurance (O'Reilly et al., 2016). Prior to analysis, data quality checks were carried out on the MDADI data to ensure data completeness and integrity (Buchanan and Scofield, 2018).

MDADI global and composite scores are recorded in the body of SLT notes, but the completed questionnaires are also available in the SLT notes and were accessed to give individual item data as well as to double check recorded composite and global scores. These were accessed following Caldicott approval.

Power calculations were not performed, as a specific sample size was not planned for a priori due to the unknown size of the UK HNC-experienced SLT pool.

Eligibility Criteria

For research strand A, SLTs self-reporting as working (or having worked) in HNC were eligible for inclusion. SLTs with no exposure to a HNC caseload are unlikely to be familiar with the HNC-specific MDADI tool and were therefore excluded. This criterion was highlighted in the questionnaire introductory email (see Appendix E). For research strand B, adult patients aged 16 or over with a diagnosis of HNC who had received treatment from the SLT team in NHS Lothian were eligible for inclusion. They must have also completed at least one MDADI during their SLT episode of care. Inclusion and exclusion criteria are documented in Table 15:

Inclusion criteria	Exclusion criteria
• Strand A: UK-based SLTs who work/have worked with head & neck cancer patients • Strand B: Patients aged sixteen or over undergoing treatment for head & neck cancer in NHS Lothian	• Strand A: SLTs who have never worked with patients with HNC • Strand B: HNC patients under the age of sixteen

Table 15: Inclusion and exclusion criteria

3.7.2 Recruitment and consent

Recruitment

Research Strand A

The link to the questionnaire was circulated via email around UK SLT professional groups: the Royal College of Speech & Language Therapists (via the RCSLT Enquiries Coordinator), the North & South England HNC 'CEN' (SLT 'Clinical Excellence Network', via the CEN secretaries), the Scottish Head & Neck cancer SLT network (via local clinical leads) and the Scottish SLT managers' network (via the NHS Lothian SLT adult service manager) on 1st November 2021. This sample was augmented by a 'snowball' recruitment effect: clinicians were encouraged to share the hyperlink around the HNC SLT community widening the scope of distribution (Gilbert and Stoneman, 2016); the questionnaire was also promoted via the author's academic account on the social networking platform Twitter™ to facilitate dispersal within the UK SLT HNC community. The sample was limited to UK clinicians to avoid any issues with cross-cultural applicability of responses.

The email participation invitation (Appendix E) included a brief background to the study, the hyperlink to the online questionnaire, and the clinician Participant Information Sheet (PIS) as an attached document (Appendix F).

Once the link was circulated, the questionnaire was kept open to responses for three months to allow adequate time for both clinician response and snowball recruitment as clinicians circulated the questionnaire more widely amongst their colleagues. A reminder email and 'tweet' requesting clinicians respond to the questionnaire was sent two months post initial circulation (Kelley et al., 2003).

Research Strand B

The MDADI data to be used in the quantitative analysis had already been collected as part of routine NHS Lothian SLT clinical practice, therefore permission to use these data was sought and granted from the NHS Lothian Caldicott Guardian in accordance with Caldicott principles concerning access to and use of confidential patient data.

Collection of data from patient records for Research Strand B took place whilst the research strand A questionnaire was open to responses.

Consent

Informed consent is at the heart of ethical research (Department of Health, 2005). The following steps were taken to ensure issues pertaining to consent were considered and current guidelines followed.

Research Strand A

To ensure participant informed consent for the questionnaire, a PIS was provided at time of recruitment (Appendix F). This PIS outlined the project aims, methods and rationale, and what was required of participants. Consent was inferred from clinicians accessing and completing the online questionnaire: a consent page constituted the first response page of the survey and participants were not able to progress until they had indicated their consent by checking the appropriate response box. All participants were assured of individual anonymity and provided with the opportunity to withdraw from the study at any time.

Research Strand B

The MDADI data were collected as part of routine clinical practice by the HNC SLT team within NHS Lothian as part of their episode of care. Consent was not contemporaneously sought from patients for use of this data for research purposes, post hoc permission to use the data for the purposes of this study was gained from the NHS Lothian Caldicott Guardian. MDADI scores were anonymised and not individually identifiable.

3.8 Data analysis

The aim of data analysis was to interpret and then synthesise the qualitative and quantitative data gathered to provide a detailed analysis of the MDADI and investigate its psychometric properties and clinical utility.

3.8.1 Qualitative analysis: Research Strand A

Questionnaire data were downloaded from the online survey platform in the form of a Microsoft Excel™ spreadsheet for processing. Demographic data were analysed with descriptive statistics to give an indication of the background of the participants.

The narrative data underwent a process of reflexive Thematic Analysis (TA) using Braun and Clarke (2021)'s 6-phase guide to performing this technique. TA provides a theoretically flexible, accessible approach to qualitative data analysis and can "provide a rich and detailed account" of data (Braun and Clarke, 2006 p.78). The process of TA involves becoming familiar with the data before generating initial codes and searching for themes, which are ultimately named and defined to produce a 'thematic map' of the analysis; it is a process of disassembling, reassembling, interpreting and drawing conclusions on the data (Castleberry and Nolen, 2018). Themes that reflect the data gathered were collated to compile a list of significant points that informed the quantitative analysis. These data also drove suggestions for potential future enhancement of the MDADI in terms of content validity and clinical utility.

Securing quality in data analysis

Ensuring quality of process and outcome in qualitative data analysis is key (Birks, 2014). Consideration of Morse (2015)'s qualitative rigour criteria of reliability and validity were kept at the core of the structured process of TA described above. TA was chosen for this study over alternatives, such as Content Analysis (Hsieh and Shannon, 2005). Vaismoradi et al. (2013) make comparison of the two techniques, and this analysis, in addition to the researcher having had prior experience of the TA and having confidence in its use, led to the decision to use TA in this study. As this study follows a Pragmatic, mixed methodology with both qualitative and quantitative data analysis and interpretation, the more highly complex traditional qualitative approaches of grounded theory or hermeneutic phenomenology were not felt to be appropriate in this case.

3.8.2 Quantitative analysis: Research Strand B

As discussed in Chapter Two of this thesis, the statistical approaches of Classical Test Theory (CTT) and Item Response Theory (IRT) can both be used to assess psychometric properties of outcome tools. To answer Research Questions Two and Three, Item Response Theory was employed, alongside COSMIN guidance as appropriate. Table 16 below shows which aspects of the MDADI were analysed quantitatively using the COSMIN framework and

an IRT statistical approach: internal consistency and structural validity. In addition, DIF is also investigated with IRT methods in this thesis, although this is not currently a specific domain within the COSMIN framework.

COSMIN Domain	Measurement property	Aspect of a measurement property	Definition	Being analysed in this thesis?
Reliability			The degree to which measurement is free from error	
	Internal consistency		The degree of item interrelatedness	Yes – COSMIN suggests presenting standard error of θ (SE (θ)) or a reliability coefficient of estimated latent trait value (Mokkink et al., 2018).
	Reliability		How 'true' differences between patients contribute to the total variance in measurement	No. Does not apply in this case as many different clinicians involved in administering MDADI.
	Measurement error		Concerns the error of an individual's score that do not relate to changes in the construct being measured	No
Validity			The degree to which a PROM measures the construct(s) it purports to measure	
	Content validity		Are all items relevant to the construct being measured? Are all key concepts included? Are the items comprehensible to the population of interest? (Terwee et al., 2018)	This data is being sought in Research Strand A (qualitative data)
		Face validity	The degree to which items in a PROM looks as though they are an adequate reflection of the construct to be measured	No
	Construct validity		The degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured	See structural validity
		Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured	Yes – MDADI is ostensibly unidimensional (only assesses swallowing related quality of life) - so this assumption will be checked by assessing IRT GRM

				model fit as per COSMIN guidance.
		Hypothesis testing	Idem construct validity	Yes – quantitative strand (structural validity)
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM	No - Not being assessed with IRT.
	Criterion validity		The degree to which the scores of a PROM are an adequate reflection of a 'gold standard'	No. Unable to address this as clinically not relevant to use two different quality of life tools in practice, therefore data unavailable.
Responsiveness			The ability of a PROM to detect change over time in the construct to be measured.	
	Responsiveness		Idem responsiveness	This study is not focusing on this aspect.
Interpretability			The degree to which one can assign qualitative meaning to a PROM's quantitative scores or change in scores.	No – Not being assessed with IRT.

Table 16: COSMIN property quantitative analysis plan

The data analysis packages Stata™ and R™ were used to analyse quantitative data. As MDADI items are polytomously scored via a Likert scale, for the main IRT analysis Samejima (1969)'s Graded Response Model (GRM) of IRT was used. The GRM allows analysis of items from multiple-item tests which use 'multiple ordered-response categories' (Chang and Reeve, 2005) such as Likert scales; in the case of the MDADI: the response categories are *strongly disagree/disagree/no opinion/agree/strongly agree*.

3.8.3 Data synthesis

It is essential in a mixed methods study that qualitative and quantitative data is combined to generate knowledge beyond that which a single method study could have provided (Franz et al., 2013). Data integration from both strands needs to be conceptualised at the planning stage and carried out in concordance with this plan consistently (Guetterman et al., 2019). A lack of transparency in reporting compromises overall study quality (Irvine et al., 2020). As the COSMIN framework does not include aspects of tool clinical utility, I chose to augment the framework with clinical utility items as part of the analysis for the purposes of this study. This decision will be revisited in the discussion in Chapter Seven of this thesis when the contents of the COSMIN framework will be reconsidered in light of the results of this study.

Data integration

As a mixed methods study, integration of the quantitative and qualitative aspects is planned at the stages of design, methods and interpretation to enhance the results gained and maximise the advantages of the mixed methods approach (Fetters et al., 2013). At the design level, this study took a convergent, interactive approach where quantitative and qualitative data are collected and analysed simultaneously with the potential for findings from one strand to influence the focus of data collection of the other strand (Creswell, 2014). In this study, data collected from the qualitative questionnaire strand had the potential to influence factors explored during the quantitative strand, e.g., when exploring the potential for DIF in the MDADI.

At the methods level, data integration occurred through 'merging' (Fetters et al., 2013), that is the qualitative and quantitative databases were brought together for comparison and synthesis. Moseholm and Fetters (2017) describe this approach as 'data diffraction': where qualitative and quantitative strands in a study generate different 'cuts of data' that will illustrate different aspects of a central phenomenon – that is in this case the MDADI tool.

At the level of data interpretation, this study integrated data through narrative, using a 'weaving' approach with quantitative and qualitative findings where there is potential for both types of data to be presented on a theme-by-theme basis (Fetters et al., 2013).

As discussed in the preceding literature review, several tools for quantitatively assessing the psychometric validity of patient reported quality of life outcome tools exist (Francis et al., 2016, Aaronson et al., 2002, Terwee et al., 2007, Mokkink et al., 2019). However, no assessment including the MDADI has yet used the COSMIN tool (Mokkink et al., 2019) despite it being considered the most widely used and comprehensive tool currently available (Lorente et al., 2020).

In this study, the content of the COSMIN tool was used as a framework to guide both qualitative and quantitative data collection and analysis. In addition to the COSMIN domains, extra aspects of tool 'clinical utility' were considered in the qualitative research phase as discussed in the preceding Literature Review, as these data are not collected in the COSMIN tool but have relevance to use of the MDADI in clinical practice. The potential to create a shorter form of the MDADI was explored by combining qualitative and quantitative data to create a short version that incorporated IRT, DIF, content validity and clinical utility information.

These combined data give an indication of the psychometric strengths and weaknesses of the MDADI and its clinical utility, and what the focus for future enhancement of the MDADI would need to be.

CHAPTER 4: QUALITATIVE RESULTS

This study constitutes a mixed methods analysis and evaluation of the MDADI HNC dysphagia-related quality of life outcomes assessment tool. The Research Questions that the following results and analysis presented in Chapters Four, Five and Six will answer are:

1. What is the content validity and clinical utility of the MD Anderson Dysphagia Inventory from a UK SLT perspective?
2. What are the results of an exploration of the psychometric properties of the MDADI on a UK population using Item Response Theory?
3. What are the potential factors which might result in differential item functioning?

Question 1 will be addressed by the following qualitative analysis, with Questions 2 and 3 being addressed by the ensuing quantitative analysis presented in Chapter Five. Chapter Six will present a synthesis of qualitative and quantitative results.

4.1 Survey piloting

Piloting is an essential part of survey design and allows for potential issues with format, wording and flow to be identified and rectified prior to live circulation (Story and Tait, 2019). It also allows the researcher to check that participants will interpret and respond to questions as intended (Bowden et al., 2002).

To ensure overall validity, prior to wider circulation the survey was piloted online on 8 piloters to check that it was feasible in terms of length and structure and that the content and flow made sense (Patrick et al., 2011). In addition to the survey itself, the PIS, invite email and survey instructions were also piloted to ensure clarity (Story and Tait, 2019). SLTs who work for NHS Lothian in dysphagia, but not in the field of HNC, were approached to pilot the tool, so as not to reduce the number of potential respondents, but to capitalise on their experience as clinicians with insight into outcome measurement and PROM tools. Comments from piloters were incorporated into the final survey design. Piloters were asked for specific feedback on layout, functionality, flow, wording and typographic errors, and also to give an estimate of time taken to complete the survey. In addition they were asked to consider whether the tool met Bowden et al. (2002) and Sullivan and Artino (2017)'s criteria for formulating questionnaire items previously presented in the Methods section of this thesis. Specific feedback provided is summarised in Table 17 below:

Issue category	Comments
Typos	• Q 33 MADADI should be MDADI • inconsistency in whether there is a hyphen in 'patient-reported' tool in the intro • no hyphen in 'patient-reported' in the consent session • In the section 'what happens to the data?' 'andthen' is all one word • Self-esteem is missing a hyphen • Page 2 Consent: 2a) should it be, "within **or** before"? • Page 2 Consent: 2d) should it be, "leaves" (plural)? • Item 34 of the questionnaire: should it be "time take**n**" instead of "time take"?
Wording	• On page 1 of the online survey there is a question asking if responses will be anonymous. First word of answer is "no" which for an instance made me think it was not anonymous could change it to "personally identifiable information will not be requested in this survey"? • might be helpful for HNC to be written out again in full if this is for patients, but maybe it's just SLTs completing it? I think this won't be an issue as all clinicians filling this out will be familiar with the acronym – plus it is spelled out in full initially
Formatting	• Question 10 (1) is in bold, but none of the others are
Functionality	• some questions say 'required' and there is one that says 'optional', but the other ones say neither – so are they optional? • My opinion would be to have as little as possible that is mandatory as most of your responders will hopefully be trying very hard to fill in the majority of points.
Time taken	• 30 mins (x 6) • 10 mins just to read through • 45 mins

Table 17: Summary of questionnaire pilot feedback

Following receipt of these comments, survey wording, formatting and typography was amended to incorporate the suggested changes. A non-substantial, non-study-wide review minor amendment was granted via IRAS for these changes.

4.2 Survey responses – demographic data

The format of the survey meant that several different types of data were generated: basic demographic data on respondents' clinical experience and use of the MDADI, item-level MDADI data, and data produced in response to specific questions around clinical utility of the tool and a request for general comments.

Survey data were downloaded from the online survey website in Microsoft Excel™ spreadsheet format. Descriptive data for the survey respondents' location, clinical experience and use of the MDADI are summarised in the following tables.

Years of experience and country of work

The total number of respondents to the online survey was 31. There were 19 responses from England,12 from Scotland, and no responses from clinicians in Wales or Northern Ireland. Respondents comprised a wide range of years of clinical experience working with people with HNC (1-34 years) with a mean of 12.42 years.

Use of the MDADI in clinical practice

Of the 31 survey respondents, 20 were using the MDADI in their clinical practice at time of survey completion.

Currently using MDADI	Freq.	Percent
No	11	35.48
Yes	20	64.52
Total	31	100.00

Table 18: Respondents' MDADI use

Pattern of use of MDADI in practice

Clinicians were asked to indicate, in a free-text box, at which points in a patient's cancer pathway they had used the MDADI. Table 19 demonstrates the different points at which respondents used the tool and shows considerable variation in practice in terms of timing of MDADI use.

Timing of MDADI	N at this timepoint
Pre-treatment	22
Directly post-surgery	1
Immediately post treatment	1
3/52 post treatment	2
1/12 post treatment	1
6/52 post treatment	2
Every 6-8/52 during follow-up	1
3/12 post treatment	8
6/12 post treatment	13
1y post treatment	7
When clinically indicated	1
Before any block of therapy	2
After any block of swallow therapy	1
At time of SLT discharge/end of SLT episode of care	5
As indicated by any research study protocols	1

Table 19: Timing of MDADI use

In summary, the group of survey respondents comprised clinicians with a range of years' HNC experience, from two countries within the UK. Demographic data confirmed that the MDADI is being used in clinical practice, but that the pattern of tool use varies greatly. The results of analysis of the narrative data on the content of the MDADI generated by the 31 survey respondents will now be presented.

4.3 Thematic analysis results of survey data

As described in Chapter Three of this thesis, the qualitative analysis of survey data was carried out using the 'reflexive Thematic Analysis' technique as described by Braun and Clarke (2021). I followed their six steps as outlined in Table 20 below.

1	Dataset familiarization
2	Data Coding
3	Initial theme generation
4	Theme development and review
5	Theme refining, defining and naming
6	Writing up

Table 20: Braun & Clarke's reflexive Thematic Analysis process

My reflexive Thematic Analysis of the survey data aimed to allow me to answer Research Question 1:

> What is the content validity and clinical utility of the MD Anderson Dysphagia Inventory from a SLT perspective?

And also, to highlight data relevant to Research Question 3:

> What are the potential factors which might result in differential item functioning?

As outlined in the preceding chapter, a hoped-for outcome of this study is generation of data that could guide or inform future development of the MDADI tool. With this in mind, and facilitated by the detailed, practical data generated by clinicians who responded to the survey, the following analysis focuses on both a 'macro' overview of the MDADI tool, and a more 'micro' or granular analysis, allowing issues with specific MDADI items to be explored and addressed through the qualitative data. As illustrated by the research diagram in Chapter Three, these qualitative data also then feed into the quantitative analysis by highlighting potential factors that might result in differential item functioning within the tool.

4.3.1 Inductive vs deductive analysis

In this mixed methods study, the data generated by the survey had to work hard in several ways. Firstly, they had to power an inductive qualitative analysis of the themes generated by UK clinicians reflecting on the MDADI and its use. In addition to this, the data would also be used deductively to explore the clinical utility of the tool in the context of the clinical utility criteria as generated in the literature review of this thesis. Finally, data from the survey also had the potential to inform the quantitative analysis arm of this study, particularly in terms of highlighting issues that might contribute to differential item functioning that could be analysed statistically.

I took a mixed inductive and deductive approach to my data analysis, as I felt this would give the fullest and most detailed consideration of my dataset. A precedent exists for this mixed approach, for example in Byrne (2021)'s worked example of a reflexive thematic analysis. I began with an inductive approach, exploring codes and themes from the data, trying to minimise preconceptions around what I expected to come up. Following an inductive analysis of the data and initial code generation, I then went back to the beginning and this time approached the whole dataset from a deductive standpoint, starting with the context of my literature review of clinical utility, considering the data with specific reference to these criteria.

The reflexive nature of Thematic Analysis

A cornerstone of Thematic Analysis as described by Braun & Clarke is the 'reflexive' nature of the process. The researcher should systematically and robustly consider their position within the analysis and what preconceptions, potential biases and perspectives they are bringing to the process of the analysis.

In my position as a practising SLT who uses the MDADI tool, I am considered an 'insider researcher' when analysing this data (Braun and Clarke, 2021). This is advantageous in that it gives me a deep and nuanced understanding of the topic at hand but is also potentially a source of bias as my years of clinical practice have meant I have developed my own opinions about the MDADI. Therefore, throughout the analysis I had to be vigilant for the potential for this to excessively influence my analysis of the data.

To facilitate self-reflection and keep track of the analysis process, I kept a 'reflexive diary' throughout. This helped me document thoughts and ideas as well as providing an opportunity to reflect on the process, and to identify and consider reactions I had to certain data items when I read respondents' comments with my 'clinician hat' on.

The diary proved to be invaluable, both as a tool to highlight and hopefully minimise my own biases from influencing the analysis, but also as a refresher and summary tool. At the start of every session of analysis, I re-read through all my previous notes, and over time this helped me to recognise patterns, facilitating the procedure of coding and theme generation.

What now follows is a four-part description of the process and results of the survey data qualitative analysis:

1. Inductive analysis
2. Deductive analysis: clinical utility
3. Practical suggestions for tool amendment
4. Data that will inform subsequent quantitative analysis

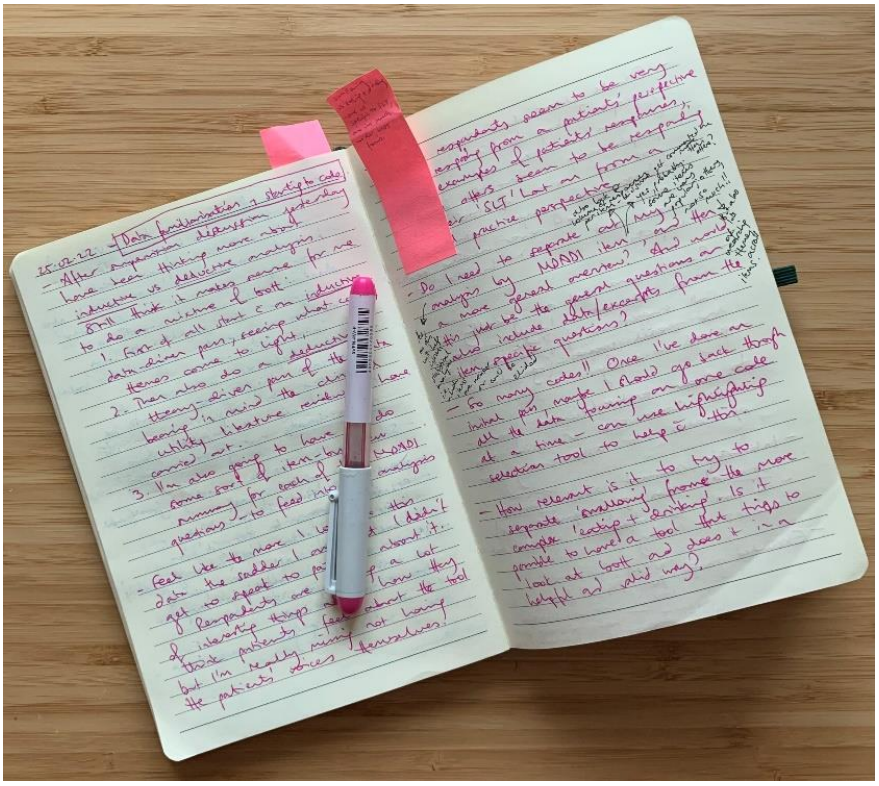
4.3.2 Inductive reflexive Thematic Analysis results

In this section I will present a narrative description of the process of the inductive reflexive Thematic Analysis and the results this generated. The outcome of the deductive analysis of the data informed by the literature on clinical utility, and analysis of the data for potential to inform the quantitative analysis arm of this study, will be presented in the sections following this.

Dataset familiarisation and initial coding

Survey responses were imported into NVIVO 12™ software to facilitate analysis. Accessing qualitative text-format data via a screen was a novel experience for me, my prior experience of thematic analysis having been purely pen-and-paper based. I found this meant I had to focus harder and re-read the data more than I anticipated. Also, as I was learning how to use NVIVO for the first time as I went along, initially getting familiar with the dataset and thinking about initial code generation took time to get going. Keeping a reflexive diary of notes about my thoughts and feelings during the process was invaluable, as initially due to work commitments I only had one day a week where I could devote time to the analysis, so reading my previous notes really helped to get myself ‘back in the zone’ for an analysis session. Figure 5 shows a page of notes from my reflexive diary:

Figure 5: My reflexive diary



I had been concerned that with 31 respondents, there might be a dearth of data, however as I read and re-read the responses, I felt increasingly impressed by (and grateful for) the level of engagement from the survey respondents. Many had obviously taken a significant amount of time to consider the survey questions and provide detailed, thoughtful answers. I found that I had a strong emotional response about this; also, about the fact that I could see my own clinical experience and experiences of my patients reflected back to me during the analysis process. Challengingly also however, there were other perspectives and narratives that I could see coming through in the data that were different to my own. I was very aware of not wanting my own clinical 'frame' or stance to silence these voices and made sure to keep honest notes in my diary about when these conflicts occurred, so that I could be aware of my bias and try to prevent it from unfairly colouring my coding and analysis.

The first codes I generated were semantic in nature, focussed on surface level, explicit concepts evident in the data; for example there was a clear thread of clinicians highlighting how useful the MDADI can be in clinical practice as a 'conversation starter' to start helping to unpick the complex impact that a swallowing problem has on a patient's life. Likewise, there were explicit references to the fact that many items in the tool were potentially exclusory to patients whose dysphagia was so severe that they were unable to eat and drink at all. As I continued to study the data, more latent themes came through, particularly around the concept of swallowing versus eating, and the fact that, as patients with HNC experience so

many other side effects of treatment that directly or indirectly impact on their swallowing, the focus of some MDADI items may not be sufficiently specific.

There were also many ‘practical’, constructive suggestions throughout the data for change, development and potential improvement of the MDADI; although I hadn’t explicitly asked in the survey for such suggestions, there were a great many of these made and I found this very exciting and inspiring as an extra potential output for this study.

Although my aim was to focus on inductive coding initially, acknowledging I would be doing further analysis of the dataset with my deductive a priori clinical utility criteria in mind, nevertheless as I became more familiar with the dataset, I felt there were motifs that were related to both latent codes concerning clinical utility and the quantitative strand of my overall study. In terms of clinical utility, additional aspects that I hadn’t previously considered were evident in the data, and it also became clear that clinicians were highlighting issues they had observed that I could identify as being potential sources of differential item functioning, therefore feeding into the quantitative strand of my study.

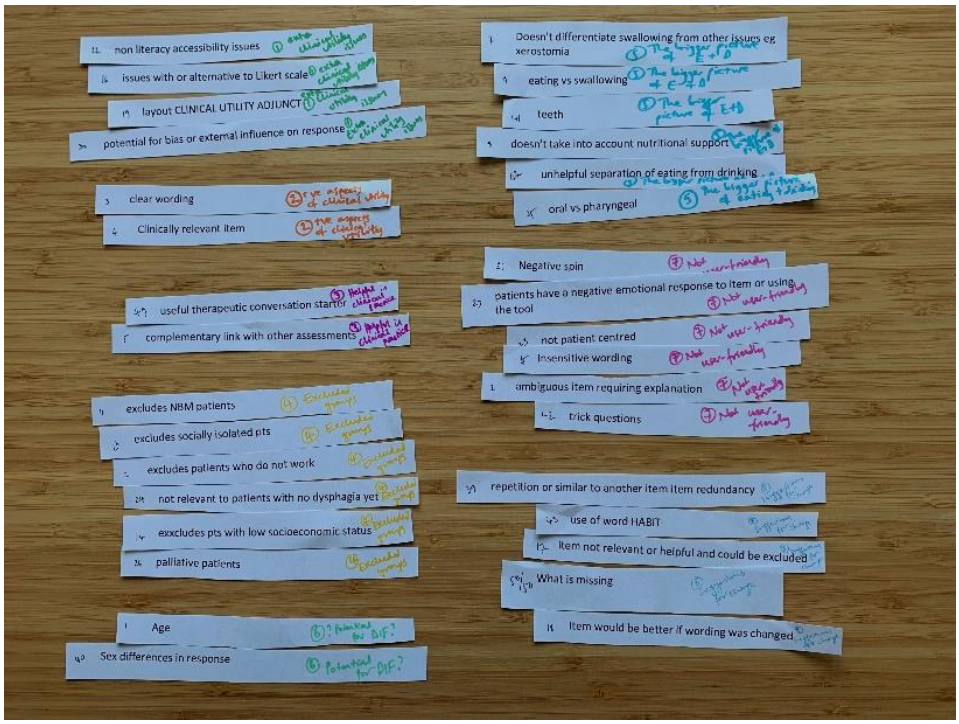
Initially I identified more than 50 codes; a long list that felt unmanageable and rather overwhelming. As I read and re-read the codes and dataset however, it became clear that there were significant areas of overlap between some of my initial codes; some felt too narrow or fragmented, whereas others continued to feel clear and distinct. I found looking at a list of codes on a screen rather alienating, so reverted to printing off all the codes, cutting them up on strips of paper and ‘playing around’ with them on a tabletop. This made me feel more connected to the process and allowed me to more easily identify codes that didn’t ‘feel right’: codes that were too broad or narrow, and to see patterns among the codes that were the beginning of theme evolution.

Initial theme generation, development and review

Once I had a workable set of codes, I started to think about potential themes. Physically manipulating the codes written on pieces of paper felt right to me, allowing me to form and reform potential groupings. Thinking more specifically about themes allowed me to see that some codes I had included in my shortlist were not sufficiently evidenced or meaningful for continued inclusion, and these were subsequently dropped from the analysis.

Figure 6 shows a photo of the stage where I had identified codes that I felt happy captured the different inductive meanings in the data and was able to start grouping them together into proto-themes.

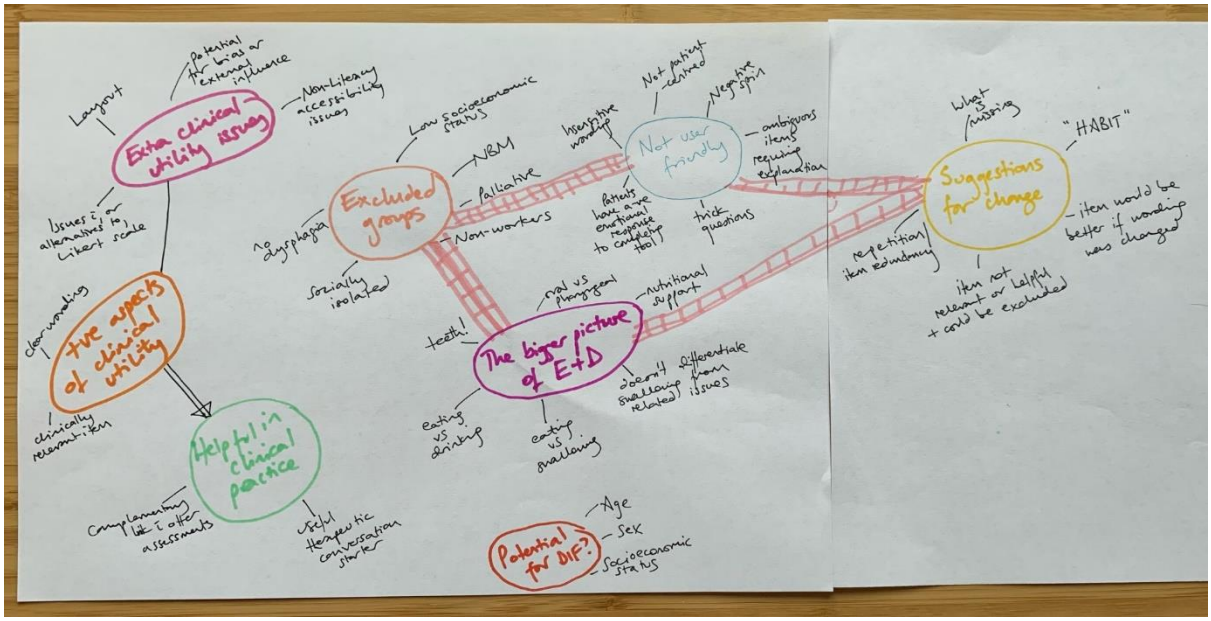
Figure 6: Initial grouping of codes into proto-themes



Theme refining, defining and naming

Once I had generated a group of proto-themes, I drew a mind map to help me visualise the themes and their components, and to facilitate thinking about potential links or overlap between these themes to help in the refining process. Figure 7 shows a proto-mind map of initial themes with accompanying codes.

Figure 7: Mind map of initial themes



The act of drawing this mind map helped me further refine my analysis; through this process I was able to see that the themes could be reduced without losing nuance of meaning, and

that the theme titles could be clearer. It also helped me to think about the overall narrative of my analysis and the most appropriate order to present my findings.

Code and theme generation audit trail

The process of code reduction involved starting off with 54 codes which were then reduced to 38, followed by the addition of 12 deductive clinical utility codes taking the total back to 50. Through the process of initial generation of eight themes, code numbers were reduced to 33 (not including clinical utility deductive codes). Through this process both code titles and theme names were continually developed and refined. Appendix G contains full lists of codes represented at each stage.

Seven themes and 32 codes (not including the 12 deductive clinical utility codes) survived this process; the seven themes are summarised in Table 21 below.

	Theme name	**Characteristics**
1	“a conversation starter”	How the MDADI is useful in clinical practice
2	Practical issues	Issues around using the tool in practice – linked to Clinical Utility
3	Potential for DIF?	Patient subgroups identified who might respond differently to MDADI items
4	The bigger picture of eating and drinking	Lack of focus in the MDADI about whether it is assessing swallowing, eating and drinking, other issues, or everything!
5	“not user-friendly”	How the MDADI is perceived to be negative, emotive and non-patient-centered
6	Excluded groups	Patient subgroups excluded by MDADI item wording or content
7	“not quite where we need it to be”	Practical suggestions for changes and improvements to the tool that would make it more useful and patient-centered

Table 21: Summary of themes generated from inductive analysis

Themes 2-7 could be seen to form part of an overarching theme, which comprises factors that could be seen to negatively affect the content validity of the MDADI. Specifically, that is that there are many aspects of the MDADI requiring further analysis and development to make it a more clinically relevant, inclusive and useful tool. Themes and their contributing codes and relationships are illustrated in the thematic map presented in Figure 8 below, and evidence for each theme from the data will now be presented. A larger version of Figure 8 is presented in Appendix H.

Figure 8: Mind map of codes mapped to themes

Theme 1: "a conversation starter"

This theme pulled together codes identifying the MDADI as a useful adjunct to clinical practice for SLTs working with people with HNC. Despite issues with the tool identified in other themes, many items in the tool were felt to be relevant and using the MDADI was still felt to add overall value to therapist-patient consultations. Clinicians acknowledged "*It's a useful QOL tool where few others exist. It was devised with HNC patients in mind*" (Respondent 13). There was a common thread that the MDADI acted as a "*conversation starter*" (Respondent 4) when used during clinical sessions. Respondent 5 described how the MDADI "*allows you to discuss a number of different topics that may not come up during informal discussion or a patient may not consider it to be an important part of their care*". Respondent 30 highlighted that the MDADI forms part of a thorough clinical evaluation, complementing other assessment tools: "*It provides information which assists clinical decision-making alongside other tools/measures*".

Theme 2: Practical issues

Codes forming this theme highlighted practical issues encountered when using the MDADI. Given the tool length and response format, layout of the MDADI as presented on paper was highlighted as an issue potentially affecting patients' engagement with and responses to the tool: "*it looks very cluttered on the page and not very user friendly. People feel overwhelmed with it*" (Respondent 22). Reduced accessibility of the tool for patients with cognitive impairment, mental health difficulties, or for non-speakers of English were also highlighted by many respondents. Respondent 3 suggested that "*patients with poor literacy would need extra support and also patients with a learning disability sufficient that they were unable to understand the material*". Accessibility issues specifically due to issues with literacy are discussed later in this chapter as this is a domain of clinical utility assessment.

A number of clinicians highlighted the fact that there may be potential for bias or external influence on patient responses as a result of needing support to complete the MDADI, or having carers or relatives present at the time: for example, for MDADI item 3 ("People have difficulty cooking for me"): Respondent 29 suggests that there is "*often bias if person who cooks is sat with the patient!*".

The Likert response format of the MDADI was highlighted by many survey respondents as an issue that made use of the MDADI frustrating for both clinicians and patients and had potential to affect the validity of patient responses. This is discussed in more detail later in this chapter in the section considering data relevant to the clinical utility of the MDADI.

Theme 3: Potential for DIF?

Respondents spontaneously identified specific items in the MDADI where subgroups of patients might respond differently, potentially leading to item bias (and DIF). For example, reference to 'embarrassment' in item 2 of the MDADI ("I am embarrassed by my eating habits") and 'self-esteem' in item 18 ("I have low self-esteem because of my swallowing problem") caused clinicians to question whether male and female patients might respond to these items differently: "*not well received by the "blokes" on our caseload, many of whom do not actively discuss mood/self-esteem with their therapists*" (Respondent 15). Age and socioeconomic status were also highlighted as demographic variables that might influence item response: "*Most of my patients are older so this isn't always helpful*" (Respondent 24, commenting on item 9); "*Can see the potential use but does not often highlight difficulties in our clinical caseload who often cannot afford to eat out*" (Respondent 15, commenting on item 8).

Theme 4: The bigger picture of eating and drinking

Swallowing ability forms part of a bigger physical, functional, emotional, psychological and social entity of 'eating and drinking'. The MDADI was designed to assess 'dysphagia-specific' impact on patients' quality of life, however the survey responses richly spoke to the fact that the tool frequently mixes swallowing impairment with other, more overarching aspects of eating and drinking, to the point that it is often not clear which aspect is being assessed. Respondent 19 described this as "*poor wording-eating habits doesn't equate to swallowing*". Likewise Respondent 7 identified ambiguity in referring to 'eating' rather than swallowing:

> *"The phrase 'eating habits' can be interpreted in many ways, and not necessarily relevant to swallowing. For example, some of my bariatric patients, or patients that wish to lose weight, will comment on this."*

A common theme throughout the dataset was the scope for issues other than oropharyngeal dysphagia to affect patients' responses to items. Respondent 4 described it thus:

> *"Most patients say my swallowing doesn't limit me- my issue is the pain/ saliva/ appetite/RT [radiotherapy] side effects. Often need to discuss this and so unsure how valid the response then is."*

Respondent 13 suggested the MDADI "*isn't always sensitive to patients who are in pain rather than a physiological problem e.g., following dental extractions.*" The impact of dental extractions, which are a common part of pre-HNC treatment workup, on MDADI responses, were frequently mentioned by respondents: "*If patients have had recent dental extractions this can influence their responses - sometimes need to guide them to think about oral intake prior to dental extractions*" (Respondent 17).

Theme 5: "not user-friendly"

This constituted the strongest theme, reported by respondents across all items of the survey.

Comments on the negative tone of the MDADI abounded throughout the dataset, with one respondent going so far as to describe the tool as "*negative, emotive, soul destroying*" (Respondent 14). Respondents found this tone to lack patient-centredness, commenting how it often cast patients in a passive or disempowered position. Respondent 10 evinced this in their comment:

> *"Some aspects are negative and focused on the potential views of other people, this can be upsetting for patients, it also focuses on swallowing difficulty and not … adaptations a patient may be effectively making".*

Likewise, Respondent 1 commented on the lack of a positive, empowering framing of issues in the tool:

> *"all the questions are framed in the negative - would be helpful to have some more positive statements e.g. I enjoy eating, I can find things I can eat on a menu, swallowing is not a concern for me".*

The use of the word swallowing 'problem*'* throughout the MDADI was frequently noted, with concerns that this negative framing had potential to "*skew the question*" (Respondent 19). Respondents often commented on other examples of negative or insensitive wording used in MDADI items, some feeling it was unsuitable to the extent that they stopped using the tool in their clinical practice:

> *"If [people] are really struggling then the questions are insensitive - reminds [people] of everything they are missing and likely to damage clinician-client bond, as this adds insult to injury. Largely why I stopped using it - didn't add to my clinical care and only those with mild-mod issues seemed well enough in themselves to complete" (Respondent 2).*

Respondent 21 echoed this sentiment with concerns about patients having a negative emotional response to using the tool: "*I can see that some of the questions may raise negative emotions that were not present prior to using the tool. Might be why I'm not keen to use it*".

Respondent 30 suggested that the negative, insensitive tone of the MDADI could be counterproductive, and that "*a more compassionate tone here may yield more representative responses*".

Items 5 and 15 in the MDADI caused consternation amongst respondents. They were commonly described as "*the trick double-negative questions*" (Respondent 29) that "*catch patients out*" (Respondent 19). This is because these are 'reverse scored' items, as explained by Respondent 17:

> *"I ask patients to read particularly carefully as they can often say disagree when they mean agree and vice versa due to trend in answers in the rest in the questionnaires i.e., in most questions "agree" indicates an issue whereas in this example "agree" is a positive thing".*

Throughout the MDADI tool, respondents highlighted items that were felt to be ambiguous, requiring explanation, and therefore being open to interpretation potentially affecting the validity of responses: "*Some questions are worded in a confusing way and the answers then*

need to be checked" (Respondent 7). Many respondents described how they had to 'step in' to help patients complete the tool due to ambiguously worded items: "*the wording of some questions is confusing therefore requires clarification from SLT*" (Respondent 18). In addition, often it was not just concern about patients misunderstanding or being confused by items, but also SLTs themselves: "*Patients don't understand the question and nor do I*" (Respondent 19).

Theme 6: Excluded groups

A recurring theme throughout the dataset was the identification of subgroups of patients who would not be able to answer some, or any, of the MDADI items due to their health or social status at time of assessment.

Respondents identified issues with items that referred to 'eating out' as potentially exclusory to patients of lower socioeconomic status who weren't in a position to afford to do so, such as item 8 ("I do not go out because of my swallowing problem") which "*does not often highlight difficulties in our clinical caseload who often cannot afford to eat out*" (Respondent 15).

Likewise, item 9 ("my swallowing difficulty has caused me to lose income") was felt to exclude patients who either couldn't work or who had retired. This was summarised by Respondent 12 who noted:

> *"I have had a number of respondents omitting this item or annotating it to say that they are not in employment. The predominance of over 60s or 65s in the HNC population tends to make this item slightly less relevant."*

Several items were also highlighted as being exclusory to patients who lived alone or were socially isolated, with the items' focus on friends, family and social interactions. These items were described as "*open to misinterpretation/non-representative answers as so many of our patients live alone/ do all the cooking themselves/don't have anyone who cooks for them*" (Respondent 17).

Significant issues throughout the tool were highlighted by multiple respondents in terms of using the tool as a baseline measurement with patients prior to their HNC treatment, often before they have any symptoms of swallowing difficulty. Most items presume a swallowing problem and therefore clinicians have frequently experienced non-dysphagic patients not knowing how to answer an item, potentially skewing or invalidating their results; in addition to causing anxiety and trepidation about what might be on the horizon in terms of future treatment side effects. Respondent 14 summarised this in their general comment about the tool:

"difficult to use with someone who was distressed, low in mood and also someone who doesn't have swallowing difficulties as it might highlight what may happen in the future and be really scary."

This concern was echoed by many respondents, including Respondent 17:

"It can also concern people who do not have swallowing difficulties, as they worry that the questions are an indication of what they will face in future e.g., they might not be able to eat out, enjoy a meal with friends, be embarrassed about their eating etc".

Patients in a palliative or end-of-life stage of HNC were also identified as a group that the MDADI would be challenging to use with. Clinicians indicated that such a lengthy, potentially distressing tool was not appropriate to use with this patient group, "*where it may seem like a tick-box exercise rather than focusing on their needs at that time*" (Respondent 7).

Finally, patients whose dysphagia is so severe that it has been recommended they be 'nil by mouth' (NBM) were highlighted as a group for whom responding to the MDADI would be extremely challenging. This constitutes a group of patients with dysphagia severe enough to warrant having been advised to avoid eating or drinking at all due to concerns over swallow safety and potential catastrophic negative health sequelae that might result from aspiration (food and drink 'going down the wrong way') or choking. This cohort of patients must rely on non-oral, enteral nutrition to meet their nutritional requirements. A common theme amongst respondents was that the MDADI was not appropriate to attempt in this situation and could not be ethically used with these patients: "*I wouldn't use this with someone who is NBM as a result of cancer/treatment as I feel it would be pretty insensitive*" (Respondent 17). This then means that the group who potentially have the greatest reduction in dysphagia related QoL, due to not being able to eat or drink at all, are excluded from having this impact measured, as summarised by Respondent 20:

"I don't use it with patients who are nil by mouth as I feel it's unfair - they can't answer many of the items and it is upsetting. This is a big issue though as they may be that patient group whose quality of life is most impacted by their dysphagia!!".

These subgroups constitute important, common subsections of the HNC patient group, and the fact that the MDADI is inaccessible to them means a significant proportion of HNC patients are potentially excluded from measurement of their dysphagia-related quality of life with this tool.

Theme 7: "not quite where we need it to be"

Throughout the survey, respondents made numerous comments on how the MDADI could (and perhaps should) be changed, with practical suggestions for rewording, modification, elision or removal of items. Several items were commonly described as "*irrelevant*" (Respondent 20), in addition many items were considered to be repetitious: "*several questions are very similar and would be good if they could be reduced*" (Respondent 8). The consensus was that shortening the tool by removing some similar items would make it more user-friendly, producing more valid data, as illustrated by Respondent 18:

> *"I feel many questions are similar and could be removed as the questionnaire is extremely long and patients often rush through it. If it was shorter I feel patients would take more time to consider their answers".*

Many suggestions were made for items that instead of being removed, could rather be improved with alterations in wording. Specific words in particular were disliked by respondents: "*Irritated is a strong word - frustrated would be more helpful*" (Respondent 22); the recurring use of the phrase "eating habit" was also flagged: "*are eating habits and swallowing problems the same? [It's] ambiguous for some people*" (Respondent 23).

Respondents made numerous suggestions for topics or items that they felt were missing from the MDADI and that would make welcome, relevant additions as issues which had significant potential to impact patients' quality of life:

> *"Think it would be good to ask specifically about safety of swallowing and efficiency. Any food avoidance. Impact of pain. Impact of taste changes. Impact of xerostomia. Impact of dental extractions/dental issues." (Respondent 13).*

Other additions were suggested in many other respondents' notes, such as adding items

> *"about avoiding textures; something about effect [of] teeth/being edentulous; questions about taste disturbance and xerostomia. [query] question about chest infections/impact these may be having on life" (Respondent 6).*

A common thread through the dataset was demonstration of an appetite for development and improvement of the MDADI. Respondent 24 declared that the MDADI is "*not good enough to be the only outcome measure*", with Respondent 14 hoping that "*there must be something better*". Respondent 10 summarised this as follows:

"I think it is useful to have a patient reported measure - I think the MDADI could be adapted/updated to better reflect patient experiences particularly patients having treatment for HNC."

This sentiment was echoed by Respondent 13: "*I think it needs re-vamped with a more up-to-date focus on the recovery of swallowing for HNC patients*".

Respondent 27 summed up both the challenges presented by using the MDADI as well as the potential benefits the tool can bring to SLT sessions, for example by facilitating meaningful clinical conversations:

"You need to be ready for the fact that this can make patients realise the extent of the impact of the dysphagia on their lives. Sometimes they have not really considered these questions before -e.g., others seem irritated, and it is always helpful to make time to talk through this after they have completed it - Tend to say something along the lines of " thank you for completing it, how did you find it? Did any of the questions or your answers surprised you or upset you? Is there anything you'd like to talk about? I notice you've put x for y question. I imagine that's very difficult... and it does often open up good conversations."

In summary, the inductive analysis contributes much to answering Research Question 1 and indicates that clinicians using the MDADI have concerns about aspects of its content validity and clinical utility. Results of the deductive analysis regarding the clinical utility of the MDADI will now follow.

4.3.3 Deductive clinical utility analysis results

The deductive analysis of these data with respect to the clinical utility of the MDADI had the twofold potential to 1. illustrate specific issues with the MDADI tool itself in terms of its clinical utility, and 2. generate results that might confirm, refute or expand on my literature-review-based summary of important aspects of clinical utility. Braun and Clarke (2021) suggest that deductive coding of data can "enrich the empirically based understanding of a theoretical concept" (p. 57).

Data adding to knowledge of the MDADI's clinical utility profile

The contribution of the qualitative data from this study to the understanding of the clinical utility of the MDADI is summarised in Table 22 below, and analysis results for each domain included in this analysis will then be individually presented.

RELEVANCE domain	Pre-existing data	Results from qualitative analysis
R1 Does the tool impact on clinical decision making?	Tuomi et al. (2020) suggest that a cutoff score <60 is useful to identify patients in need of swallow intervention	Yes, it has potential however multiple suggestions to improve validity as presented in Results
R2 Are there potentially issues with tool use with patient subgroups?	No previously published data	Yes – multiple issues
R5 Are questions likely to be distressing for patients?	No previously published data	Yes – multiple issues
R6 Is it possible to use a proxy for patients with special needs?	No previously published data	There may be issues of bias or coercion if using a proxy, or if family members are present whilst a patient completes the tool given the emotive nature of the content
R7 Is the tool acceptable to patients – would they be willing to reuse the scale in the future?	No previously published data	Many patients express reluctance to reuse the tool due to length and repetitive items
R8 Are the items relevant?	Explored statistically by Lin et al. (2022)with factor analysis but with no qualitative component	Unlikely all are relevant - see item level qualitative analysis and also quantitative analysis
R9 Is the recall timeframe appropriate?	No previously published data	Yes – 7-day recall timeframe corroborated by clinicians
USABILITY domain	**Pre-existing data**	**Results from qualitative analysis**
U1 Does the tool have an appropriate literacy level?	Only published work is looking at US English (Zraick et al., 2012)	Concern from clinicians that tool is too linguistically complex
U4 Is the time taken to administer acceptable?	No previously published data	Varied responses but predominant theme that tool is too long
U5 Is the scoring process straightforward?	No previously published data	Varied responses but predominant theme that scoring is too complex
U6 Is the time taken to score acceptable?	No previously published data	Varied responses but predominant theme that scoring takes too long

Table 22: Mapping existing MDADI knowledge to clinical utility domains

R1 Does the tool impact on clinical decision making?

This domain is similar in nature to the inductive Theme 1 previously discussed. Clinicians were specifically asked about impact on clinical decision making with a dedicated question within the survey, and responses to this item were varied in nature. Whilst some respondents did not feel the MDADI impacted on their clinical decision making, the majority indicated it was a useful tool, particularly as it had potential to flag up previously undisclosed issues and act as a starting point for further discussion and assessment, and also assist with other aspects of SLT management: "*it can help focus discussion … help focus therapy or assist in the discharge planning with patients*" (Respondent 13).

R2 Are there potentially issues with tool use with patient subgroups?

This is a significant area of weakness for the MDADI as previously highlighted and detailed in the inductive Theme 4 ("excluded groups").

R5 Are questions likely to be distressing for patients?

This is another area of significant issue for the MDADI as highlighted in the inductive Theme 6 ("not-user friendly").

R6 Is it possible to use a proxy for patients with special needs?

Respondents noted that they had "*made adjustments*" (Respondent 18) or that they had sometimes had to "*provide extra support*" (Respondent 3) when using the tool with patients with special needs, but no detail was provided about what this constituted. The use of a proxy brings up questions of ethics and potential for bias however when a patient is asked to complete a psychosocially sensitive tool such as the MDADI. Respondents indicated that the presence of a carer or relative might influence a patient's responses, or that carers/relatives might respond 'on behalf' of a patient with an answer that may not necessarily reflect the patient's experience: "*if they have a family member with them that person might jump in and say disagree/strongly disagree*" (Respondent 17).

R7 Is the tool acceptable to patients – would they be willing to reuse the scale in the future?

Respondents indicated that they had encountered patients who had completed the MDADI at the pre-treatment stage, when they were not yet experiencing dysphagia, describing it as "*a waste of time*" (Respondent 9); "*Some say it's not relevant to them*" (Respondent 13). Respondent 10 noted "*Patients have previously commented that they did not like completing the questionnaire*", whilst Respondent 6 expressed that "*if they have persistently low scores and are asked to complete it again for data collection purposes only, this could be frustrating for them."* – which could have implications for use of the tool for research purposes. Several other clinicians however indicated that if the clinical relevance and usefulness of the tool was sufficiently explained to patients, they were happy to complete it: "*generally patients are happy to engage and complete when it has been introduced and explained as an outcome measure*" (Respondent 10).

R8 Are the items relevant?

This domain speaks to content validity and is addressed in the inductive Themes 5, 6 and 7 ("the bigger picture of eating and drinking", "not user friendly" and "suggestions for change").

R9 Is the recall timeframe appropriate?

A specific survey item addressed this domain and many respondents felt that the 7-day recall timeframe of the MDADI is appropriate, particularly when at some stages in a patient's

journey their symptoms may be changing quickly. Respondent 27 noted "*Yes - often these questionnaires go for "over the last month" but in the early treatment stages, things change quickly so maybe 7 days is more appropriate*". One caveat was indicated by respondent 17 "*If patients have had recent dental extractions this can influence their responses - sometimes need to guide them to think about oral intake prior to dental extractions*". This ties in with the inductive Theme 5 ("The bigger picture of eating and drinking") and the difficulty clinicians have in teasing out patients' responses to isolate patients' swallowing function rather than other related factors, such as their dental status.

U1 Does the tool have an appropriate literacy level?

The literacy level of the MDADI was concerning to several survey respondents. Respondent 3 stated "*It is quite complex language at points*" and multiple respondents indicated that patients with literacy issues needed extra support or adjustments to the tool to help them complete it: "*I would always make adjustments to allow all patients to use this tool including support with reading/writing*" (Respondent 18).

U4 Is the time taken to administer acceptable?

Whilst a minority of clinicians commented that the "*length is okay*" (Respondent 23), many more clinicians made note of how long the MDADI takes to complete and how this is "*too long in everyday practice*" (Respondent 25). This was to the extent that sometimes the MDADI is "*not always possible to fit into sessions and busy clinics*" (Respondent 28), or "*there are too many statements to carry out the whole thing during a verbal session"* (Respondent 5), limiting its use over the phone or via videocall. Some respondents also noted that the time taken to administer the tool was not acceptable for patients as well as clinicians: the MDADI is "*too long for patients*" (Respondent 6); "*It's quite long and I think patients become tired of completing questionnaires*" (Respondent 1).

U5 Is the scoring process straightforward?

Many respondents echoed the statement that "*scoring is ok once clearly explained and doesn't take long*" (Respondent 6), and some had taken practical steps to aid the process: "*have created an Excel spreadsheet to help speed up scoring*" (Respondent 15). However, other respondents noted that the "*scoring is a bit clumsy*" (Respondent 19), and the tool is "*difficult to score*." (Respondent 20). In particular this is due to the two 'reverse scored' items (5 and 15), meaning many clinicians reflected the statement "*I often need to check the scoring*" (Respondent 27).

U6 Is the time taken to score acceptable?

The response picture was mixed for this domain, with some respondents stating, "*scoring is quick*" (Respondent 13), whereas others felt scoring took too long: "*It can take ages to do and to score. Scoring the composite score is time-consuming"* (Respondent 20).

In this summary of the deductive analysis of the data, overall there can be seen to be some overlap between inductive codes and themes and the deductive a priori clinical utility criteria, which I would argue adds weight and evidence to their use as robust assessment domains.

4.4 Practical suggestions for MDADI item-level change

Survey respondents provided many practical suggestions at MDADI item level for potential changes to the tool, aimed at making it easier to use, and more user-friendly. As this was the case, I have chosen to summarise these data separately, in addition to the inductive and deductive analysis of the data. The practical suggestions made by clinicians who use the MDADI in their practice add an extra dimension to the analysis of the overall validity and clinical utility assessment of the MDADI and these data could contribute to future development of the tool.

Practical suggestions from clinicians took three forms: wording suggestions at individual item level, comments on tool layout, and suggestions concerning changes to the current Likert response format.

Item level suggestions

Wording throughout the MDADI was highlighted as being problematic, falling within previously presented Themes 6 and 7 ('not user-friendly' and 'suggestions for change'). Clinicians suggested the language used across the tool is overly negative. Many respondents specifically posited that the term 'ability' should be substituted for 'problem' or 'difficulty' to make the MDADI more inclusive for patients at the pre-treatment stage who are not yet presenting with dysphagia, and to make the tone generally less negative. Making this change across the tool would have a direct impact on the wording of 8 items: numbers 6, 8, 10, 12, 14, 16, 17 and 18.

In addition to this, there were multiple comments and suggestions around lack of consistency in terminology regarding swallowing versus eating. In the MDADI 'swallowing' and 'eating' are used interchangeably (see Theme 5 'the bigger picture of eating and drinking'), however clinicians consistently made the argument that this was not appropriate and made items less valid or clinically relevant, and that the term 'swallowing' should be used consistently across the tool, rather than 'eating'. As discussed in Theme 6 ('not user-friendly'), clinicians felt that the wording of multiple items was not patient-centred, and overly

insensitive or exclusory. Indeed, some items were felt to be sufficiently negative and unhelpful as to be potentially 'droppable' from a future iteration of the tool. Specific examples of this included item 11 ('people ask me 'why can't you eat that'') and item 12 ('other people are irritated by my eating problem').

Considering respondents' comments at an item-by-item level as above also highlights suggestions that multiple items overlap in meaning, with clinicians frequently suggesting that certain items could be elided. In addition, one item in particular, item 9 ("my swallowing difficulty has caused me to lose income") was so frequently found to be irrelevant to patients that a strong case for removing this item from the tool was made by multiple respondents.

Tool layout

The length of the MDADI means that to fit on 2 sides of A4 paper, font size can be a maximum of 12 points. Clinicians commented that two sides of A4 paper is too long, and that the layout looked 'busy' creating an off-putting overall impression. It was suggested that a tool that would fit on one side of A4 would be welcomed in terms of improved clinical utility. Respondent 22 describes the layout of the MDADI thus: "*it looks very cluttered on the page and not very user friendly. People feel overwhelmed with it*". Respondent 12 echoes this point: "*I have had the feeling that [patients] find the look of it a bit daunting".* Respondent 4 concurred, stating "*I think some patients are put off by it being two pages*". One clinician (Respondent 30) felt the need to say sorry to patients for the length of the tool: "*I often find myself apologising to patients, particularly when it extends over the page*".

Likert response modality

Clinicians made comments regarding the Likert response format used in the MDADI across the survey, but also in response to the specific item regarding the Likert scale (question 30 in the survey). The most cited issue with the Likert scale was the descriptors used, particularly the inclusion of 'no opinion' as the middle descriptor. As stated by Respondent 14: "*I don't think anyone has no opinion of the kind of questions that are being asked*", with many others suggesting that 'no opinion' should be changed to something indicating ambivalence or neutrality, such as 'neither agree nor disagree', 'unsure', 'irrelevant' or 'it varies day to day'. Respondent 6 describes their experience: "*I find patients don't vary their answers between e.g., strongly agree versus agree and will stick with the same whichever they go with, right through*". Respondent 19 concurred:

> *"patients tend to go for one extreme or the other and use no opinion if it's neither one nor the other-so it's not that they have no opinion but that the other options don't work for them."*

Conversely Respondent 9 described how in their experience: "*the majority of patients tick agree or disagree, avoiding the end of each scale.*"

A shift from expressing opinion to frequency (e.g., always/never instead of agree/disagree) was also mooted. Table 23 summarises the Likert scale changes or alternatives suggested by respondents. Some respondents suggested a complete break from the use of a Likert scale in a future iteration of the MDADI. More accessible, non-linguistic options such as a 1-10 numerical rating scale, a visual analogue scale, or use of pictographic symbols similar to those used in other tools such as the CARE measure (Mercer et al., 2004) or SUDS scale (Wolpe, 1969) were suggested as potential alternatives. As Respondent 5 described: "*It's difficult to make [patients] choose a closed answer. I think a 1-10 scale would be better as there's more ability to account for change over time.*"

Suggestions for changes to Likert scale descriptors	• Always/often/sometimes/not often/never rather than strongly agree → strongly disagree • Alternative to 'no opinion', e.g., 'unsure', 'varies day to day', 'neither agree nor disagree' • Nonlinguistic indicators
Suggestions for alternative to Likert scale	• 0-6 like the SUDS tool • Numbers e.g., 1-10, rather than descriptors, with key at beginning of tool • Non- linguistic indicators like short version of Care Measure
Other suggestions	• Include 'irrelevant' option to omit item without affecting scoring • Different visual layout

Table 23: Summary of suggested Likert scale alternatives

Tool layout and response modality were evidenced to impact on clinical utility sufficiently to merit inclusion as additional parameters of clinical utility domains; this will be discussed further in Chapter Seven of this thesis.

4.5 Data that could inform quantitative analysis

During the thematic analysis process, several issues that could also be explored statistically in the quantitative arm of this study were highlighted. First the theme around potential for DIF provides further incentive for the analysis of this using IRT, specifically around the variables of age, sex and socioeconomic status for items 2, 3, 8, 9, 18 and 20.

Secondly, indications from this analysis that items from different subscales may overlap in meaning, and that some items could be elided or removed, provides impetus for a statistical analysis of the structure and internal consistency of the MDADI. In a tool that has been highlighted as overly lengthy, 8 out of the 20 items have been indicated as potentially clinically insufficiently distinct, and one item (item 9) was specifically singled out as being potentially removable, rather than ripe for elision with other items. The suggestions made by respondents regarding item elision are summarised in Table 24 below. Of note is that items flagged as having potential for elision or removal represented all three MDADI subscales:

functional, emotional and physical; however, items that clinicians felt were similar were not always from the same pre-determined subscale as demonstrated by the colour coding in the table.

Item number	Potential overlap- item number(s)
6	2, 5
8	14
10	16
11	2
14	1
15	8
18	2,5
19	7

Table 24: Items suggested for elision or removal; Key: Emotional = pink, Physical = green, Functional = blue

Potential for item elision or removal, and the structure of the MDADI subscale, are issues that were therefore analysed mathematically using IRT in the quantitative arm of this study.

4.6 Qualitative analysis summary

Analysis of the data generated by the online survey has painted a complex picture of the 'state of the art' of clinicians' use and opinions of the MDADI tool in their clinical practice:

1. The MDADI is being used in clinical SLT practice in the UK and is not just a research tool.
2. It can be a clinically useful and relevant tool, but UK clinicians have concerns about its content validity. A more constructive, inclusive, shorter, person-centred iteration of the tool would be welcomed by UK SLTs.
3. There are many issues with the clinical utility of the MDADI identified by UK clinicians, which would apply both to its use in research settings as well as clinical practice.
4. There is indication that there may be factors resulting in DIF within the MDADI, and that reanalysis of structure and internal consistency are warranted. The results of an IRT assessment of this will be presented in the next chapter.

Items 1-4 listed above will be discussed further in the context of the relevant wider literature in the Discussion chapter of this thesis.

CHAPTER 5: QUANTITATIVE RESULTS

Quantitative analysis of the patient MDADI data collected for this study allowed study Research Questions 2 and 3 to be addressed:

> **Research Question 2:** What are the results of an exploration of psychometric properties of the MDADI on a UK population using Item Response Theory?
>
> **Research Question 3:** What are the potential factors which might result in differential item functioning?

The results of the analysis will now be presented.

5.1 Research Question 2

As described in the Literature Review and Methods chapters of this thesis, the psychometric properties of the MDADI have yet to be analysed quantitatively using IRT statistical techniques. Table 25 below summarises aspects of psychometric property analysis of the MDADI that are currently lacking in the literature and that the COSMIN framework provides guidance for. The IRT based analyses of these parameters will be presented in this chapter.

COSMIN Parameter	Definition
Structural validity	Whether the dimensionality of the construct being measured by a PROM is adequately reflected by PROM tool scores
Internal consistency	Whether items that propose to measure the same construct generate similar scores – are items interrelated

Table 25: COSMIN parameters identified for MDADI IRT quantitative analysis

The preceding qualitative analysis highlighted an additional parameter meriting quantitative assessment at an MDADI item level: that is, analysis of potential item redundancy in the MDADI. IRT analysis provides the opportunity to assess the MDADI at an item level, and to identify item redundancy and therefore facilitate the creation of a shorter form of the tool.

Prior to presentation of MDADI IRT analysis results, demographic data describing the patient cohort from whom MDADI data were collected will now be reported.

5.1.1 Demographic, diagnosis and treatment data

Anonymised data from 302 patients were included in this quantitative analysis of the MDADI. These data represent both pre-treatment and six months post-completion-of-treatment swallow outcomes data from a cohort of patients with HNC treated with curative intent at the Edinburgh Cancer Centre, Scotland, UK, between 2016 and 2021.

30.46% of patients were female and 69.54% male, with an age range of 19-89 years and a mean age of 61. Scottish Index of Multiple Deprivation (SIMD2020) data were also collected for patients as a surrogate for socioeconomic status. SIMD scores represent “the extent to which an area is deprived across seven domains: income, employment, education, health,

access to services, crime and housing" (Scottish Government, 2020) and range from 1-10 with a lower score indicating a higher level of relative deprivation.

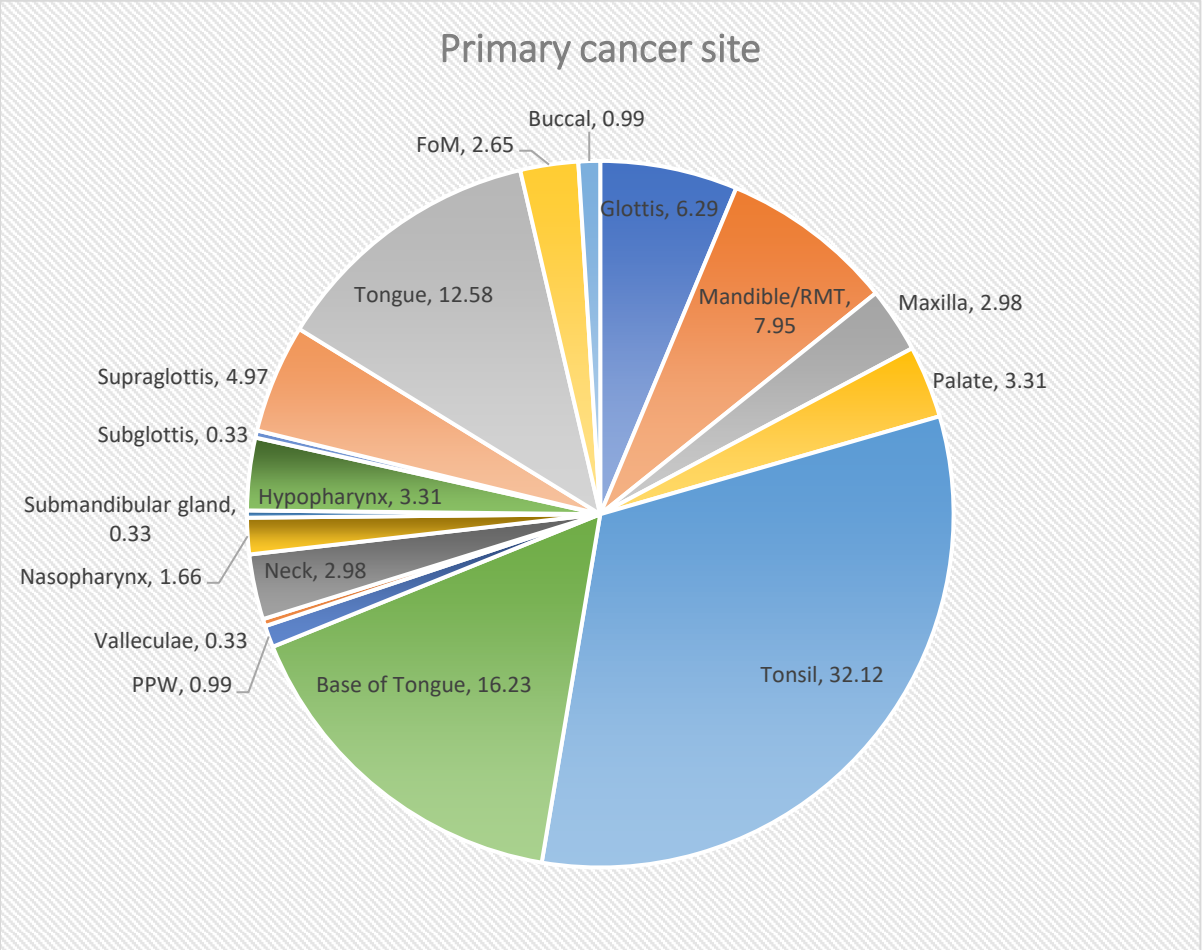
Characteristic	Value
Subjects (n)	302
Gender (%)	
Female	30.46
Male	69.54
Mean age (years)	61.29
Range	19-89
Working age (%)	58.28
Retirement age or more (%)	41.72
SIMD (%)	
1	3.32
2	9.97
3	12.29
4	9.97
5	9.90
6	11.96
7	5.98
8	10.63
9	7.64
10	18.94
Higher area-level deprivation (SIMD 1-3) (%)	25.58
Lower area-level deprivation (SIMD 4-10) (%)	74.42

Table 26: Demographic characteristics of patient cohort

For the purposes of later DIF analysis, SIMD was broken down into two groups: higher area-level deprivation (SIMD 1-3) and lower area-level deprivation (SIMD 4-10). Patients' ages were also categorized into working vs non-working age. Working age for women was taken as ≤60y and for men ≤65y, and retirement age or more for women was >60y and for men >65y. These data are all summarized in Table 26 above.

Patients in the cohort all had a diagnosis of primary HNC and received treatment with curative intent.

Figure 9: Breakdown of location of primary cancer site (%) Key: PPW = posterior pharyngeal wall; RMT = retromolar trigone; FoM = floor of mouth



As displayed in Figure 9, a range of primary cancer sites are represented, reflecting the diversity within the HNC diagnostic group. Oropharynx (tonsil, soft palate, posterior pharyngeal wall and base of tongue) cancers together dominated the group at 52.65%.

Figure 10 shows the distribution of cancer T staging for the patient group. T staging constitutes a score of 0-4, with a higher number indicating a larger, more advanced primary tumour (Brierley et al., 2017). All T stages were represented in the group, with the most common T stage being T2, comprising 33.11% of the group.

Figure 10: Cancer T staging

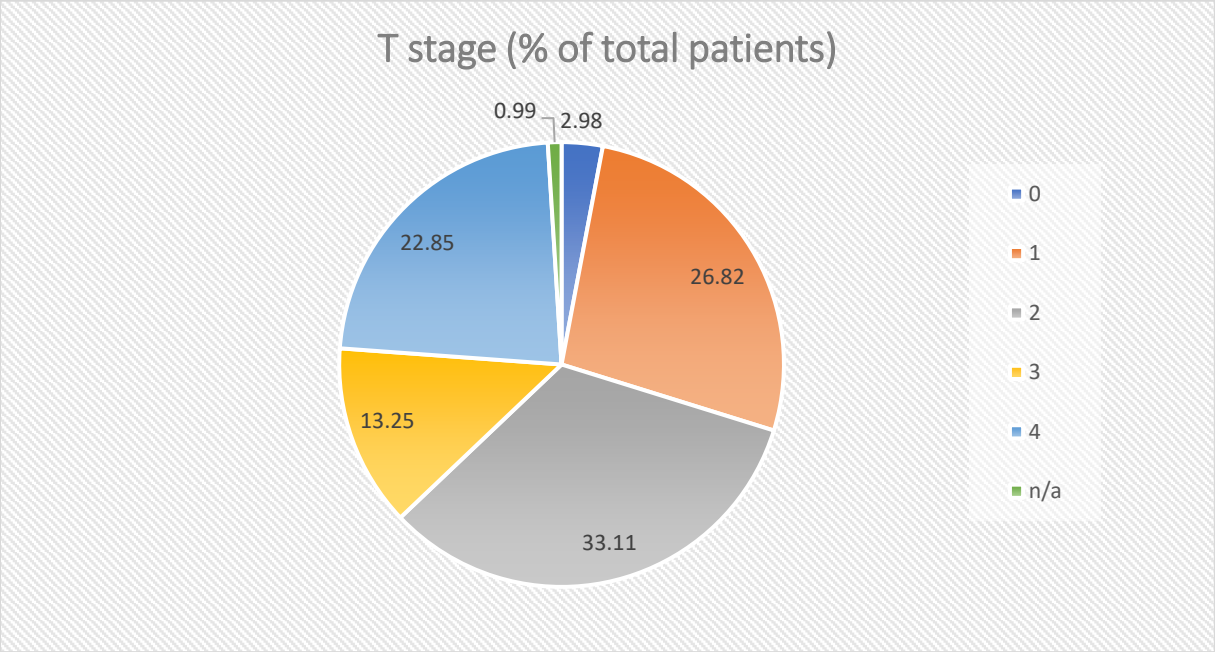


Table 27 below demonstrates the complexity of treatment options for HNC patients. Patients may receive one type of treatment or a combination, including treatments that are considered 'neoadjuvant' or 'adjuvant' to their primary curative treatment. As can be seen below, oncological treatment (radiotherapy +/- chemotherapy) was the most common primary treatment in the cohort. Radiotherapy treatment varies in intensity, with the 'Gray' (Gy) unit ionizing radiation dose figures below ranging from 55-75Gy, and treatment length ranging from 20-35 'fractions' (i.e., individual treatment sessions).

Characteristic	Value
Treatment type (Freq.)	
Neoadjuvant chemotherapy	98
Primary radiotherapy	142
Primary chemoradiotherapy	158
Primary surgery	96
Adjuvant radiotherapy	52
Adjuvant chemoradiotherapy	60
Number of primary treatment radiotherapy fractions (Freq.)	
20	8
25	2
30	169
33	3
35	14
Primary radiotherapy dose (Gy) (Freq.)	
55	9
60	5
64	2
65	161
70	17
75	1

Table 27: Oncological treatment detail

The subgroup of patients who had surgery as their primary treatment modality underwent either ENT or maxillofacial surgery, as illustrated in Table 28 below. 11 different types of surgery were represented in the cohort.

Characteristic	Value
Type of surgery (freq.)	
ENT	12
Maxillofacial	84
Type of resection (freq.)	
Wide local excision only	8
Hemiglossectomy	20
Subtotal glossectomy	3
Neck dissection	3
Wide local excision with flap reconstruction	6
Laser	7
Floor of mouth resection	7
Mandibulectomy	24
Maxillectomy	9
Palate resection	1
Partial glossectomy	7
Unknown	1
Number of patients requiring flap reconstruction	77

Table 28: Surgical treatment detail

Within this group, many patients (n=77/96; 80.2%) who underwent primary surgery required surgical tissue flap reconstruction, indicating more extensive surgery with the potential for higher functional impact.

5.1.2 MDADI data

Within NHS Lothian, SLT clinicians attempt to collect MDADI data with patients receiving curative treatment for HNC at pre-treatment and 6 months post completion of treatment timepoints.

MDADI Global and Composite scores

Table 29 shows the mean and range of MDADI scores, for both the Global (MDADI-G) and Composite (MDADI-C) scores, at the two timepoints. The minimum and maximum possible scores are 20 and 100 respectively for both MDADI-G and MDADI-C; lower scores represent lower dysphagia-related QoL. Six months post-treatment mean MDADI-G and -C scores were lower than pre-treatment mean scores.

Variables	Obs	Mean	Std. Dev.	Obs Min	Obs Max
Pretreatment MDADI-G	288	86.74	21.46	20	100
Pretreatment MDADI-C	285	85.99	15.71	23.16	100.00
6m posttreatment MDADI-G	178	74.8	26.56	20	100
6m posttreatment MDADI-C	162	73.09	17.76	28.4	100.00

Table 29: Mean MDADI-G and -C scores across datapoints

MDADI subscale data availability

The MDADI consists of three subscales: emotional, physical and functional, scores for which are calculated individually prior to calculating the overall MDADI-C score. Emotional subscale scores can range between 6-30, physical between 8-40, and functional between 5-25. Table 30 below shows a summary of data regarding subscale scores for the MDADI at both datapoints for patients where item level data was available. Means and ranges in Table 30 differ compared with the previous table as a smaller subgroup of patients had subscale level data available.

Variables	Obs	Mean	Std. Dev.	Obs Min	Obs Max
Pretreatment					
Emotional	275	26.04	4.78	6	30
Physical	273	33.74	7.33	8	40
Functional	275	22.09	3.69	8	25
Global	277	87.51	29.68	29	100
Composite	271	86.25	15.80	23.16	100
6 months post-treatment					
Emotional	148	22.77	5.47	8	30
Physical	147	28.39	7.12	15	40
Functional	146	19.55	4.65	7	25
Global	146	74.93	25.92	29	100
Composite	145	74.64	16.98	38.95	100

Table 30: Summary of MDADI subscale score data

As demonstrated in Table 30, more subscale data was available at the pre-treatment datapoint than at the 6 months post-treatment datapoint. This is in keeping with this data having been collected as part of routine clinical practice: in accordance with Scottish Government guidelines, 90% of patients should be seen by an SLT pre-treatment, reflecting the higher values. At the 6 month point the NHS Lothian SLT team collect data from as many patients as possible within the constraints of busy clinical practice, reflecting the lower values.

Table 30 shows that mean MDADI global, composite and subscale scores are lower at 6 months post-treatment compared with pre-treatment means. The standard deviation for all global and composite scores is large reflecting the variability of these scores across patients at both datapoints.

MDADI item level data availability

Item-level data was essential for the IRT quantitative analysis of the MDADI, as this allowed analysis of the tool at an item-by-item level. However, item level data must be accessed from the original questionnaire filled in by patients, and this was not physically available for all patients in the cohort for data collection purposes. Of the overall cohort of 302 patients, MDADI item level data was available for 271 patients at pre-treatment and for 145 patients at the 6-month post-treatment datapoint. 133 patients had item level data available at both datapoints.

Data error checking

416 MDADI questionnaires with item level data were available for this analysis. Individual questionnaires were located by the author and scores entered into a Microsoft Excel™ spreadsheet. Errors in data transfer from clinical records to databases are a known issue (Mays and Mathias, 2019). To assess the author's error rate in transcription from SLT clinical records to the research database in this study, the following procedure was carried out. Once all item scores for the 416 questionnaires had been inputted, 20% (n=83) were rechecked for data input errors. 83 MDADI questionnaires equates to 1660 individual items (83x20 items per questionnaire). During the error checking process, three item level errors were found, equalling a 0.18% error rate. This compares favourably with manual transcription error rates reported in the literature (Hong et al., 2013, Paulsen et al., 2012).

A separate but related issue is the potential for errors made by clinicians in original MDADI-C score calculations in the clinical notes accessed for MDADI data. MDADI-C score calculation = $(\left(\frac{x}{19}\right) x20)$ where $x =$ the summation of Likert scores for items 2-20. To monitor and correct potential errors, the research database included a formula to automatically calculate the MDADI-C score from item level data. This identified 103 errors in clinician calculation of MDADI-C scores (=24.76% error rate) which were subsequently corrected by the formula prior to quantitative analysis of MDADI-C data.

Missing data analyses

The number of patients with MDADI item level data available is lower than that for patients who have only a MDADI-G and/or MDADI-C score. This is due to the nature of the prospective data collection which occurred as part of routine clinical practice, hence paper copies were often disposed of once MDADI-G and -C scores were recorded meaning that item-level data was no longer accessible. The descriptive statistics summarising the data that were collected do not highlight any markedly unusual situations in terms of demographic data and are in keeping with what would be clinically expected across the two datapoints.

Missing data can be a significant issue when assessing the psychometric properties of a tool, potentially increasing standard error and biasing estimates of test performance (Kalkan et al., 2018). However, as IRT analyses tools at an item level rather than tool level, missing item level scores are ignored in IRT models, do not contribute to the analysis and do not require to be extrapolated or imputed.

The results of IRT test- and item-level analysis carried out to answer Research Questions 2 and 3 will now be presented.

5.1.3 Research Question 2 results

What are the results of an exploration of psychometric properties of the MDADI on a UK population using Item Response Theory?

As proposed in the Methods chapter of this thesis, the IRT quantitative analysis of MDADI data in this study will map to parameters set out in the COSMIN tool. Table 31 below outlines the analyses which will now be reported, and which aspects of IRT will be used to perform these analyses.

COSMIN Parameter	Definition	IRT analysis approach
Structural validity	Whether the dimensionality of the construct being measured by a PROM is adequately reflected by PROM tool scores	Proving unidimensionality of overall MDADI tool with GRM model fit at both datapoints
Internal consistency	Whether all items are interrelated	Calculating Test Information Function with standard error of measurement data for the overall MDADI tool at both datapoints

Table 31: COSMIN parameters identified for MDADI IRT analysis

The preceding qualitative analysis highlighted the potential for item redundancy in the MDADI. Therefore, results of IRT statistical analysis of MDADI item level data to inform potential item reduction will also be presented as per Table 32 below:

Potential item level issue	Rationale	IRT analysis approach
Item redundancy	potential to create shorter form of the MDADI as suggested in qualitative data	Taking the approach described by Sekely et al. (2018), considering a theoretical approach to item reduction, considering the information values of each MDADI item at both datapoints

Table 32: Additional IRT analysis of MDADI item-level data

IRT analysis

IRT uses the term 'latent trait', commonly termed 'θ', to describe the construct being assessed by the tool under analysis (Reise et al., 2005). In the case of the MDADI, θ = dysphagia-related QoL, where a higher θ value = better dysphagia-related QoL.

An IRT analysis produces data on different item-level properties. When considering responses to each item in a tool, data on three characteristics of items or a whole tool can be generated:

- **a** = item 'discrimination', or 'slope' parameter - that is the ability of an item to discriminate between respondents with low and high values of θ.
- **b** = item 'difficulty' or 'threshold' parameter – items with a high value of b are only endorsed at high levels of θ.

- **item ‘information’** = a metric that combines item or test discrimination and difficulty.

5.2 Structural Validity

The COSMIN study design checklist (Mokkink et al., 2019) states that analysis of structural validity is only relevant for PROMs that are based on a ‘reflective’ model, that is “a model in which all items are a manifestation of the same underlying construct” (p.9) rather than a ‘formative model’ which incorporates multiple constructs. Chen et al. (2001) do not use this ‘reflective/formative’ terminology in their MDADI origin paper. However, they do suggest that MDADI-C summated scores calculated from all subscales reflect "better day-to-day functioning and better QoL" (p.871); also, they state "the questions within the scale are consistently assessing the same issues" (p.875) which suggests the MDADI can be considered to be based on a reflective model.

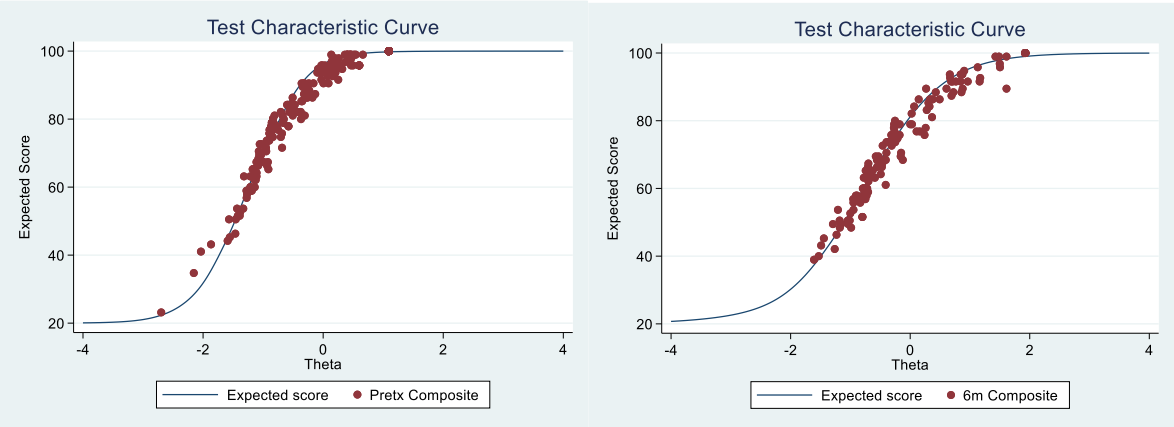
The COSMIN tool suggests factor analysis as the CTT method, and ‘model fit’ as the IRT method, of assessing structural validity. It is of note that Lin et al. (2022) have previously proven the unidimensionality of the MDADI via the CTT method of factor analysis. Analysis of structural validity with IRT is carried out through analysis of IRT model fit: that is, does the IRT analysis support the assertion that the MDADI overall, and its subscales, are unidimensional, or reflective, i.e., measuring one construct.

There are four assumptions made when applying the GRM to the analysis of a tool: that the tool is unidimensional, that the GRM is a good ‘fit’ for the data, that items in the tool are locally independent, and also ‘monotonic’ (Stover et al., 2019). These factors all provide evidence that a tool measures what it purports to measure, i.e., has structural validity.

As per the COSMIN tool, the structural validity of the MDADI at both datapoints will now be analysed with IRT by assessing the fit of the GRM to the tool.

Graded Response Model fit testing

Figure 11: Comparison of model θ to observed θ scores across datapoints



As illustrated by the above figure, there is a pattern of tight fit of model θ to observed θ scores on the MDADI at both pre-treatment and 6 months post-treatment datapoints, characterised by the close grouping of MDADI-C scores around the 'expected score' line on the graphs. This confirms the fit of the GRM model and the unidimensionality of the MDADI, and therefore its structural validity. Tabulated numeric detail of the Stata™ GRM model fit testing results can be viewed in Appendix H.

5.3 Internal Consistency

The COSMIN group (Prinsen et al., 2018) define internal consistency as "the degree of interrelatedness among the items" (p.1153) within a tool.

IRT analysis provides 'graphic representations' of item and tool properties by producing graphs for item-level and tool-level information and characteristics (Jean-Pierre et al., 2014). Item Information Functions (IIFs) are IRT graphs which demonstrate the precision and information of items. These can be summed to produce a Test Information Function (TIF) summating results for all items in a tool or test, enabling visualisation of which θ ranges are best covered by the tool with the least standard error of measurement. Standard error increases at the extremes of the range where the tool provides less information.

The COSMIN study design checklist (Mokkink et al., 2019) suggests that IRT analysis of a tool's internal consistency should take the form of considering the tool TIF and Standard Error (SE) data for θ. The SE indicates what proportion of information the test is providing is true; in the case of the MDADI this is information about dysphagia-related quality of life. Higher SE might indicate that more than one construct is being measured, however this should not be the case if the MDADI is truly unidimensional as previously discussed.

The TIFs in Figure 12 below show the MDADI test information and standard error plotted against θ for the whole test, at both pre-treatment and 6 months post-treatment datapoints.

Figure 12:TIFs for both datapoints

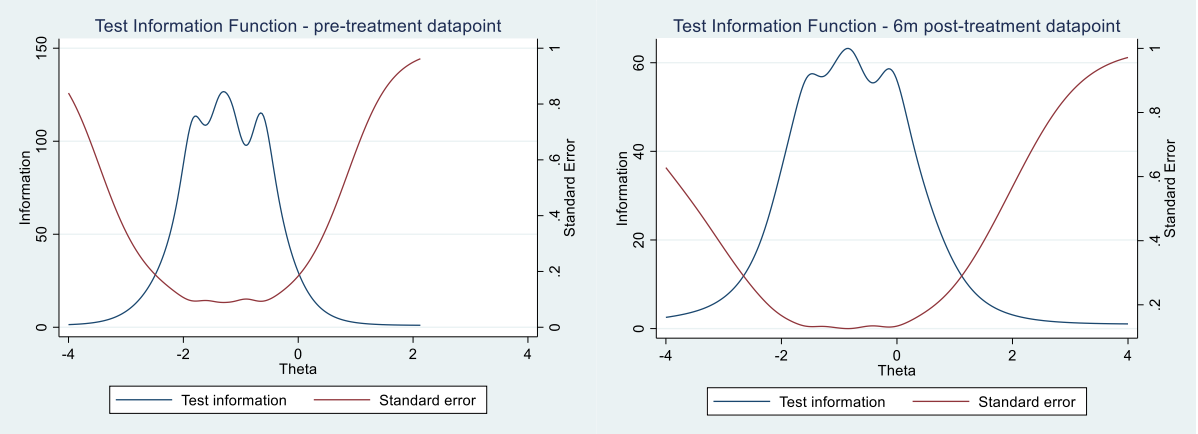


Figure 12 illustrates that at the pre-treatment datapoint, the MDADI provides almost no information for θ > 1, with most information in the range of θ = -2 – 0, whereas at 6 months post-treatment, this range is closer to -2 - +1. This reflects a potential ceiling effect in the tool when used at the pre-treatment stage. As can be seen on the calibration of the Y axes, the overall information provided by the MDADI is higher at the pre-treatment datapoint than at 6 months post-treatment.

	Standard Error at Test level over different values of θ						
Datapoint	**θ = -3**	**θ = -2**	**θ = -1**	**θ = 0**	**θ = 1**	**θ = 2**	**θ = 3**
Pre-treatment	.3418391	.1066665	.0987697	.1838407	.2464923	.9498423	Not estimable
6 months post-treatment	.3772944	.1664111	.1274631	.1336228	.2614183	.5677811	.8610921

Table 33: Summed MDADI Standard Error values

As demonstrated in Table 33, IRT analysis allows exploration of how a test varies in precision over different values of θ. This is a key difference between CTT and IRT, where CTT assumes constant SE within a test, but IRT can perform a more nuanced assessment. As the pre-treatment and post treatment data are being analysed separately the θ values are scaled differently. That is, the θ values at the two time points are not directly comparable. θ=0 is the average condition-specific dysphagia-related QoL at that time point.

When viewed in conjunction with the graphic representation of SE in Figure 12 above, it is evident that in comparison with the pre-treatment datapoint, the MDADI has considerably lower information and precision at the 6-month post-treatment datapoint than the pre-treatment datapoint. The TIF shows that at this 6-month post-treatment timepoint the MDADI is most precise for a range of θ of approximately -2 to 1.

At the pre-treatment datapoint there is less variability/SE for higher θ compared with the 6 months post-treatment datapoint. This is due to a ceiling-type effect as there are many patients bunched together at this datapoint with high θ – in keeping with the fact that many

patients do not have a dysphagia-related QoL impairment prior to their treatment commencing, as they may be asymptomatic or minimally symptomatic. At the 6 months post-treatment point scores are shifted towards lower θ, as at this point the population has lower dysphagia related QoL. SE is lowest for θ of approximately -1 - -2, meaning the test at this timepoint is most accurate for negative θ, that is poorer dysphagia-related QoL.

The IRT analysis indicates the MDADI does not provide information on people with θ>2 in the pre-treatment period, so SE is not estimable. That is, the MDADI does not differentiate between extremely good and very good dysphagia related QoL in the pre-treatment period.

In addition to TIF and SE information, the COSMIN user manual also suggests that a 'reliability coefficient of estimated latent trait value' can be calculated to illustrated internal consistency (Mokkink et al., 2018). A latent trait value reliability coefficient is an IRT equivalent to a CTT Cronbach's alpha metric. Table 34 below displays these coefficients for the current data. These are high values which indicate good internal consistency.

Datapoint	**Empirical reliability coefficient value**
Pre-treatment	0.860
Post-treatment	0.943

Table 34: MDADI empirical reliability coefficient values

The more detailed assessment of SE that IRT facilitates, and the complex results presented here with respect to the MDADI, show that a tool's internal consistency measurement is not straightforward. The MDADI functions differently at different timepoints and across different levels of θ. This would not be apparent with a simple Cronbach's alpha score alone, which is the CTT equivalent to this IRT analysis. CTT analysis therefore misses the nuance demonstrated here, that is that tools can vary in their internal consistency across different levels of the latent trait being assessed, and at different timepoints in a patient's healthcare journey.

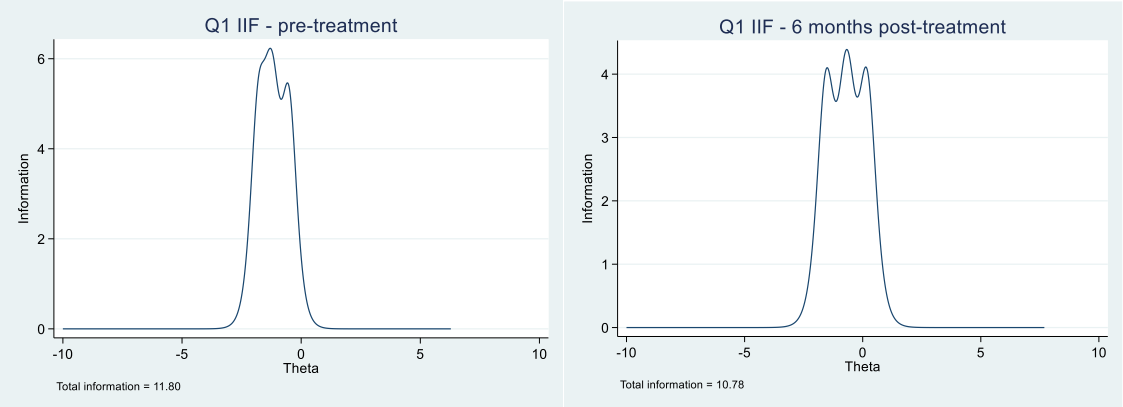
5.4 Item redundancy

An investigation of item redundancy and potential for item reduction was not an a priori plan of this study, but rather something that emerged as worth pursuing during the qualitative data analysis. As this was the case, following qualitative data analysis I interrogated the literature for approaches to item redundancy analysis and the process of item reduction, and that of Sekely et al. (2018), discussed in the next section, seemed the most well-considered and replicable and was chosen as the model to follow.

The qualitative data generated by this study identified the total number of items in the MDADI as a negative aspect of the tool, impacting on its clinical utility. Therefore, IRT analysis of potential item redundancy was carried out to ascertain which items had weaker properties and therefore would be 'top of the list' if looking to remove items to make a shorter tool. In addition, the impact on overall tool information resulting from item removal was explored.

Item information is the focus of this analysis, as both discrimination and difficulty feed into this metric. Item information was calculated for MDADI items at both datapoints. Item Information Functions (IIFs) illustrate the information each item provides. An example of these is given in Figure 13 below:

Figure 13: IIFs for MDADI item 1 at both datapoints



As demonstrated in Figure 13, the total information for item 1 at the pre-treatment datapoint is 11.80, and 10.78 at the 6 months post-treatment datapoint. Table 35 below shows the results of all the item information calculations, for both datapoints.

	Item information			
Item	**Pre-treatment**	**% of total test information**	**6 months post-treatment**	**% of total test information**
1	11.80	4.92	10.78	6.33
2	8.59	3.58	10.12	5.94
3	7.27	3.03	6.85	4.02
4	8.50	3.54	5.83	3.42
5	4.02	1.68	4.05	2.38
6	10.22	4.26	9.27	5.44
7	10.09	4.21	9.13	5.36
8	19.18	8.00	8.80	5.17
9	9.20	3.84	5.79	3.40
10	7.50	3.13	3.72	2.18
11	10.79	4.50	7.63	4.48
12	15.94	6.65	9.12	5.35
13	10.75	4.48	4.03	2.36
14	20.65	8.62	15.87	9.32
15	5.97	2.49	5.67	3.33
16	9.25	3.86	6.96	4.09
17	11.16	4.67	7.43	4.36
18	31.38	13.09	15.62	9.17
19	7.23	3.02	4.10	2.41
20	20.20	8.43	19.57	11.49
Total information	**239.69**		**170.34**	

Table 35: Item information values

Table 35 demonstrates that individual items within the MDADI have greatly varying information values, and that the overall test information differs greatly between datapoints. At the pre-treatment datapoint the item information range is 4.02 – 20.65, whilst at the 6 months post-treatment datapoint the total information per item range is 3.72 – 19.57. As the table indicates that MDADI items provide different levels of information at different timepoints, datapoints were examined separately in terms of item profile and candidacy for removal when considering item redundancy.

Empirical vs Theoretical item reduction

Sekely et al. (2018) suggest consideration of both ‘theory-driven’ and ‘empirically-driven’ approaches when choosing items that could be dropped to form a shorter scale. A ‘theory-driven’ approach maintains the subscale structure of a tool. The rationale for this is that the distinction between subscale content was considered sufficiently significant during tool inception to incorporate into the scale, and therefore should be maintained. Conversely, an ’empirically-driven’ approach considers all items across the tool, dropping the weakest items irrespective of subscale membership. In an empirically driven approach, it would be technically possible to ‘lose’ an entire subscale if the comprising items were found to be weaker.

As previously discussed, the MDADI is divided into 3 separate subscales: emotional, physical and functional. Each subscale has a different number of items. Therefore, if taking a

theoretical approach aiming to maintain proportional subscale representation, subscales can be weighted in terms of overall tool structure, and the number of items allocated accordingly. As Table 36 below indicates the items that comprise each subscale, and detail of how many items per subscale would be included in a 5-item theoretically driven version of the MDADI, weighted by percentage subscale makeup of the overall tool.

Subscale	Item numbers	Total	% of total	No. of items in 5 item theoretical scale
Emotional	2,5,6,8,12,18	6	31.6	2
Physical	4,7,10,11,13,16,17,19	8	42.1	2
Functional	3,9,14,15,20	5	26.3	1

Table 36: MDADI subscale item ratios

A 5-item shortened version of the MDADI was generated for comparison with the extant literature on potential for MDADI item reduction (Lin et al., 2022). A theory-driven approach only was taken given the structure of the MDADI. As previously presented data demonstrates that the properties of the MDADI differ between the pre-treatment and 6 months post-treatment datapoints, item redundancy/removal was considered separately for each datapoint.

Item redundancy – pre-treatment datapoint

At the pre-treatment datapoint, the ranking of items by total information is as demonstrated in Table 37 below.

	Pre-treatment
Highest – lowest information item rank order	18, 14, 20, 8, 12, 1, 17, 11, 13, 6, 7, 16, 9, 2, 4, 10, 3, 19, 15, 5

Table 37: Ranking of MDADI items by total information at pre-treatment datapoint

The items have been colour-coded to illustrate where items from each subscale fall within the ranking: yellow for the Emotional subscale, blue for the Physical subscale and green for the Functional subscale. Item 1 is the 'Global' question and is not incorporated into any subscale.

Each subscale will now be considered individually with respect to information values for its constituent items.

Emotional subscale

Figure 14: Pre-treatment emotional subscale items IIFs comparison

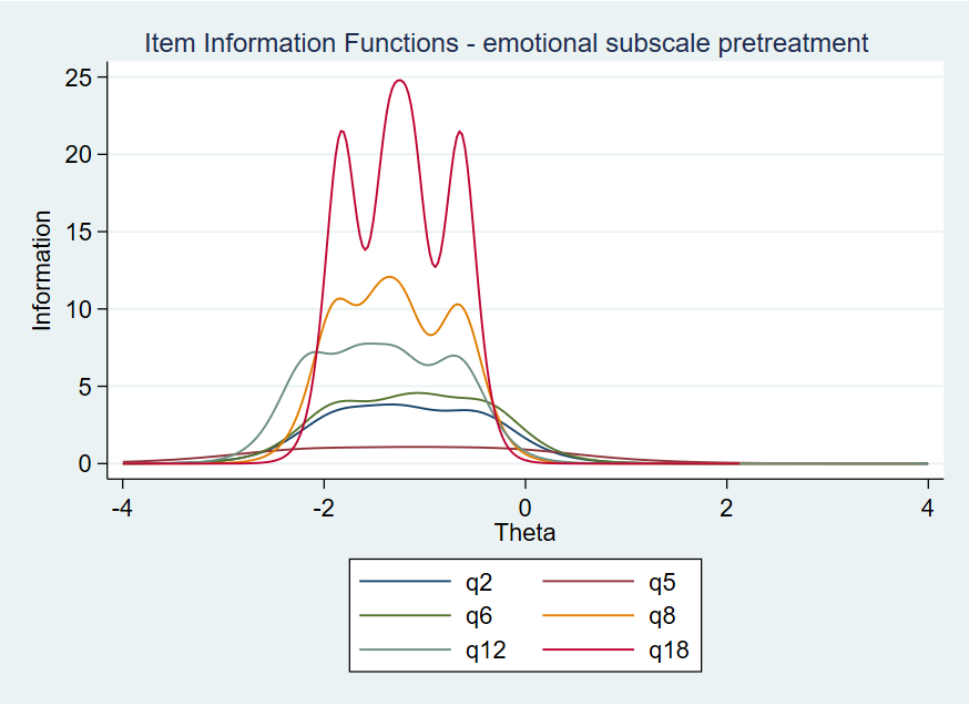


Figure 14 above shows the superimposed IIFs for the items which comprise the emotional subscale. It illustrates that items 18, 8 and 12 have the highest discrimination and information of the six emotional subscale items. Although item 18 has the highest discrimination and information, it covers a narrower range of -θ than other items. Table 38 below shows the ranking of items by their information values.

Rank	Item	Content	Information value
1	18	I have low self-esteem because of my swallowing problem	31.38
2	8	I do not go out because of my swallowing problem	19.18
3	12	Other people are irritated by my eating problem	15.94
4	6	I am upset by my swallowing problem	10.22
5	2	I am embarrassed by my eating habits	8.59
6	5	I do not feel self-conscious when I eat	4.02
		Total information	**89.33**

Table 38: Pre-treatment emotional subscale sorted by information

Physical subscale

Figure 15: Pre-treatment physical subscale items IIFs comparison

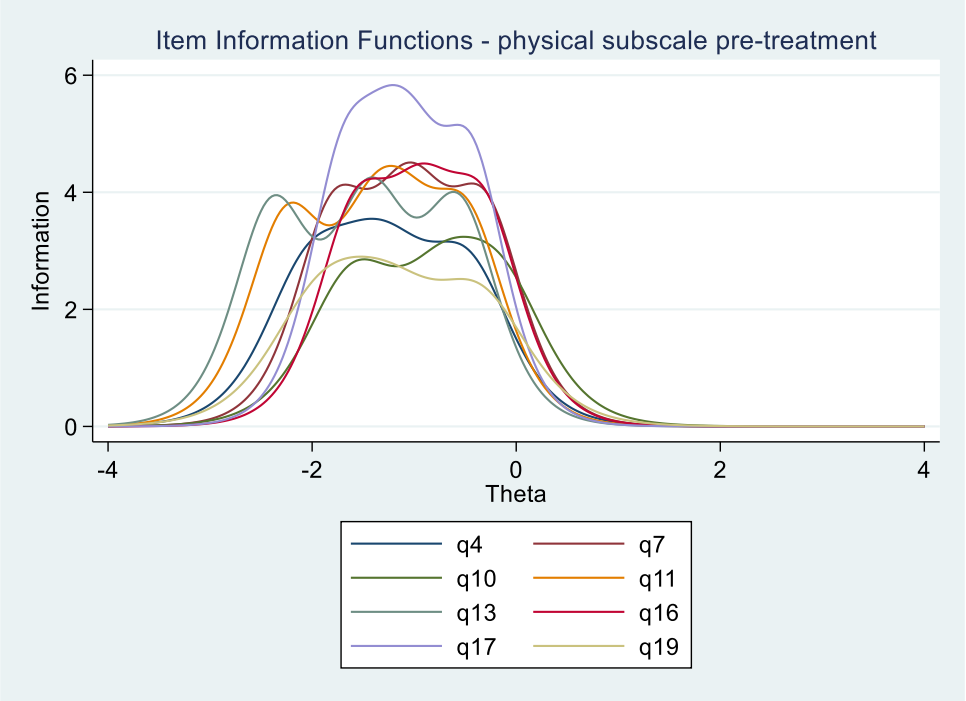


Figure 15 above shows that item 17 has higher information and discrimination than other items in the physical subscale at the pre-treatment datapoint, but that it also covers a narrower range of difficulty in the -θ range. Table 39 below shows the ranking of items by their information values.

Rank	Item	Content	Information value
1	17	I cannot maintain my weight because of my swallowing problem	11.16
2	11	People ask me 'why can't you eat that?'	10.79
3	13	I cough when I try to drink liquids	10.75
4	7	Swallowing takes great effort	10.09
5	16	I limit my food intake because of my swallowing difficulty	9.25
6	4	Swallowing is more difficult at the end of the day	8.50
7	10	It takes me longer to eat because of my swallowing problem	7.50
8	19	I feel that I am swallowing a huge amount of food	7.23
		Total information	**75.27**

Table 39: Pre-treatment physical subscale sorted by information

Functional subscale

Figure 16: Pre-treatment functional subscale items IIFs comparison

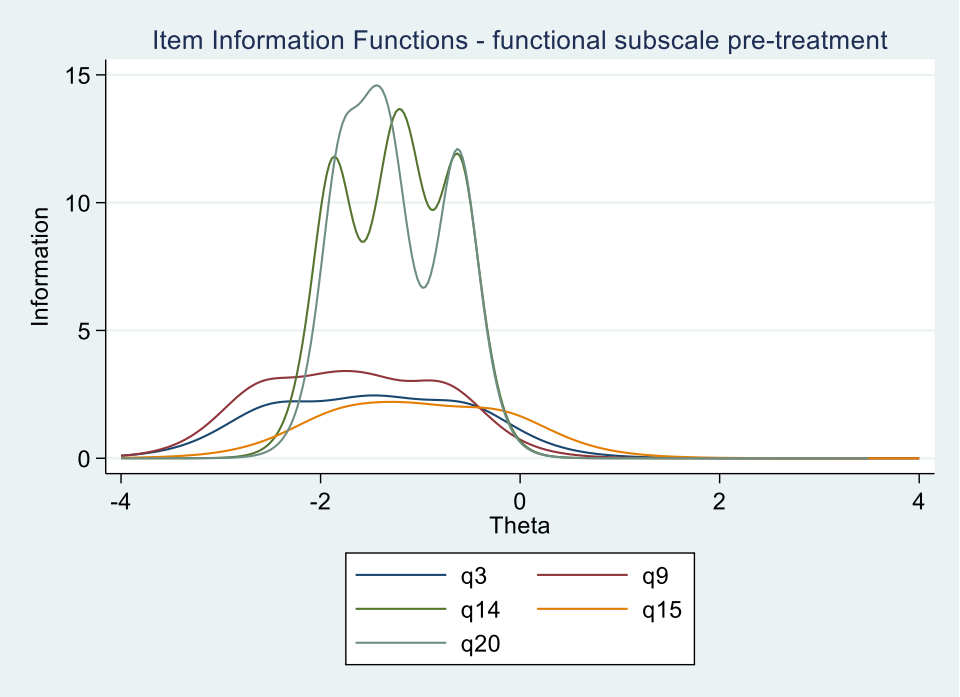


Figure 16 above shows that items 14 and 20 have higher information and discrimination than other items in the functional subscale at the pre-treatment datapoint; however, they both cover a narrower range of difficulty in the -θ range. Table 40 below shows the ranking of items by their information values.

Rank	Item	Content	Information value
1	14	My swallowing problems limit my personal and social life	20.65
2	20	I feel excluded because of my eating habits	20.20
3	9	My swallowing difficulty has caused me to lose income	9.20
4	3	People have difficulty cooking for me	7.27
5	15	I feel free to go out to eat with my friends, neighbours, relatives	5.97
		Total information	**63.29**

Table 40: Pre-treatment functional subscale sorted by information

Considering the information values presented above for items in each of the subscales, taking into account the item ratios, theoretically driven item selection for shortened MDADI forms would be as displayed in Table 41 for the pre-treatment datapoint:

Items	5 item scale
8	I do not go out because of my swallowing problem
11	People ask me 'why can't you eat that?'
14	My swallowing problems limit my personal and social life
17	I cannot maintain my weight because of my swallowing problem
18	I have low self-esteem because of my swallowing problem

Table 41: Content of item-reduced MDADI for pre-treatment

Item redundancy – 6 months post-treatment datapoint

At the 6 months post-treatment datapoint, the ranking of MDADI items in terms of total information is as displayed in Table 42:

	6 months post-treatment
Highest – lowest information item rank order	20, 14, 18, 1, 2, 6, 7, 12, 8, 11, 17, 16, 3, 4, 9, 15, 19, 5, 13, 10

Table 42: Ranking of MDADI items by total information at 6 months post-treatment datapoint

The items have been colour-coded to illustrate where items from each subscale fall within the ranking: yellow for the Emotional subscale, blue for the Physical subscale and green for the Functional subscale. Item 1 is the 'Global' question and is not incorporated into any subscale.

Each subscale will now be considered individually with respect to information values for its constituent items.

Emotional subscale

Figure 17: 6-month post-treatment emotional subscale items IIFs comparison

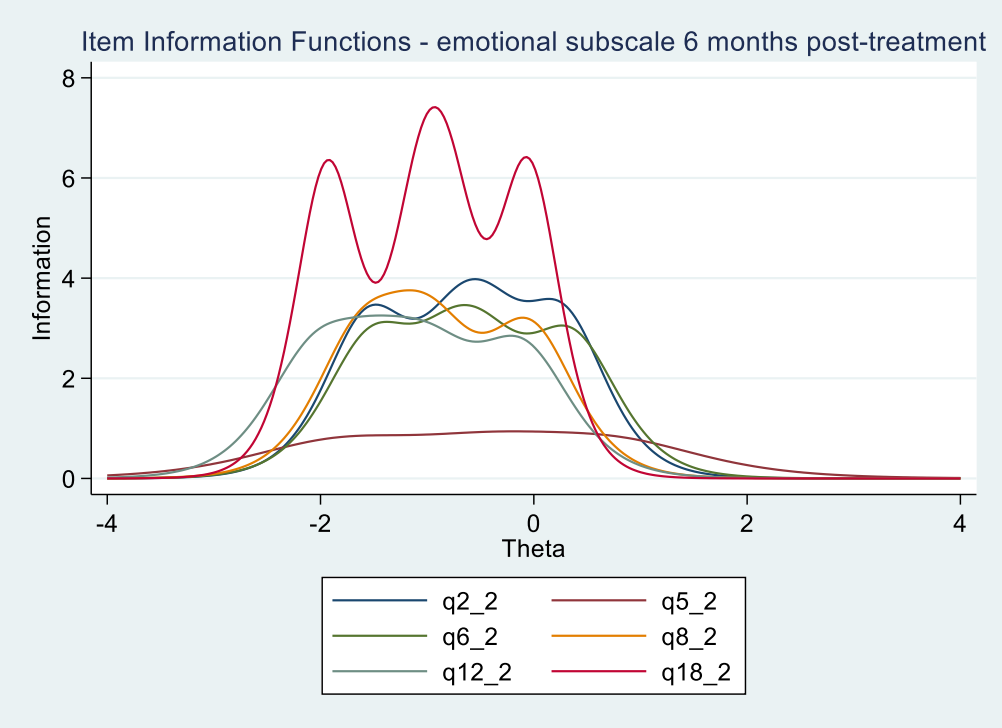


Figure 17 above shows that item 18 has higher information and discrimination than other items in the emotional subscale at the 6 months post-treatment datapoint. Item 5 has considerably lower information and discrimination than other items in this subscale at this datapoint, although it covers a wider range of θ than other items, particularly with respect to +θ values in comparison with item 18. Table 43 below shows the ranking of items by their information values.

Rank	Item	Content	Information value
1	18	I have low self-esteem because of my swallowing problem	15.62
2	2	I am embarrassed by my eating habits	10.12
3	6	I am upset by my swallowing problem	9.27
4	12	Other people are irritated by my eating problem	9.12
5	8	I do not go out because of my swallowing problem	8.80
6	5	I do not feel self-conscious when I eat	4.05
		Total information	**56.98**

Table 43: 6 months post-treatment emotional subscale sorted by information

Physical subscale

Figure 18 below shows that items 19, 10 and 13 have considerably lower information and discrimination than other items in the physical subscale at the 6-month post-treatment datapoint. Item 7 has the highest information of all items. Items 7, 16 and 17 cover a narrower range of θ than other items.

Figure 18: 6-month post-treatment physical subscale items IIFs comparison

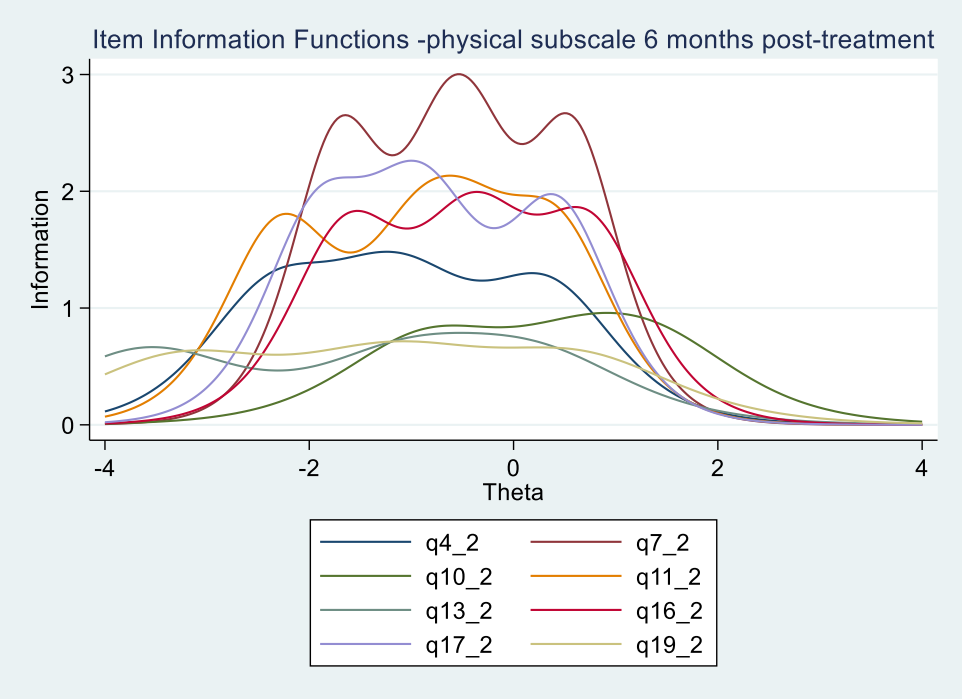


Table 44 below shows the ranking of items by their information values.

Rank	Item	Content	Information value
1	7	Swallowing takes great effort	9.13
2	11	People ask me ‘why can’t you eat that?’	7.63
3	17	I cannot maintain my weight because of my swallowing problem	7.43
4	16	I limit my food intake because of my swallowing difficulty	6.96
5	4	Swallowing is more difficult at the end of the day	5.83
6	19	I feel that I am swallowing a huge amount of food	4.10
7	13	I cough when I try to drink liquids	4.03
8	10	It takes me longer to eat because of my swallowing problem	3.72
		Total information	**48.83**

Table 44: 6 months post-treatment physical subscale sorted by information

Functional subscale

Figure 19: 6-month post-treatment functional subscale items IIFs comparison

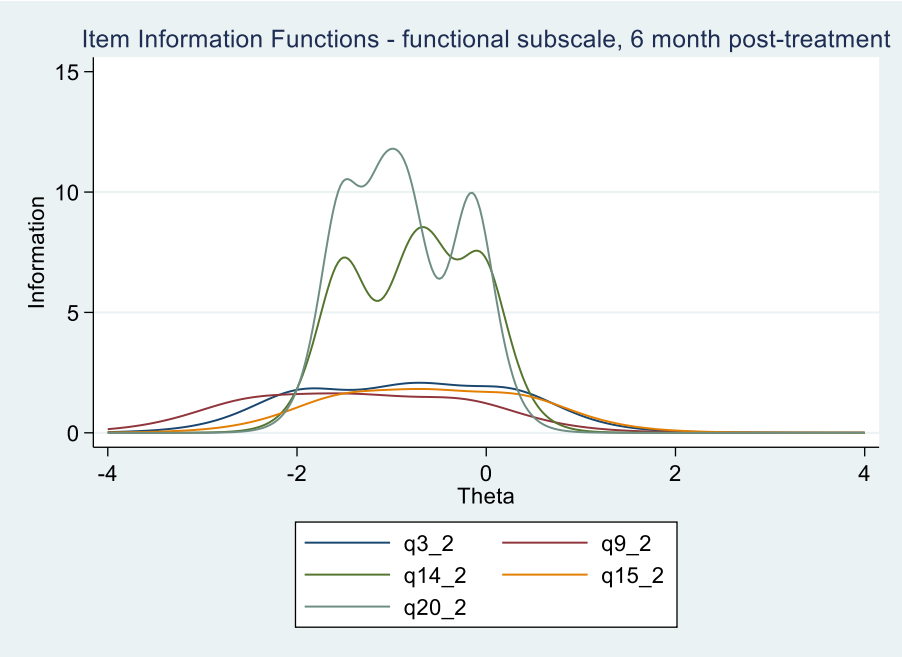


Figure 19 above shows that items 14 and 20 have higher information and discrimination than other items in the functional subscale at the 6 months post-treatment datapoint. Items 14 and 20 cover a narrower range of θ than other items in this subscale at this datapoint.

Table 45 below shows the ranking of items by their information values.

Rank	Item	Content	Information value
1	20	I feel excluded because of my eating habits	19.57
2	14	My swallowing problems limit my personal and social life	15.87
3	3	People have difficulty cooking for me	6.85
4	9	My swallowing difficulty has caused me to lose income	5.79
5	15	I feel free to go out to eat with my friends, neighbours, relatives	5.67
		Total information	**53.75**

Table 45: 6 months post-treatment functional subscale sorted by information

Taking into account the item ratios, theoretically driven item selection for a shortened MDADI form would be as follows for the 6-month post-treatment datapoint:

Items	5 item scale
2	I am embarrassed by my eating habits
7	Swallowing takes great effort
11	People ask me 'why can't you eat that?'
18	I have low self-esteem because of my swallowing problem
20	I feel excluded because of my eating habits

Table 46: Content of item-reduced MDADI for 6 months post-treatment

Summary of theoretically driven MDADI shortforms

Table 47 below outlines the results of MDADI item reduction taking a theoretical approach, for both datapoints, based on item information scores for the 19 MDADI composite items. The highest scoring items from each subscale as per the above ratio weightings were selected. Items are presented in numerical order rather than ranking.

Scale length	Pre-treatment	Sum of information	% Information of test total	6m post-treatment	Sum of information	% Information of test total
5 item theoretical	8, 11, 14, 17, 18	93.16	40.88%	2, 7, 11, 18, 20	62.07	38.90%

Table 47: 5 item 'theoretical' MDADI across datapoints

As demonstrated in Table 47, there are differences across timepoints for proposed scale content providing evidence for using different versions of the MDADI at pre-treatment and post-treatment timepoints.

In terms of retained test information content, the 5-item scales retain a greater proportion of information than their proportion of items, that is, 25% of total items provide more than 25% test information.

For this study, final suggestions for item redundancy/removal will be presented later in this section, following presentation of data concerning DIF. At that point, results of qualitative and quantitative data analysis will be synthesised to generate overall recommendations for MDADI item reduction.

5.5 Research Question 3

What are the potential factors which might result in differential item functioning?

The COSMIN tool, in the section considering 'measurement invariance', encourages statistical assessment of Differential Item Functioning (DIF) when using IRT methods (Mokkink et al., 2019). IRT allows individual items within a tool to be analysed for DIF, that is whether respondents with the same level of θ respond differently to a specific item due to the influence of other variables or characteristics. The qualitative data generated in this study indicated that there was potential for DIF in multiple items of the MDADI.

5.5.1 Identification of variables potentially associated with DIF

The preceding qualitative analysis generated data that informed and drove analysis of DIF in the MDADI. Firstly, group characteristics/variables that might result in DIF were identified by questionnaire respondents. These were coded as such in NVIVO™; for DIF analysis the NVIVO datafile was revisited to ensure all items identified as being potential sources for DIF were included in the analysis. In total, 6 MDADI items were flagged as being potential sources of DIF: items 2, 3, 8, 9, 18 and 20.

The variables identified by respondents as being potential sources of DIF were:

- **sex:** the patient's sex might influence their response
- **age**: whether the patient was of working age or not might influence their response
- **socioeconomic status:** the patient's socioeconomic status might influence their response

Table 48 below displays the specific items flagged as potentially susceptible to DIF in the qualitative analysis, with quotes from questionnaire respondents evidencing this selection, and how these variables were coded for the quantitative DIF analysis.

Variable	Items investigated for DIF with qualitative rationale		Variable coding
Sex	**2**	**I am embarrassed by my eating habits** "*Usually only finding an impact for ladies or those who actively engage in social eating. Many of our male patients do not report feelings of embarrassment*" (Respondent 15)	Male Female
	18	**I have low self-esteem because of my swallowing problem** "*Difficult question and not well received by the "blokes" on our caseload, many of whom do not actively discuss mood/self-esteem with their therapists*" (Respondent 15)	
	20	**I feel excluded because of my eating habits** "*Can be helpful, for women especially*" (Respondent 4)	
Age	**9**	**My swallowing difficulty has caused me to lose income** "*The predominance of over 60s or 65s in the HNC population tends to make this item slightly less relevant*." (Respondent 12) "*Not relevant to retired patients or patients who don't work?*" (Respondent 20) "*Not always relevant for patients who have retired or on benefits.*" (Respondent 22) "*useful for working age*" (Respondent 29)	Working age (≤60y for females, ≤65y for males) Retirement age or more (>60y for females, >65y for males)
Socioeconomic status	**3**	**People have difficulty cooking for me** "*This varies in relevance, depending on socio economic background.*" (Respondent 22)	SIMD used as a surrogate for socioeconomic status SIMD 1-3 = higher area-level deprivation SIMD 4-10 =lower area-level deprivation
	8	**I do not go out because of my swallowing problem** "*Can see the potential use but does not often highlight difficulties in our clinical caseload who often cannot afford to eat out*" (Respondent 15)	

Table 48: Items investigated for DIF with qualitative data rationale

5.5.2 DIF analysis

The presence of DIF was analysed for these six items at both the pre-treatment and 6-months post-treatment datapoints as per the guidance in the Stata IRT manual (Stata Press, 2021). This analysis allows for calculation of whether item discrimination and difficulty differ significantly between identified groups. This potential difference can be visually represented by Item Characteristic Curves (ICCs), and they have been included in the following analysis for illustration. In addition, a likelihood-ratio test of significance, to indicate whether any difference found between groups reaches statistical significance, has been calculated and

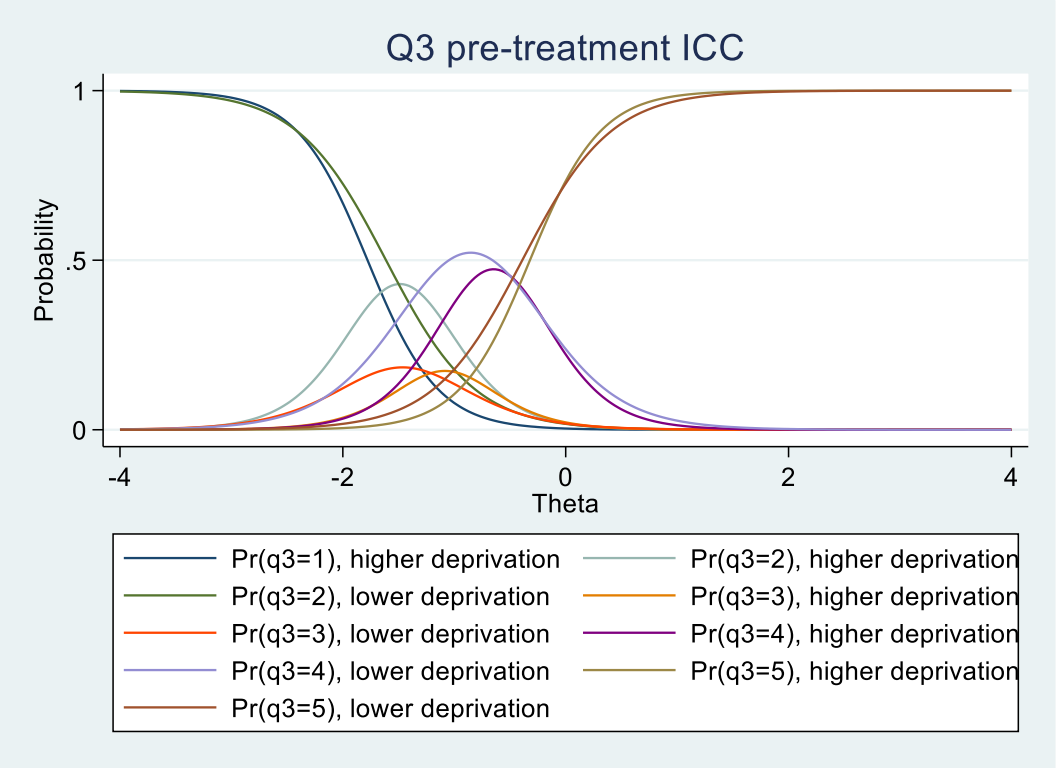
presented. Statistical significance was taken as p≤0.05 for these analyses. What follows are the results for DIF analysis for items 2, 3, 8, 9, 18 and 20 of the MDADI, grouped by pre-treatment and 6 months post-treatment datapoints.

Pre-treatment datapoint

SIMD

Item 3

Figure 20: ICC for item 3 pre-treatment



As illustrated in Figure 20, when analysed by the two different SIMD groupings of 1-3 and 4-10, responses to item 3 can be seen to differ.

The results of the likelihood ratio test for item 3 demonstrating DIF = 26.84, with p<0.001. Therefore, the null hypothesis of no DIF for item 3 at the pre-treatment datapoint can be rejected. The conclusion can therefore be made that item 3 of the MDADI at the pre-treatment datapoint exhibits DIF for the variable of SIMD category.

Item 8

Figure 21: ICC for item 8 pre-treatment

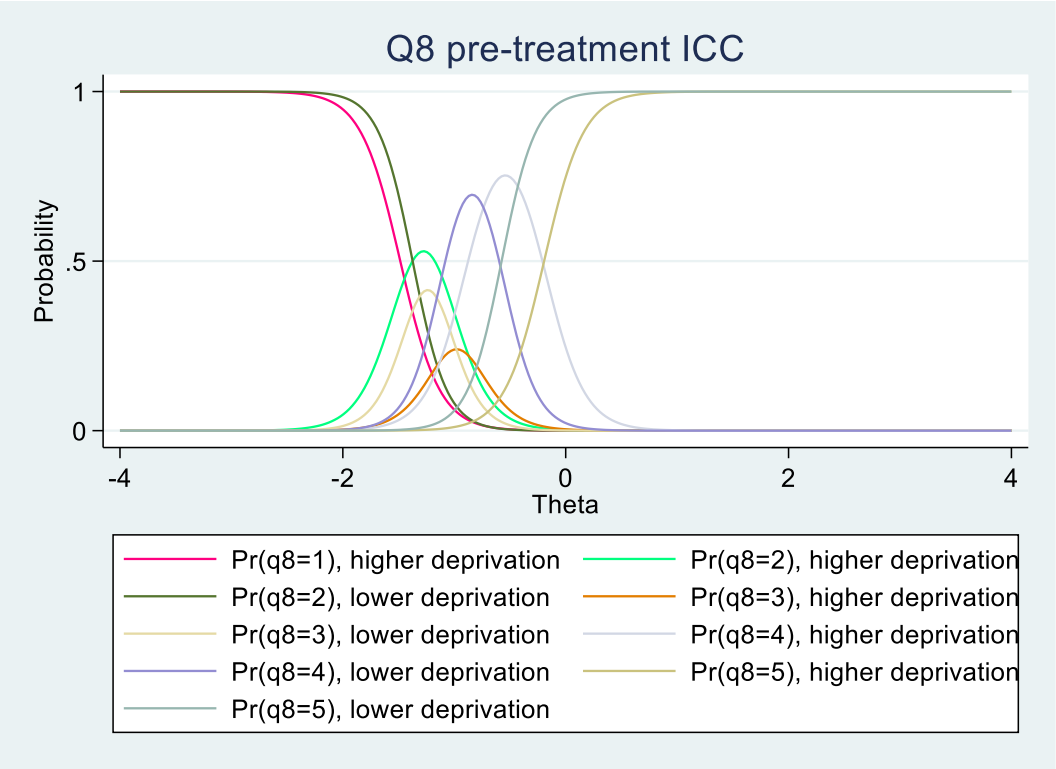


Figure 21 demonstrates the difference in responses to item 8 between the two SIMD groups.

The result of the likelihood ratio test for item 8 demonstrating DIF = 17.77 with p=0.0014. Therefore, the null hypothesis of no DIF for item 8 at the pre-treatment datapoint can be rejected. The conclusion can therefore be made that item 8 of the MDADI at the pre-treatment datapoint exhibits DIF for the variable of SIMD category.

Age

Item 9

Initial DIF analysis for item 9 failed, as for patients with 'age_cat=1' (i.e., non-working age), there were no item-level scores of <3 for this item. This then meant that location parameter b (i.e., difficulty) for categories 1 and 2 was not estimable for the group age_cat=1. The DIF analysis therefore failed because there is no comparison for the values obtained for age_cat=0. To overcome this issue, the analysis was re-run with the 7 cases where item 9 is scored as 1 or 2 dropped. It is important to note that taking this course *underestimates* the amount of DIF in item 9 as the most extreme cases have been dropped.

Figure 22: ICC for item 9 pre-treatment

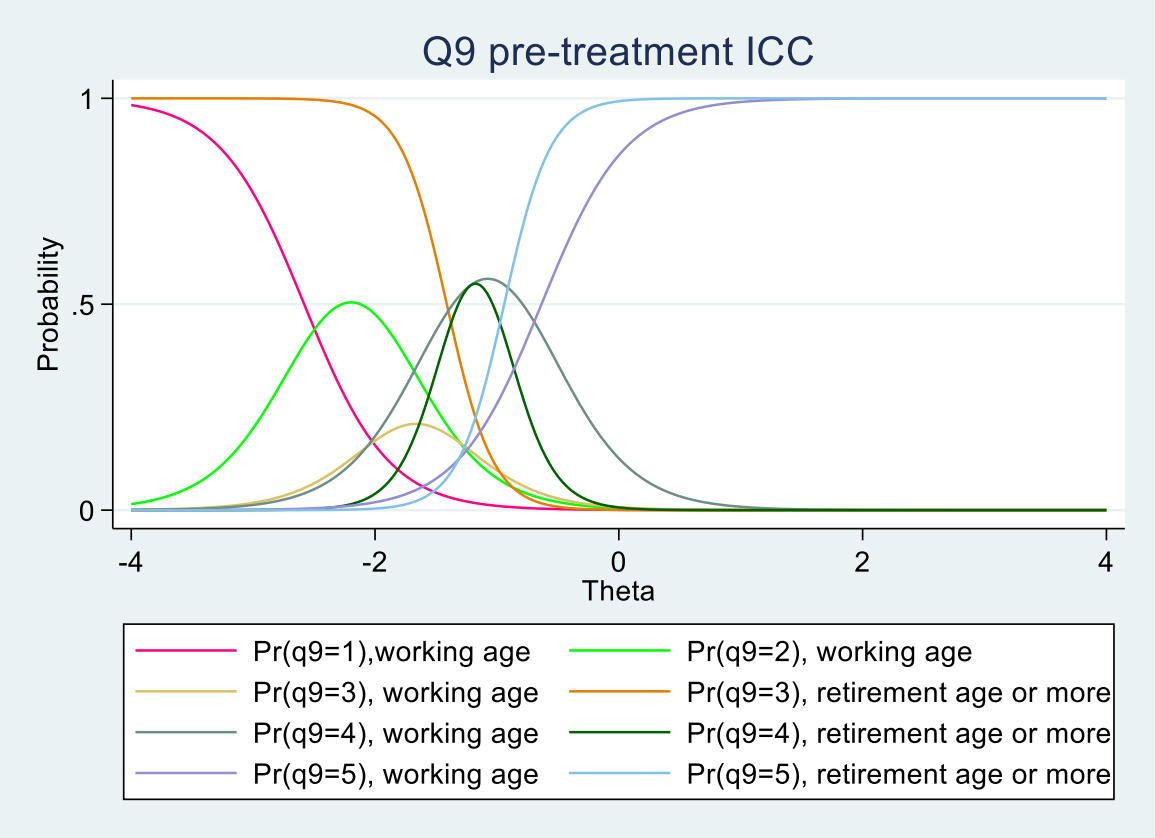


Figure 22 depicts the differences in responses between the two age groups analysed. The figure displays the results for the whole group rather than the reduced analysis group, and illustrates the issue described above: there are more lines plotted for patients of working age, as no patient of retirement age or more endorsed 'strongly agree' or 'agree' on this item ("My swallowing difficulty has caused me to lose income") at this datapoint.

The result of the likelihood ratio test for item 9 demonstrating DIF = 10.61, with p=0.014. Therefore, the null hypothesis of no DIF for item 9 at the pre-treatment datapoint can be rejected. The conclusion can therefore be made that item 9 of the MDADI at the pre-treatment datapoint exhibits DIF for the variable of working vs non-working age, and that

given the nature of the analysis, the true extent of DIF for the item will be higher than calculated here.

Sex

Item 2

Figure 23: ICC for item 2 pre-treatment

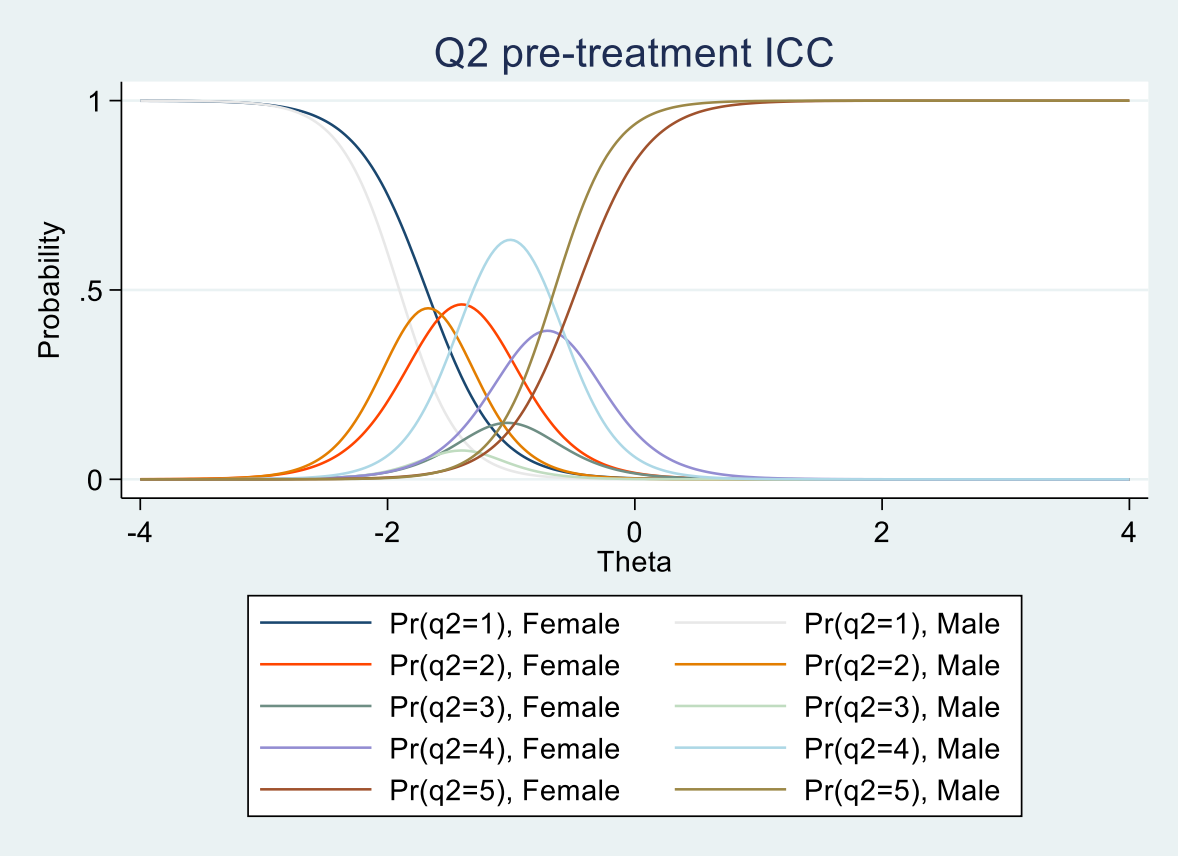


Figure 23 highlights the difference in responses between the male and female groups for item 2 pre-treatment.

The result of the likelihood ratio test for item 2 demonstrating DIF = 15.14, with p=0.0098. Therefore, the null hypothesis of no DIF for item 2 at the pre-treatment datapoint can be rejected. The conclusion can therefore be made that item 2 of the MDADI at the pre-treatment datapoint exhibits DIF for the variable of sex.

Item 18

Figure 24: ICC for item 18 pre-treatment

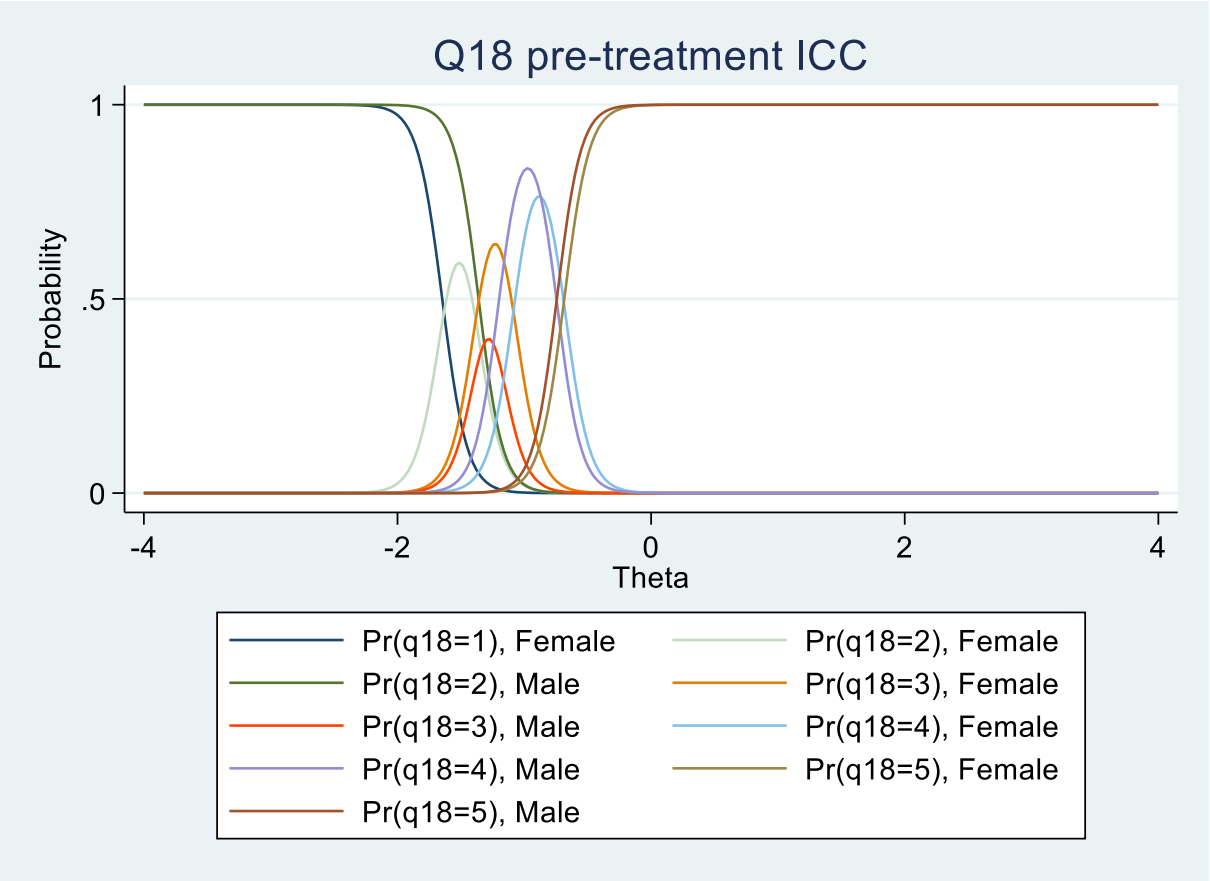


Figure 24 illustrates the difference in responses between the male and female groups for item 18 pre-treatment.

The results of the likelihood ratio test for item 18 demonstrating DIF = 27.09, with $p<0.01$. Therefore, the null hypothesis of no DIF for item 18 at the pre-treatment datapoint can be rejected. The conclusion can therefore be made that item 18 of the MDADI at the pre-treatment datapoint exhibits DIF for the variable of sex.

Item 20

Figure 25: ICC for item 20 pre-treatment

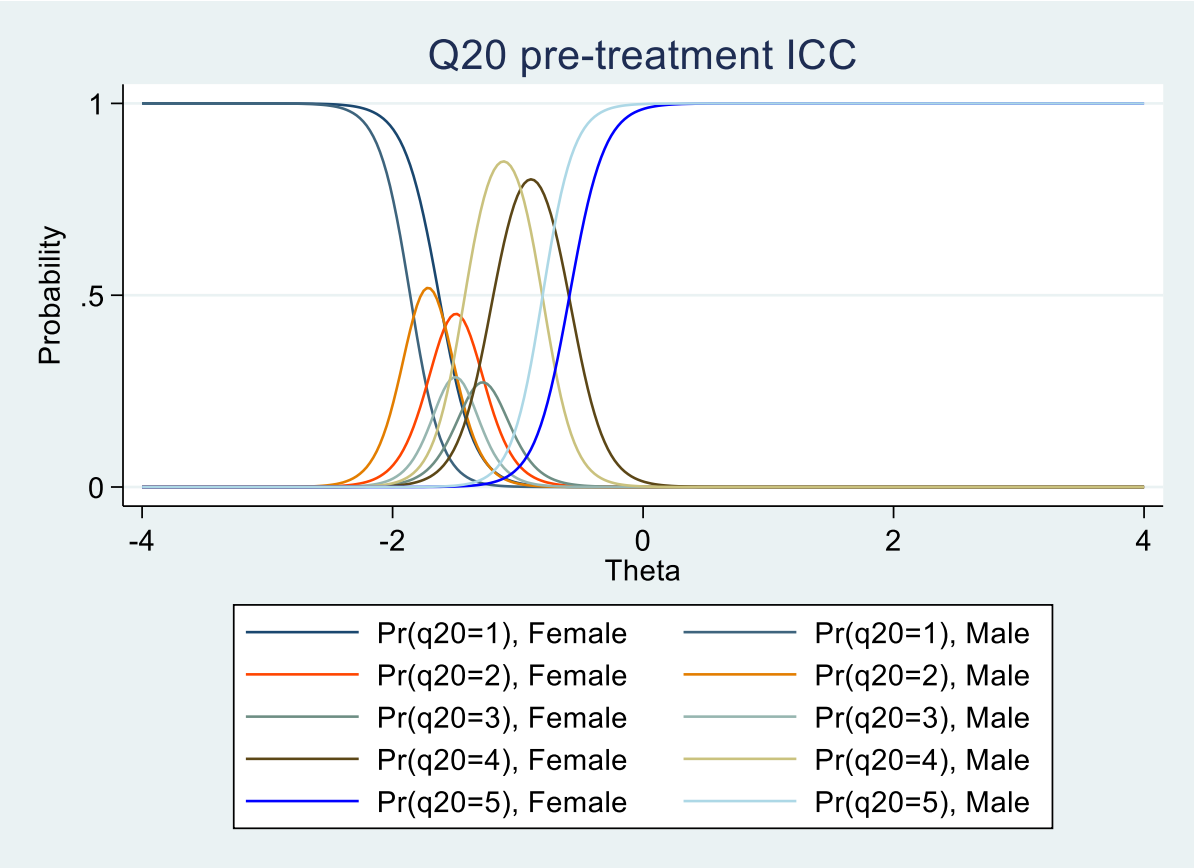


Figure 25 shows the difference in group responses between males and females for item 20 pre-treatment.

The results of the likelihood ratio test for item 20 demonstrating DIF = 12.18, with p=0.0324. Therefore, the null hypothesis of no DIF for item 20 at the pre-treatment datapoint can be rejected. The conclusion can therefore be made that item 20 of the MDADI at the pre-treatment datapoint exhibits DIF for the variable of sex.

6 months post-treatment datapoint

Results will now be presented for DIF analyses of the same selection of items, this time at the 6-month post-treatment datapoint.

SIMD

Item 3

Figure 26: ICC for item 3 at 6 months post-treatment

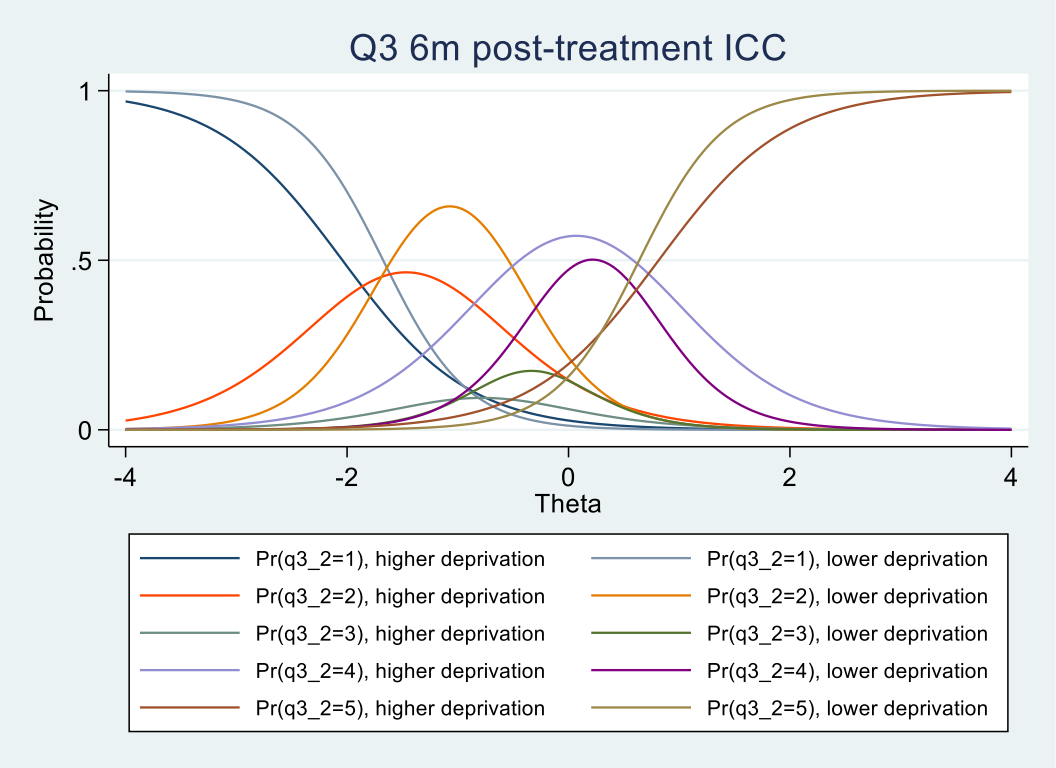
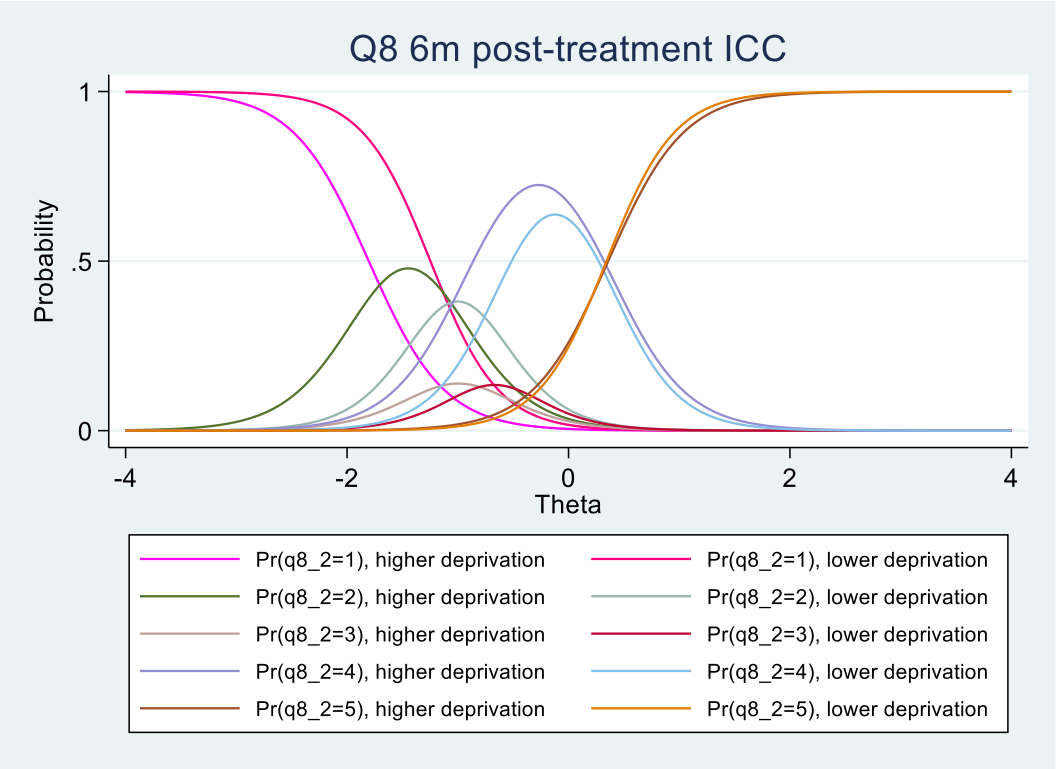


Figure 26 demonstrates that there is some difference between socioeconomic groups for responses to item 3 at the 6 months post-treatment datapoint.

The results of the likelihood ratio test for item 3 demonstrating DIF = 5.52 with p=0.3552. Therefore, the null hypothesis of no DIF for item 3 at the 6 months post-treatment datapoint must be accepted. The conclusion can therefore be made that item 3 of the MDADI at the 6 months post-treatment datapoint does not exhibit DIF for the variable of socioeconomic status.

Item 8

Figure 27: ICC for item 8 at 6 months post-treatment



The preceding figure depicts the ICC curves for responses to item 8 at the 6 months post-treatment datapoint, by socioeconomic group.

The results of the likelihood ratio test for item 8 demonstrating DIF = 3.19 with p=0.6706. Therefore, the null hypothesis of no DIF for item 8 at the 6 months post-treatment datapoint must be accepted. The conclusion can therefore be made that item 8 of the MDADI at the 6 months post-treatment datapoint does not exhibit DIF for the variable of socioeconomic status.

Age

Item 9

Figure 28: ICC for item 9 at 6 months post-treatment

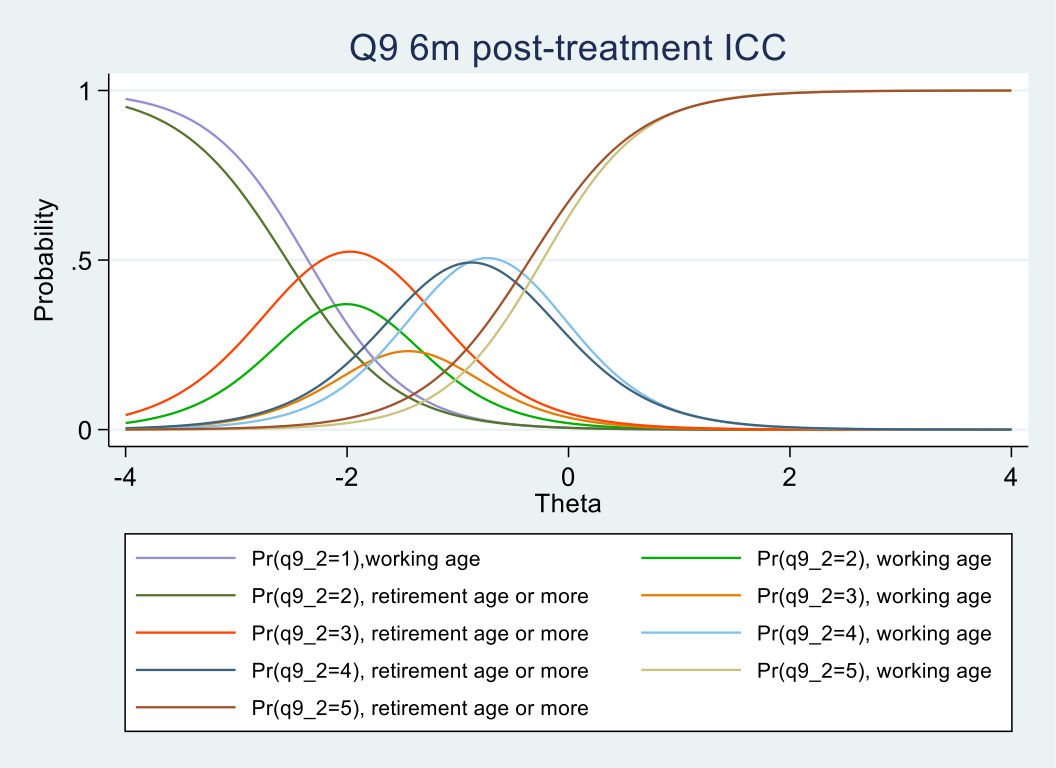


Figure 28 demonstrates the ICC curves for responses to item 9 at the 6-months post-treatment datapoint.

The results of the likelihood ratio test for item 9 demonstrating DIF for the variable of age = 19.16 with p=0.0007. Therefore, the null hypothesis of no DIF for item 9 at the 6 months post-treatment datapoint can be rejected. The conclusion can therefore be made that item 9 of the MDADI at the 6 months post-treatment datapoint exhibits DIF for the variable of age.

Sex

Item 2

Figure 29: ICC for item 2 at 6 months post-treatment

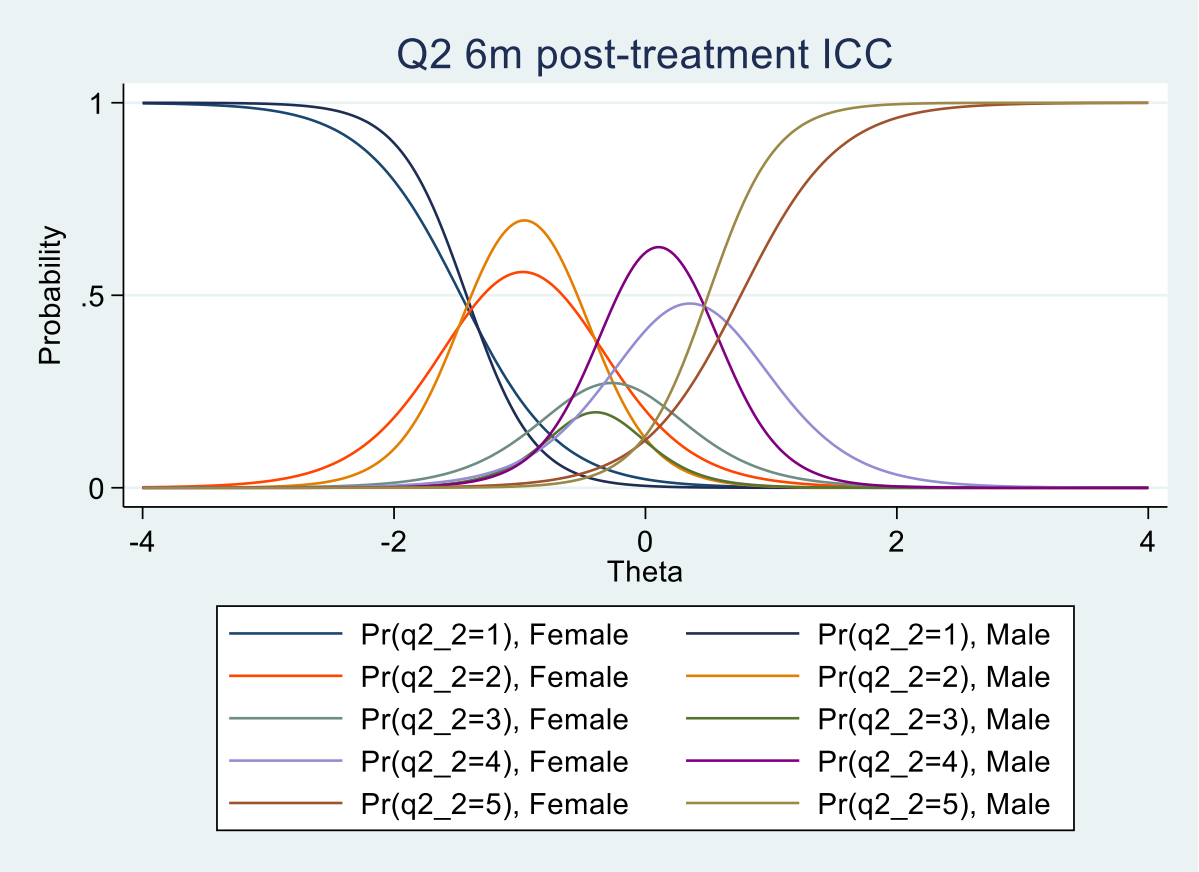


Figure 29 displays the ICC curves for responses to item 2 at the 6 months post-treatment datapoint.

The results of the likelihood ratio test for item 2 demonstrating DIF = 3.73 with p=0.5893. Therefore, the null hypothesis of no DIF for item 2 at the 6 months post-treatment datapoint must be accepted. The conclusion can therefore be made that item 2 of the MDADI at the 6 months post-treatment datapoint does not exhibit DIF for the variable of sex.

Item 18

Figure 30: ICC for item 18 at 6 months post-treatment

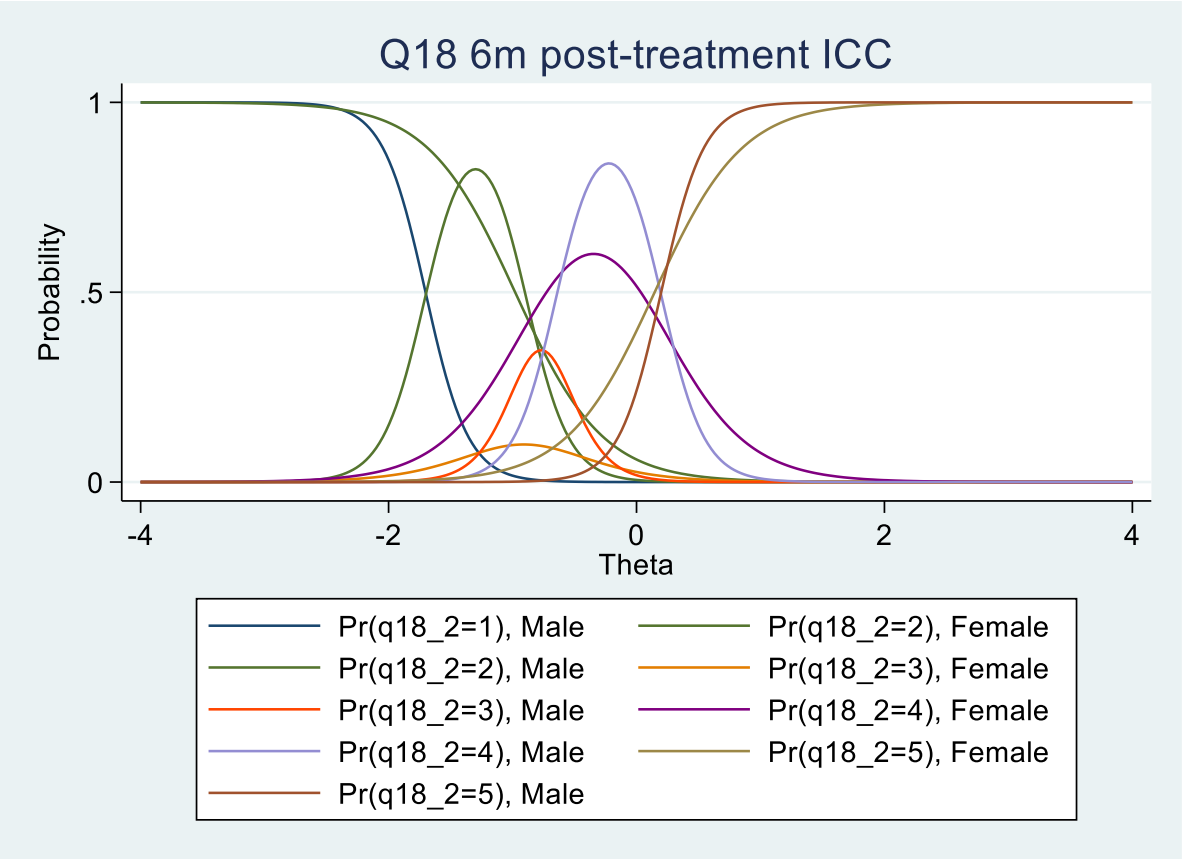


Figure 30 illustrates the different ICC curves for responses to item 18 at the 6 months post-treatment datapoint, grouped by sex.

The results of the likelihood ratio test for item 18 demonstrating DIF = 47.55 with p=0.0000. Therefore, the null hypothesis of no DIF for item 18 at the 6 months post-treatment datapoint can be rejected. The conclusion can therefore be made that item 18 of the MDADI at the 6 months post-treatment datapoint exhibits DIF for the variable of sex.

Item 20

Figure 31: ICC for item 20 at 6 months post-treatment

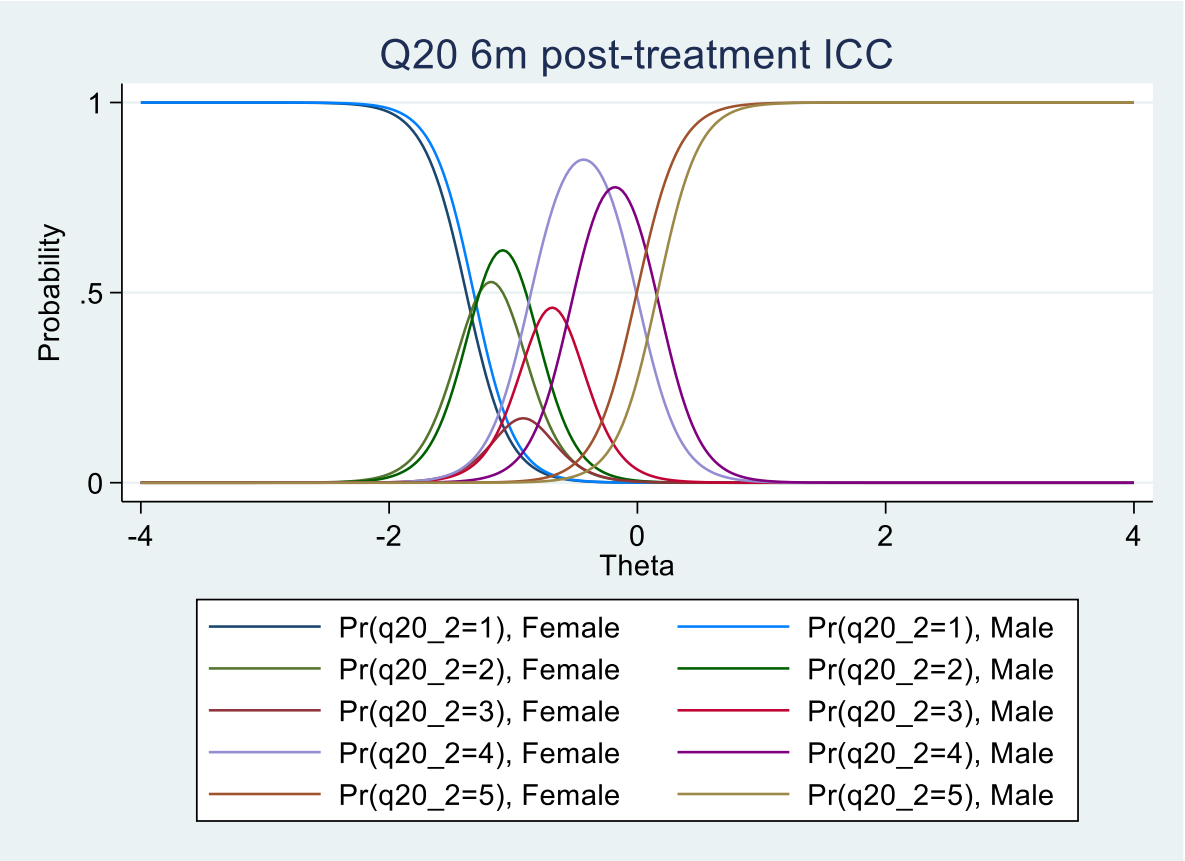


Figure 31 shows the ICC curves for responses to item 20 at the 6 months post-treatment datapoint, grouped by sex.

The results of the likelihood ratio test for item 20 demonstrating DIF = 9.88 with p=0.0787. Therefore, the null hypothesis of no DIF for item 20 at the 6 months post-treatment datapoint must be accepted. The conclusion can therefore be made that item 20 of the MDADI at the 6 months post-treatment datapoint does not exhibit DIF for the variable of sex.

5.5.3 DIF analysis summary

Table 49 below summarises the results of the DIF analyses conducted for items 2, 3, 8, 9, 18 and 20 of the MDADI at both the pre-treatment and 6 months post-treatment datapoints.

Datapoint	Variable	Item	Exhibits DIF?	P value
Pretreatment	Socioeconomic status	3	Yes	0.0000
		8	Yes	0.0014
	Age	9	Yes	0.014
	Sex	2	Yes	0.0098
		18	Yes	0.0000
		20	Yes	0.0324
6 months post-treatment	Socioeconomic status	3	No	0.3552
		8	No	0.6706
	Age	9	Yes	0.0007
	Sex	2	No	0.5893
		18	Yes	0.0000
		20	No	0.0787

Table 49: Summary of DIF analysis results

At the pre-treatment datapoint, all items identified by questionnaire respondents as potentially subject to subgroup differences were proven to be so under DIF analysis. This constitutes more than a quarter of total MDADI items being subject to DIF at this timepoint. At the 6 months post-treatment datapoint, two of the six items identified as being potentially subject to DIF were shown to be so (10% of total MDADI items).

When looking at the presence of DIF within MDADI subscales, the following is noted by datapoint.

Pre-treatment datapoint:

At this timepoint, the emotional subscale contains 50% (3/6) of items with DIF, the physical subscale does not contain items with DIF, and the functional subscale contains 60% (3/5) items with DIF. This is illustrated in the following Tables 50, 51 and 52; items with DIF are highlighted in grey.

Information Rank	Item	Content	Information value	DIF variable
1	18	I have low self-esteem because of my swallowing problem	31.38	Sex
2	8	I do not go out because of my swallowing problem	19.18	Socioeconomic status
3	12	Other people are irritated by my eating problem	15.94	-
4	6	I am upset by my swallowing problem	10.22	-
5	2	I am embarrassed by my eating habits	8.59	Sex
6	5	I do not feel self-conscious when I eat	4.02	-

Table 50: Pre-treatment emotional subscale indicating DIF

Information Rank	Item	Content	Information value
1	17	I cannot maintain my weight because of my swallowing problem	11.16
2	11	People ask me 'why can't you eat that?'	10.79
3	13	I cough when I try to drink liquids	10.75
4	7	Swallowing takes great effort	10.09
5	16	I limit my food intake because of my swallowing difficulty	9.25
6	4	Swallowing is more difficult at the end of the day	8.50
7	10	It takes me longer to eat because of my swallowing problem	7.50
8	19	I feel that I am swallowing a huge amount of food	7.23

Table 51: Pre-treatment physical subscale indicating DIF

Information Rank	Item	Content	Information value	DIF variable
1	14	My swallowing problems limit my personal and social life	20.65	-
2	20	I feel excluded because of my eating habits	20.20	Sex
3	9	My swallowing difficulty has caused me to lose income	9.20	Age
4	3	People have difficulty cooking for me	7.27	Socioeconomic status
5	15	I feel free to go out to eat with my friends, neighbours, relatives	5.97	-

Table 52: Pre-treatment functional subscale indicating DIF

6 months post-treatment datapoint

At the 6 months post-treatment datapoint, the emotional subscale contains one item with DIF constituting 16.67% of the subscale (item 18, DIF for sex). No items in the physical subscale contain DIF, and the functional subscale contains one item with DIF constituting 20% of the subscale (item 9, DIF for age). This is illustrated in the following Tables 53, 54 and 55; items with DIF are highlighted in grey.

Information Rank	Item	Content	Information value	DIF variable
1	18	I have low self-esteem because of my swallowing problem	15.62	Sex
2	2	I am embarrassed by my eating habits	10.12	-
3	6	I am upset by my swallowing problem	9.27	-
4	12	Other people are irritated by my eating problem	9.12	-
5	8	I do not go out because of my swallowing problem	8.80	-
6	5	I do not feel self-conscious when I eat	4.05	-

Table 53: 6 months post-treatment emotional subscale indicating DIF

Information Rank	Item	Content	Information value
1	7	Swallowing takes great effort	9.13
2	11	People ask me 'why can't you eat that?'	7.63
3	17	I cannot maintain my weight because of my swallowing problem	7.43
4	16	I limit my food intake because of my swallowing difficulty	6.96
5	4	Swallowing is more difficult at the end of the day	5.83
6	19	I feel that I am swallowing a huge amount of food	4.10
7	13	I cough when I try to drink liquids	4.03
8	10	It takes me longer to eat because of my swallowing problem	3.72

Table 54: 6 months post-treatment physical subscale indicating DIF

Information Rank	Item	Content	Information value	DIF variable
1	20	I feel excluded because of my eating habits	19.57	-
2	14	My swallowing problems limit my personal and social life	15.87	-
3	3	People have difficulty cooking for me	6.85	-
4	9	My swallowing difficulty has caused me to lose income	5.79	Age
5	15	I feel free to go out to eat with my friends, neighbours, relatives	5.67	-

Table 55: 6 months post-treatment functional subscale indicating DIF

This analysis shows that DIF results vary across datapoints, adding further weight to the concept that the MDADI behaves differently depending on when in a patient's cancer journey it is used, as evidenced in the results of Research Question 2 in this thesis. How the presence of DIF identified can contribute to the process of item reduction will be demonstrated in the next chapter of this thesis.

CHAPTER 6: SYNTHESIS OF QUALITATIVE AND QUANTITATIVE RESULTS

Synthesis of qualitative and quantitative data generated by Research Questions 1-3 provides suggestions for formation of shortened versions of the MDADI that have improved validity and clinical utility, for the different timepoints of pre-treatment and 6 months post-treatment.

The quantitative results of item information-lead analysis will now be integrated with quantitative results from the DIF analysis, along with qualitative results from clinicians concerning the clinical utility of specific items, to inform a process of item reduction.

6.1 Synthesising item information and DIF data

Table 56 summarises the originally proposed theoretical versions of the short-form MDADIs for both timepoints, based on item information and subscale ratios as previously presented. Question 1 of the MDADI is not included in the calculations as this is a separate, global item.

Scale length	Pre-treatment	6m post-treatment
5 item theoretical	8, 11, 14, 17, 18	2, 7, 11, 18, 20

Table 56: Theoretical MDADI shortforms

To facilitate consideration of how DIF might affect the suggested items for inclusion in MDADI short forms, Table 57 shows items ranked in order by information value but also highlighted for presence of DIF where applicable:

	Pre-treatment
Highest – lowest information item rank order	18, 14, 20, 8, 12, 1, 17, 11, 13, 6, 7, 16, 9, 2, 4, 10, 3, 19, 15, 5
	6 months post-treatment
Highest – lowest information item rank order	20, 14, 18, 1, 2, 6, 7, 12, 8, 11, 17, 16, 3, 4, 9, 15, 19, 5, 13, 10

Table 57: Ranking of MDADI items by total information at both datapoints; items highlighted in grey have DIF

When viewed in this manner, it becomes evident that 3 out of the top 5 items as ranked by information display DIF at the pre-treatment datapoint, and 1 out of 5 for the 6 months post-treatment datapoint.

The DIF results presented above add an extra dimension when considering how to form a shortened version of the MDADI. In the literature, items shown to be subject to DIF are often considered for removal from tools (Jones, 2019). In the case of this study, as length of the MDADI has been highlighted in the qualitative data as an issue, it makes sense to use presence of DIF in particular items as an additional way of providing candidate items for removal from a shortened version of the tool. If the items highlighted in grey in Table 57 above are removed from consideration, further decisions need to be made about potential item substitutions in their place.

For the purposes of this study, given the clinical focus and the fact that the MDADI was originally designed to consider and represent different facets of dysphagia related QoL, in the following synthesis I have attempted to continue as much of a 'theoretical' approach as possible: that is, trying to maintain some sort of proportional subscale representation when suggesting shortened versions of the MDADI.

6.2 Additional qualitative data for integration

Qualitative data analysis indicated that the MDADI tool is overly long; this has been addressed mathematically in this thesis, through the process of considering item reduction.

In conjunction with mathematical analyses of tool length reduction by considering item information, and then item DIF, multiple items across all subscales were also flagged by questionnaire respondents as having potential content overlap with others, as demonstrated in Table 58:

Item number	Potential overlap- item number(s)
6	2, 5
8	14
10	16
11	2
14	1
15	8
18	2,5
19	7

Table 58: Items suggested for elision or removal; Key: Emotional = pink, Physical = green, Functional = blue

Clinicians made the argument that these overlapping items could be considered for removal from the tool; therefore, these suggestions will be incorporated into the following data integration.

6.3 Integrated MDADI short form suggestions

Pre-treatment datapoint

The original theoretical 5-item pre-treatment scale was suggested as containing items 8, 11, 14, 17 and 18. However, items 8 and 18 have been shown to be subject to DIF. As items 8 and 18 are both from the emotional subscale, the remaining items from that subscale with the highest information values (12 and 6) could be substituted in their place to maintain the subscale ratio, giving a 5-item scale of items 6, 11, 12, 14 and 17. When referring to Table 58 above, none of these items have conflicts in terms of clinician-suggested content overlap.

Table 59 therefore presents this study's final suggestion for a 5-item shortform MDADI for pre-treatment use:

5 item pre-treatment MDADI shortform	
Items	**Content**
6	I am upset by my swallowing problem
11	People ask me 'why can't you eat that?'
12	Other people are irritated by my eating problem
14	My swallowing problems limit my personal and social life
17	I cannot maintain my weight because of my swallowing problem

Table 59: Integrated 5 item shortform MDADI for pre-treatment use; Key: Emotional = pink, Physical = green, Functional = blue

6 months post-treatment datapoint

The original theoretical 5-item shortform MDADI for the 6 months post-treatment timepoint comprised items 2, 7, 11, 18 and 20. However, item 18 at this timepoint has been shown to be subject to DIF; in addition, clinicians flagged items 2 and 6 and 2 and 11 as being potentially too similar in content; therefore, as this is the case, I suggest that item 2 should be removed.

As both items 2 and 18 are part of the emotional subscale, the items with the next-highest information scores from this subscale could be substituted in their stead to maintain the subscale ratio: this would be items 6 and 12. This then would give a 6-month post-treatment 5-item MDADI shortform as comprising items 6, 7, 11, 12 and 20:

5 item 6 months post-treatment MDADI shortform	
Item	**Content**
6	I am upset by my swallowing problem
7	Swallowing takes great effort
11	People ask me 'why can't you eat that?'
12	Other people are irritated by my eating problem
20	I feel excluded because of my eating habits

Table 60: Integrated 5 item shortform MDADI for 6 months post-treatment use; Key: Emotional = pink, Physical = green, Functional = blue

6.4 Additional clinical utility considerations

Some issues with the clinical utility of the MDADI tool have been addressed in the above synthesis, as they lend themselves to feeding into quantitative analysis.

However, in the qualitative data Theme seven: "Not quite where we need it to be", another item-specific issue was also raised: that of item wording. Clinician respondents suggested changes to nine items that were considered to have overly negative wording: items 6, 8, 10, 11, 12, 14, 16, 17 and 18.

This impacts on the pre-treatment 5 item MDADI: *all* included items (6, 11, 12, 14 and 17) were highlighted as having problematic wording. The inconsistency in use of 'swallowing' versus 'eating' is particularly evident when looking at the item content in this version. The 6-month post-treatment 5 item version comprising items 6, 7, 11, 12 and 20 also contains three of these problematic items (60%).

Rewording items would be a significant undertaking in terms of content re-development and then psychometric testing of the amended items; therefore, the above data have not been incorporated into the suggested content for the MDADI shortforms. However, it should be noted that in addition to the issue of item wording, tool layout and scoring as flagged by questionnaire respondents would also require to be addressed with any alternative version of the MDADI produced. These additional issues with MDADI content cannot be addressed mathematically and this will be considered in the discussion section of this thesis.

6.5 Summary of qualitative and quantitative analyses results

Analysis of the MDADI via qualitative and quantitative methods has produced a wealth of data about the strengths and weaknesses of the tool.

Qualitative data generated via an online questionnaire of UK-based SLTs working with people with HNC provided rich information regarding the content validity and clinical utility of the MDADI. The MDADI is considered to be a useful adjunct to clinical practice with this patient caseload, however significant issues with its content validity and clinical utility were highlighted, all of which impact negatively on its use with patients.

The a priori clinical utility criteria developed from a review of the literature were found to have validity in that the categories chosen were also flagged as having clinical significance in the inductive qualitative analysis. Two additional clinical utility parameters were also identified from the qualitative data.

Quantitative analysis provided novel IRT data on the psychometric aspects of structural validity and internal consistency of the MDADI. The MDADI was shown to have structural validity and good internal consistency, although internal consistency was shown to vary across levels of θ, and datapoints, with a potential ceiling effect at the pre-treatment datapoint.

Age, sex and socioeconomic status identified as potential-DIF causing variables in the qualitative data were confirmed as such on IRT analysis. Six MDADI items at pre-treatment, and two items at 6 months post-treatment, were found to be subject to DIF.

Analysis of qualitative and quantitative data enabled synthesis of shortened versions of the MDADI that would meet clinicians' suggestions of improved clinical utility and provide an opportunity to exclude items with poorer information values (higher SE), DIF, or reduced content validity. Due to the difference in results between datapoints, pre-treatment and post-treatment MDADI shortform versions were suggested.

Fusion of the qualitative and quantitative data produces a complex picture of the MDADI; the results will be explored in the context of the literature in the next chapter.

CHAPTER 7: DISCUSSION AND CRITIQUE OF STUDY

What follows is a discussion of how the results presented in Chapters Four, Five and Six answer the research questions posed in this thesis, and how these results relate to the wider context of the published literature. This contextual siting of the study results will be followed by a critique of the study's quality.

7.1 Revisiting the study aims and research questions

This study has fulfilled the aim of analysing and evaluating the MD Anderson Dysphagia Inventory on a UK-based population. The key findings with respect to the research questions are summarised below.

7.1.1 Summary of key findings

What is the content validity and clinical utility of the MD Anderson Dysphagia Inventory from a SLT perspective?

The results of this qualitative analysis of data from UK clinicians indicate that, whilst the MDADI is routinely used in clinical HNC SLT practice in the UK and has potential to be a helpful adjunct to clinical assessment, there are significant issues with its content validity and clinical utility.

What are the results of an exploration of the psychometric properties of the MDADI on a UK population using Item Response Theory?

Analysis confirmed the structural validity of the MDADI in terms of adequate fit with the Graded Response Model, as per COSMIN guidance.

IRT analysis shows that the MDADI's internal consistency varies within the tool by item and across datapoints in terms of both Standard Error (SE) and levels of θ.

What are the potential factors which might result in Differential Item Functioning?

Qualitative data indicated variables sex, age and socioeconomic status were worthy of investigation for the presence of DIF. IRT analysis showed that at the pre-treatment datapoint, MDADI items 3 and 8 are subject to DIF by the variable of socioeconomic status, item 9 is subject to DIF by the variable of working age, and items 2 and 18 are subject to DIF by the variable of sex (25% of total items). At the 6 months post-treatment datapoint, item 9 is subject to DIF by the variable of working age, and item 18 is subject to DIF by the variable of sex (10% of total items).

Synthesis of a shortform MDADI integrating qualitative and quantitative results

Whilst this was not an a priori objective of this study, during the process of data collection and analysis it became clear that making suggestions for item reduction, resulting in a recommendation for a shortened version of the MDADI, would be a relevant and important additional output. Qualitative and quantitative results were synthesised to produce a recommendation for 5-item MDADI shortforms for use at the pre-treatment and 6 months post-treatment stages. Additional issues with item wordings were flagged which impacted on the content validity and clinical utility of items even within these shortforms.

Combined qualitative and quantitative analysis of the MDADI is a novel approach that has not previously been presented in the literature, either regarding the MDADI or any other dysphagia PROM. The in-depth analysis of the MDADI taken by the current study has generated complex results which have significant implications for both SLT clinical practice, and wider HNC research practice.

Discussion of these results in the context of the extant literature will now follow.

7.2 Research Question 1

What is the content validity and clinical utility of the MD Anderson Dysphagia Inventory from a UK SLT perspective?

To date, there is no published literature exploring the content validity or clinical utility of the MDADI, and this study is the first to have investigated these areas. The results of this study corroborated the need for a dedicated outcomes tool considering swallowing related QoL in patients with HNC as an adjunct to clinical SLT practice, but also highlighted the shortcomings of the MDADI with respect to both its content validity and clinical utility.

The qualitative analysis performed in this study generated data that speak to both the content validity and clinical utility of the MDADI. Concepts of content validity and clinical utility share some commonalities. Where appropriate, related clinical utility parameters that relate specifically to, or overlap with, the MDADI's content validity, will be discussed first in tandem with the specific qualitative Themes 1, 4, 6 and 7. As Themes 2 and 5 relate more specifically to clinical utility, they will be discussed in the subsequent section.

An aside about research vs clinical use of PROM tools

Investigating the MDADI from the perspective of its use in practice by clinicians is valid, as despite the MDADI authors' assertion that it is designed to be a research tool (K. Hutcheson, personal communication, November 24, 2020), this study shows that SLTs practising in the

field of HNC are using this tool in their everyday practice across multiple timepoints during their patients' episodes of care.

This finding that UK SLTs use the MDADI in their everyday practice agrees with the results of a previous survey of HNC SLT practice undertaken by Roe et al. (2012). Osborn et al. (2019) also showed in their survey of head and neck surgeons in the USA that the MDADI was one of the most used assessment tools in their clinical practice. There is, after all, no other extant tool to assess dysphagia related QoL specifically for patients with HNC.

Many PROMs were originally developed for research trials, designed to produce data on group-level comparisons rather than informing individual-level care (Edwards et al., 2016). Whilst Kroenke et al. (2015) suggest that "most PROs gravitate from research into practice" (p.1086), Vickers and Chen (2017) voice concerns about the applicability of PROMs developed for research use to clinical practice, as they may be challenging to use in a clinical context due to potential issues with clinical utility. I would argue however that characteristics of a tool, whether it be validity, reliability or clinical utility, have great significance no matter the proposed setting of use. No tool is administered in a sterile vacuum out with the relevant concerns of its content validity and clinical utility: whether a patient is completing a tool for a clinical trial or their individualised care, its relevance and usability are paramount.

7.2.1 The content validity of the MDADI

The content validity of a tool refers to its relevance to the domain and population it focuses on (Salkind, 2010), and is widely considered to be the most important measurement property of a PROM (Terwee et al., 2018). Poor content validity has potential to impact all other aspects of tool psychometrics. To provide a framework for discussion of the content validity results of this study, I will use the COSMIN content validity assessment criteria (Prinsen et al., 2018, Terwee et al., 2018).

The COSMIN tool considers relevance, comprehensiveness and comprehensibility as inherent to content validity and suggest when rating these properties of a tool that it can be either sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?). It is necessary to keep in mind however that ultimately rating of content validity remains a subjective judgement (Prinsen et al., 2018).

Table 61 below shows a summary of my assessment of the MDADI's content validity as per the rating system presented by Terwee et al. (2018), based on the results of this thesis.

Content validity criteria	Overall rating
Relevance	
Are the included items relevant for the construct of interest?	±
Are the included items relevant for the target population of interest?	±
Are the included items relevant for the context of use of interest?	±
Are the response options appropriate?	-
Is the recall period appropriate?	+
RELEVANCE RATING	***INCONSISTENT***
Comprehensiveness	
Are all key concepts included?	-
COMPREHENSIVENESS RATING	***INSUFFICIENT***
Comprehensibility	
Are the PROM instructions understood by the population of interest as intended?	-
Are the PROM items and response options understood by the population of interest as intended?	-
Are the PROM items appropriately worded?	-
Do the response options match the question?	-
COMPREHENSIBILITY RATING	***INSUFFICIENT***

Table 61: COSMIN content validity rating results for the MDADI

Despite the MDADI's high profile within HNC research and practice, its content validity has been minimally appraised to date. Previous studies assessing the content validity of the MDADI (Ojo et al., 2012, Timmerman et al., 2014, Patel et al., 2017) take a cursory 'checklist' approach to assessing the presence of content validity consideration in development of the tool, rather than performing a de novo evaluation of the content validity of the tool, and even then, all three papers came to different conclusions as discussed in Chapter Two of this thesis.

This thesis provides the first qualitative data on the content validity of the MDADI, albeit from a clinician rather than a patient perspective, and highlights issues with all three aspects that Terwee et al. (2018) flag as inherent to content validity: Relevance, Comprehensiveness and Comprehensibility.

As no previously published literature exists specifically regarding de novo analysis of the content validity of the MDADI, I will consider the results of the current study concerning MDADI content validity in the context of the wider extant literature with reference to these three key parameters.

Relevance

The COSMIN assessment criteria for relevance are as follows:

- Are the included items relevant for the construct of interest?
- Are the included items relevant for the target population of interest?
- Are the included items relevant for the context of use of interest?
- Are the response options appropriate?
- Is the recall period appropriate?

I would suggest 'response options' and 'recall' included in the above Relevance criteria, fall within a separate domain of Clinical Utility (as defined in this thesis) rather than Content Validity, as they relate specifically to practical aspects of a tool. Therefore, I will consider these criteria as part of the Clinical Utility discussion presented later in this chapter.

Using the COSMIN assessment criteria for content validity, I have rated these 'Relevance' criteria as 'inconsistent'. Clinician respondents in this study suggest that the MDADI functions as a 'starting point', as it opens up the concept of dysphagia related quality of life for open-ended discussion between clinician and patient, rather than all the MDADI items themselves having strong content validity.

Qualitative Theme One generated in this study ("a conversation starter") corroborates existing evidence that PROMs are a relevant adjunct to clinical practice: the concept of the MDADI had clinical relevance to questionnaire respondents. In her overview of the literature regarding application of PROMs in clinical practice, Greenhalgh (2009) highlights the frequently cited potential of PROMs for opening up and guiding clinical discussions around HRQoL related issues. Benavent et al. (2021), in their study surveying physicians about the clinical utility of PROM used in gastrointestinal oncology, corroborate this finding with their study providing evidence that clinicians feel that use of HRQoL PROMs improves communication between clinicians and patients.

The HNC dysphagia literature highlights the existence of significant dysphagia-related quality of life impact post-treatment (Patterson et al., 2015, Nund et al., 2014a) and therefore it is essential that assessment tools exist that are able to assess this and provide meaningful data. This study indicates that a tool that facilitates clinicians instigating emotive and difficult conversations with patients about key aspects of cancer survivorship is clinically useful; however, the items need to be relevant, and the tool also needs to be practically useable. A major concern with the MDADI is that it potentially lacks relevance in terms of its item content with respect to specific, significant subgroups of people with HNC.

This thesis has shown that the tool is inherently exclusory to specific patient subgroups (qualitative Theme Six: “excluded groups”). A tool that aims to assess the impact of dysphagia on quality of life should be relevant to, and able to comprehensively represent all subgroups of patients with dysphagia, arguably most of all those patients with the most severe dysphagia. However, data from this study highlighted that many clinicians feel the MDADI is exclusory to patients with ‘nil by mouth’ (NBM) status who are dependent on tube-feeding due to the severity of their dysphagia. Studies of post-HNC treatment swallow outcomes show high rates of tube dependency (suggesting likely minimal or no oral intake) (Russi et al., 2012, Hutcheson et al., 2012) therefore it would seem logical that any tool that assesses HRQoL with respect to dysphagia for people with HNC would need to adequately cover patients at the most severe end of the spectrum in terms of swallowing impairment. The items in the MDADI, focussed as many of them are on the act of eating or drinking, automatically exclude people who are NBM from responding and therefore completing the tool. Brotherton and Judd (2007) performed a literature review which demonstrated that enteral feeding has a significant impact on patients’ quality of life. Whilst the authors found factors impacting QoL to be complex in nature and wider than dysphagia or eating and drinking issues and included wider aetiologies than just HNC, dysphagia was the indication for enteral feeding in most studies reviewed and therefore could be seen to be potentially linked.

In their pilot study investigating the relationship between nutritional status and psychological distress in patients with HNC, Gosak et al. (2020) presented preliminary findings that there is a relationship between levels of anxiety and depression and reduced nutritional status, with an association with dependence on tube-feeding. Although this study does not specifically identify whether dysphagia was present or not, in the HNC population there is commonly a direct link between tube-feeding and dysphagia.

Patients with the most severe dysphagia also need assessment of their dysphagia-related QoL so their support and survivorship needs can be understood and addressed. However, whatever the level of dysphagia-related QoL in the subgroup of patients who are NBM, a tool capable of assessing this is a necessity for clinical practice.

There is a bioethical concern with PROMs that fail to meet the needs of more vulnerable subgroups, and this issue has been explored in other areas of healthcare, such as by Hagell et al. (2009) in their exploration of PROMs use with people with Parkinson’s Disease. They found two commonly used PROMs to be perceived more poorly by older people or people with more severe symptoms.

Studies looking at treatment outcomes for new HNC treatment regimens (D'Andréa et al., 2022, Charters et al., 2021, Petkar et al., 2016, Mehanna et al., 2017) continue to use the MDADI as an outcome measure. However, the study protocols do not provide data on the plan for using the MDADI with participants whose dysphagia is so severe that, as suggested may be the case by this study's analysis, it would preclude clinicians from using the MDADI tool for example with patients who were NBM. It is not clear in these studies whether such patients were excluded from the analysis, potentially skewing the results through non-response bias, or included in the analysis, potentially also skewing the results given the suggestion in the data of this current study that the MDADI has reduced validity for the subgroup of patients who are NBM.

Comprehensiveness

The COSMIN criterion for comprehensiveness is "are all key concepts included" (Terwee et al., 2018). In Table 61 above, based on the results of this thesis, I have rated this as 'insufficient' as follows.

Qualitative Theme Four in this study ("The bigger picture of eating and drinking") highlighted issues with the comprehensiveness of the MDADI. Specifically, the concern that the MDADI is not sufficiently clear about what it is assessing, due to a lack of definition of concept in terms of dysphagia versus the wider process of eating and drinking. To robustly assess a concept, that concept first requires to be clearly defined. This thesis has shown that clinicians find the wording of the MDADI inconsistent: the terms 'swallowing' and 'eating' are used interchangeably without obvious rationale. It could be argued that the MDADI's scope is *too* comprehensive and needs to be reined in to focus specifically on dysphagia, or conversely not comprehensive enough, in that it does not give explicit reference to other swallowing-adjacent issues such as xerostomia or reduced dentition, which may be inseparable from patients' eating, drinking and swallowing experience. There remains an issue of scope, clarity and definition when it comes to eating, drinking, swallowing and dysphagia that is highlighted in this study but has yet to be addressed adequately in research, literature or clinical practice.

Dysphagia as an impairment of oropharyngeal swallow function is the narrowest scope of concept. In her classic and still often-cited text on swallowing disorders, Logemann (1998) defines dysphagia as "difficulty moving food from mouth to stomach" (p.1); however in recent years a wealth of qualitative research on the lived experience of people with dysphagia, and long-term follow-up outcomes research with patients who have had treatment for HNC, has increased understanding of the far-reaching implications of dysphagia and the interrelated constellation of symptoms that impact on the process and experience of eating, drinking and

swallowing. The complex wider concept of eating and drinking for people post-treatment for HNC encompasses additional factors such as dentition, saliva production, taste, appetite, pain and the potential need for nutritional support. This makes assessment of dysphagia and dysphagia related QoL in HNC challenging, as clinicians and researchers try to capture the diverse and nuanced aspects of this complex issue.

In the literature, swallow physiology adjacent issues such as dysgeusia (altered taste) and xerostomia (reduced saliva) are frequently grouped under the term ‘dysphagia’, for example by Nund et al. (2014a) in their qualitative study investigating the lived experience of post-HNC dysphagia. Similarly Bressan et al. (2017) in their systematic review and meta-analysis of the impact of HNC treatment on QoL found that dysphagia, xerostomia, dysgeusia and oral mucositis all impacted significantly on patients’ quality of life and that these symptoms do not occur in isolation.

In terms of how to better define these issues in their interrelatedness, in their review of the literature Ganzer et al. (2015) use the term ‘eating experience’ to encompass the complex, multifaceted physical, social and emotional impacts on eating drinking and swallowing post HNC (they separate out ‘dysphagia’ as one element of the eating experience).

Ottosson et al. (2013) refer to ‘food, eating and meals’ as a meaningful term to describe a variety of issues in their qualitative study of patients’ experiences post HNC treatment, whereas another qualitative, co-produced study carried out with people with HNC by Burges Watson et al. (2018) suggests ‘altered eating’ as a more appropriate umbrella term. These results suggest that perhaps it is more realistic to aim to assess a range of eating, drinking and swallowing related issues rather than singling out swallowing (or dysphagia) when assessing related QoL in the HNC population.

Chan et al. (2019) recognise the limitations of excluding the wider picture of HNC side-effects that impact on eating, drinking and swallowing and therefore have developed a novel assessment tool, the Head and Neck Cancer Survivors’ Assessment of Mealtimes (HNSAM). HNSAM items were developed from qualitative data gathered from interviews with HNC survivors and speak to the common themes experienced by these patients which all had an interrelated impact on their mealtimes. The HNSAM focusses on mealtime-related eating and drinking changes, rather than an overarching assessment of dysphagia-related QoL, incorporating dysphagia-adjacent symptoms such as reduced saliva, appetite, sensory issues and the impact of dentition.

The inconsistency in terminology in the MDADI needs to be addressed. This thesis has highlighted that the mixing of the terms ‘swallowing’ and ‘eating’ across items in the tool confuses users, disappoints clinicians and negatively impacts on its clinical validity.

Even when considering a narrower, physiologically focussed concept of dysphagia, this thesis has shown that there are also problems with the specifically 'swallowing related' items in the MDADI. Clinicians suggested the items suspected to be concerning pharyngeal efficiency (7: 'Swallowing takes great effort'/19: 'I feel that I am swallowing a huge amount of food') were insufficiently clear, and the item concerning swallow safety (13: I cough when I try to drink liquids) did not take into account the high rates of silent aspiration in this clinical group (Langerman et al., 2007) and therefore lacked validity. In their study comparing MDADI results with swallow efficiency and safety physiological parameters, Kirsh et al. (2019) demonstrated a lack of correlation between MDADI scores and degree of physiological impairment. The authors suggests that patients' awareness of swallow dysfunction is not equivalent to the level of physiologic decline as observed by formal instrumental swallow assessment. This therefore has implications for the above-mentioned items of the MDADI and concurs with questions raised in this thesis about the validity of the 'physical' MDADI subscale (which items 7, 13 and 19 all belong to). This was mirrored in the quantitative results of this thesis, which showed that the MDADI 'physical' subscale items contained lower information than other subscale items. In their recent systematic review exploring the relationship between affective symptoms and oropharyngeal dysphagia in HNC patients, Krebbers et al. (2022) emphasise that, whilst most of the published literature shows a correlation between dysphagia and affective symptoms, the correlation relationship is not always positive in nature, and many other factors may influence affective symptoms in addition to dysphagia.

The significance of these mismatches to clinical practice is for example that if a patient with minimal physiological swallow impairment reports high dysphagia related impact on QoL, this needs to be probed further by the SLT, understood, and appropriately managed, for example through counselling, support and education. It might perhaps be better for the MDADI to focus on evaluation of the functional and emotional impact of EDS difficulties rather than trying to include a 'physical' element to its questioning, as this width of scope may reduce its validity.

Comprehensibility

The COSMIN criteria for comprehensibility are as follows:

- Are the PROM instructions understood by the population of interest as intended?
- Are the PROM items and response options understood by the population of interest as intended?
- Are the PROM items appropriately worded?
- Do the response options match the question?

As per Table 61 above, I have evaluated this parameter as ‘insufficient’ based on the results presented in this thesis. The content of this parameter of content validity has strong links with some of the clinical utility criteria developed in this thesis. I would suggest that therefore the concept of comprehensibility should be considered as part of the wider topic of clinical utility of a tool. The results of the clinical utility assessment of the MDADI will be discussed in the next section.

7.2.2 Clinical utility

The concept of clinical utility and its assessment

As described in the literature review of this thesis, the concept of 'clinical utility' is currently poorly defined, even though the clinical utility of outcomes tools has great impact on their relevance and usability. The review of the literature presented in this thesis generated a more comprehensive taxonomy of clinical utility characteristics than exists in the published literature to date. To recap, clinical utility has been separated into two distinct but related domains which I have defined as follows:

> **'Relevance'**- the relevance of the tool to both patients and clinicians, including aspects such as whether there are clinical benefits to using the tool, whether all items within the tool are pertinent to the clinical area, whether the tool covers all patients in the target group irrespective of level of disability.
>
> **'Usability'** - more practical aspects of PROM use such as number of items, time taken to complete and score, readability, available translations and whether there are cost or legal implications such as copyright involved in using the tool.

Aspects of both Relevance and Usability can be seen to have direct links to the 'comprehensibility' dimension of content validity described by Terwee et al. (2018).

The synthesis of the clinical utility taxonomy developed in Chapter Two of this thesis with qualitative data generated by the current study confirmed the validity of the domains included, as well as suggesting two additional criteria of 'tool layout' and 'response modality' in the 'usability' domain as well as an amendment to include the potential for improving tool accessibility as necessary in the 'relevance' domain.

Table 62 below outlines the amended clinical utility domains synthesised from the literature review and the qualitative analysis.

RELEVANCE	USABILITY
R1 Does the tool impact on clinical decision making? **R2** Are there potentially issues with tool use with patient subgroups? **R3** Is normative data available? **R4** Is data available on standard error of measurement (SEM) and minimal important change (MIC)? **R5** Are questions likely to be distressing for patients? **R6** Is it possible to use a proxy or make the tool more accessible for patients with special needs? **R7** Is the tool acceptable to patients – would they be willing to reuse the scale in the future? **R8** Are the items relevant? **R9** Is the recall timeframe appropriate?	**U1** Does the tool have an appropriate literacy level? **U2** Are there any intellectual property issues? **U3** Is there a cost involved with using the tool? **U4** Is the time taken to administer acceptable? **U5** Is the scoring process straightforward? **U6** Is the time taken to score acceptable? **U7** Are translations into other languages available? **U8** Is there a plan for managing/interpreting missing responses? **U9** Is the layout of the tool acceptable? **U10** Is the response modality of the tool acceptable?

Table 62: Summary of amended clinical utility factors Key: red = criteria not addressed in this discussion; green = suggestions for amendment/addition

This thesis has presented a novel synthesis of the concept of clinical utility with regards to PROMs. Clinicians involved in this study made comments evidencing that although they might not call it 'clinical utility', characteristics of this concept were very much relevant to the use of a tool in practice, and in the specific case of the MDADI there were many issues with its clinical utility which had a negative impact, sometimes to the extent that teams made the decision to no longer use the tool.

The approach taken in this thesis of generating qualitative data from clinicians to substantiate the literature-generated initial domains is innovative and in contrast to the approach taken by other authors. Previous studies have either not included stakeholders in their clinical utility criteria development or have given limited detail of how they did (Kroenke et al., 2015, Nic Giolla Easpaig et al., 2020, Turner et al., 2020, Aiyegbusi et al., 2018, Montgomery et al., 2020).

Manduchi et al. (2022) describe a protocol for a systematic review published after the literature review for this thesis was carried out, that is currently underway with the aim of assessing the psychometric properties of PROMs specifically for dysphagia in HNC using the COSMIN methodology. In addition to the COSMIN criteria, they mention 'feasibility' (including completion time, and type and ease of administration) as an additionally important aspect of a tool's properties which is worthy of assessment; however, they do not propose a formalised way of evaluating this, instead proposing to provide 'descriptive detail'.

An additional relevant study that has been published since commencement of this thesis is that of Benavent et al. (2021), which describes their investigation of the physician-perceived utility of a HRQoL PROM used in the management of patients with gastrointestinal neuroendocrine tumours. They surveyed 36 clinicians with an ad hoc questionnaire they

devised to explore three aspects of clinical utility: 1. The impact of the tool on clinical and therapeutic decision-making, 2. The impact of the tool on doctor-patient communication, and 3. Questionnaire characteristics and their impact on clinical utility. However, the authors provide no detail as to how they developed the content of their survey, nor justification for the items included.

Nevertheless, there are some similarities in content between Benavent's tool and the criteria described in this thesis: Benavent etal consider impact on clinical decision making, relevance of items, the 'understandability' of the tool, and whether "the characteristics of the questionnaire enable its routine use in clinical practice" (p.40), however they do not elaborate on what those characteristics might be. In this thesis the clinical utility domains developed consider this in much greater detail, e.g., response modality, layout, time taken to administer, scoring, cost. Benavent etal's tool used a Likert rating response system for each of the 14 items in their clinical utility assessment tool, rather than eliciting 'free text' type responses, thereby limiting the information they could elicit from physicians, as well as the usefulness of responses due to lack of definition of which variables might be included in the 'characteristics of the questionnaire'.

In summary then, the clinical utility domains proposed in this thesis are more detailed than in extant studies, and recent papers that consider clinical utility, whilst their criteria are not as comprehensive, do substantiate the choice of parameters presented here.

The clinical utility of the MDADI

Of the a priori clinical utility criteria, not all were possible to investigate qualitatively via the survey method in this study, for example the existence of normative MDADI data, or data on the MDADI MCID (R3 and R4 respectively in Table 62 above). Criteria not possible to investigate, or previously addressed in the Literature Review, are highlighted in red in the table.

7.2.2.1 Relevance

Many aspects of the relevance domain of clinical utility also speak to a tool's content validity, and therefore have already been discussed earlier in this chapter. Of the aspects listed in Table 62, items not addressed elsewhere and within the scope of this study are R5 (Are questions likely to be distressing for patients?), R6 (Is it possible to use a proxy for patients with special needs?) and R9 (Is the recall timeframe appropriate?).

R5 Are questions likely to be distressing for patients?

This has a direct link to the “Not user friendly” theme in the qualitative analysis of this study. This study presents evidence that the tone of the MDADI is overly negative and therefore potentially unnecessarily distressing for patients to complete. This finding concurs with Greenhalgh et al. (2017)’s assertion that there is potential for the act of completing a PROM to cause patient distress: items in the tool may remind patients of the ways their life has changed and the impact that their symptoms have on many aspects of their life. However, this thesis also suggests that a change in tone and rewording of items to be more inclusive, empowering and less negative might go some way to make the patient’s experience of completing the MDADI less upsetting. Clinicians in this study made many practical suggestions for how MDADI items could be reworded, however these would need to be ratified and corroborated in further research involving people with HNC.

Data in this thesis which suggests that the MDADI is not very user-friendly echo findings of PROM studies from other areas of healthcare. Boyce et al. (2014), in their systematic review of professionals’ experiences of using PROMs, highlighted that patient distress was reported in many studies they reviewed. This finding was also corroborated by Duncan and Murray (2012) in their systematic review of barriers and facilitators to outcome measurement by AHPs which highlighted concerns that PROMs might cause distress.

R9 Is the recall timeframe appropriate?

Results of this study suggest that the 1-week recall timeframe of the MDADI is appropriate. Published evaluations of other dysphagia related QoL tools do not investigate this aspect. Studies from the wider evidence base present a mixed picture of the appropriacy of this recall period. Magasi et al. (2012) suggest this period has ecological validity and incorporate it into their patient-reported outcomes measurement system. Conversely, in their study looking at a tool measuring oral health related quality of life, Waller et al. (2016) found no difference in SEM for tool scores in a comparison between groups of patients who completed the tool using either a 7-day or a 1-month recall period. However, as the MDADI is used at various timepoints during a patients’ cancer journey, and sometimes over multiple closely timed datapoints (e.g., 1 month, 3 months and 6 months post treatment in research studies), during periods where acute toxicities and QoL may be rapidly changing, a short period for recall would seem to be appropriate. However, further formal investigation is warranted to confirm this assertion.

Relevance domain amendment

R6 Is it possible to use a proxy for patients with special needs?

A relevance-related criterion of clinical utility generated a priori was “Is it possible to use a proxy for patients with special needs?”. Data generated by the survey provided evidence for altering this criterion to be more inclusive of other ways of making tools more accessible for patients with special needs, but that use of a proxy per se has inherent complexity. Ideally tools are accessible to as many patients as possible, without any group being blocked from using the tool. Use of a proxy is one solution to this, however data generated in this study suggested that the use of a proxy or having someone else present while a patient completes the tool might influence their responses. This supports previously published data: in their qualitative study exploring clinician and patient opinion of PROMs used in rare diseases, Aiyegbusi et al. (2020) describe how respondents in this study raised concerns that use of a proxy had to be done with care to selection of someone who could truly reflect the service users’ opinions.

However, it is not just for patients with special needs that care needs to be taking when considering proxy responses. Patterson et al. (2013)’s qualitative study exploring the impact on carers of caring for someone with post-HNC treatment dysphagia provides evidence that dysphagia can put relationships under stress. Carers can feel ‘responsible’ for patients’ oral intake, which causes stress for the carer and can be experienced as nagging or coercive by the patient. This could mean that if a carer is present when a patient is completing a PROM, there may be potential for unspoken or overt influence on a patients’ responses. Patterson’s data also suggested that carers and patients did not necessarily have the same concerns, suggesting that carers completing an MDADI on behalf of a patient might not produce valid data.

Therefore, I would argue that this criterion could be expanded to include other ways of increasing tool accessibility. How exactly to amend the tool to facilitate accessibility, for example for patients with cognitive impairment or a learning disability was not described in the data. This issue has wider implications than just for the MDADI: it has relevance to any PROM. Therefore, I suggest going forward this criterion of clinical utility ‘relevance’ is amended to:

> “Is it possible to use a proxy *or make the tool more accessible* for patients with special needs?”

7.2.2.2 Usability

This domain also links to the qualitative theme of 'Practical Issues'.

U1 Does the tool have an appropriate literacy level?

This thesis presents clinicians' concerns that the literacy level of the MDADI may be too complex for many people with HNC, thus affecting their ability to complete it. This finding provides qualitative validation to Zraick et al. (2012)'s quantitative analysis of the 'readability' of the MDADI, which found the MDADI to be the 'most difficult to read' of all of the swallowing-related PROMs examined in their study.

U4 Is the time taken to administer acceptable?

This thesis evidences that clinicians value tools that are short and succinct. In their literature review of barriers to using PROMs in routine cancer care, Nguyen et al. (2021) found that time taken to administer the tool was the most frequently cited barrier to PROM use by both patients and clinicians. Popular PROMs in other areas of SLT practice such as the Voice Handicap Index (VHI), a 30 item voice-related PROM, (Jacobson et al., 1997) have recently been investigated for item reduction to make shorter, less time-consuming versions such as the 10 item VH10 (Rosen et al., 2004). Many currently used swallow-related QoL tools are lengthy (e.g., 31 items for the HNSAM, 44 items for the SWAL-QOL) with implications for administration time. The length of the 20-item MDADI has previously been highlighted as potentially overly long (Lin et al., 2022) and this thesis confirms that clinicians (and by proxy patients) agree with this assertion.

U5/U6 Is the scoring process straightforward and time taken to score acceptable?

This study demonstrated a mixed response regarding the MDADI scoring process, with some respondents finding it challenging and others denying it was an issue. However, taken in the context of the results from the analysis of the quantitative data, which found a clinician MDADI-C score calculation error rate of 24.76%, this would appear to confirm that ease of scoring of the tool may be an issue. This 24.76% did not constitute 'transcription errors' as described by Hong et al. (2013), but rather errors of calculation when working out the MDADI-C score from subscale scores. In the only comparable study in the literature, Lin et al. (2022) did not provide any data on such a scoring accuracy check, referring instead only to 'missing data'.

This high error rate implies the possibility of incorrect scores being subsequently recorded in clinical notes or entered into research databases. The MDADI-C calculation which is prone to error as demonstrated in this thesis is a weakness of the MDADI. One potential solution to this would be for teams to use a webapp or spreadsheet with an inbuilt MDADI score formula

to allow automatic score calculations from individual patients MDADI item scores, as suggested by clinicians in this study.

There is no guidance in the literature for what constitutes an acceptable time to take for scoring of a PROM, however it could be argued that characteristics such as inclusion of 'reverse scored' items within the MDADI overcomplicate its scoring, adding to time taken and difficulty of the scoring process (Kroenke et al., 2015). The MDADI contains two 'reverse scored' items that were consistently highlighted in this study as being problematic, both in terms of potentially catching patients out and causing difficulties with tool scoring. This confirms Sexton-Radek and Simmons (2018)'s suggestion that reverse-scored items can lead to item misinterpretation. Issues with reverse-scored items have been corroborated in other studies, such as Carlson et al. (2011)'s investigation of a depression scale used with older people. This study found reverse-scored items in the measures to be more frequently responded to atypically and to have lower internal consistency.

Usability domain additions

Thematic Analysis of the survey data generated an entire theme around clinical utility-related issues which impact on the use of the MDADI in practice (Theme 2: 'Practical issues'). This included identification of two additional criteria that I propose be added to the clinical utility taxonomy, under the 'Usability' domain as follows.

U9 Is the layout of the tool acceptable?

Layout is related to the length of a tool, which also feeds into potential rationale for item reduction, but also concerns aesthetics and how a tool may look potentially 'off-putting' as evidenced in this thesis. Layout was not an issue that came up in the review of the literature around clinical utility parameters. However, data generated from the survey in this thesis did form a motif around the layout of the MDADI, indicating the amount of text and layout of the 20-item, two-page tool are an issue. As layout is again a practical aspect of a tool, has been indicated by clinicians to be a meaningful aspect of tool use, and is not already covered in other clinical utility criteria, I suggest it is added to the 'usability' domain as a clinical utility criterion with globally applicable relevance.

U10 Is the response modality of the tool acceptable?

On Thematic Analysis of the survey data, it was clear that the response modality of the MDADI (i.e., the Likert scale) had a significant impact on its clinical utility in the eyes of clinicians, who suggested that this format had the potential to affect patients' responses, therefore not giving a true reflection of their situation. Clinicians highlighted their experience of 'anchor effect' in patients' responses to the tool. 'Anchor effect' for Likert scales, where

respondents are less likely to endorse the ‘extreme’ ends of the scale, is a well-documented phenomenon (Bishop and Herron, 2015) negatively affecting the validity of responses.

Respondents in this study suggested that alternative response modalities to a Likert scale might produce more reliable and valid data, for example visual analogue scales (VAS). Evidence does exist in the literature to support this clinical instinct: Voutilainen et al. (2016) compared patient responses to satisfaction surveys via VAS and Likert scales, and found VAS less susceptible to confounding factors and ceiling effects than Likert scales. Bishop and Herron (2015) also suggest that VAS may be easier to assess statistically.

Data presented in this thesis evidence the fact that response modality can affect the usability of a tool. As this is a practical aspect of a PROM as defined above, which is not covered by the existing clinical utility criteria, I suggest it is added as an extra clinical utility criterion in the ‘usability’ domain.

7.2.3 Summary of Research Question 1 discussion

In summary, this thesis argues that the content validity and clinical utility of the MDADI are flawed. Based on the COSMIN assessment criteria for rating content validity, the overall rating for the MDADI is ‘insufficient’. The concept and assessment of clinical utility has been developed via a novel qualitative approach which has not previously been used to examine existing dysphagia related QoL tools. Clinical Utility as a concept, and the contents of the Clinical Utility assessment taxonomy devised in this study, have received initial corroboration in the data analysis and undergone a subsequent refinement. The clinical utility of the MDADI has been found to be problematic across multiple parameters of the taxonomy.

7.3 Research Question 2

What are the results of an exploration of psychometric properties of the MDADI on a UK population using Item Response Theory?

The results of this study with respect to Research Question 2 will now be discussed. This section will consider the structural validity and internal consistency of the MDADI.

7.3.1 Structural validity

Mokkink et al. (2010c) present the COSMIN definition of structural validity as "the degree to which the scores of an … instrument are an adequate reflection of the dimensionality of the construct to be measured" (p.743). Chen et al. (2001) assert that the MDADI subscales are unidimensional and assesses different aspects of the same construct: "four subscales were developed to tap the different effects of dysphagia on QoL … [the] questions within the scale are consistently assessing the same issues" (p. 875).

With regards to IRT analysis of structural validity, COSMIN requires that the "chosen model fits well to the research question" (Mokkink et al., 2018 p.48).

IRT analysis in this thesis confirmed the structural validity of the MDADI in terms of adequate fit with the Graded Response Model. IRT analysis of the MDADI has not been performed previously therefore it is not possible to directly compare this finding with other results in the literature.

For the comparable CTT analysis, the COSMIN guidelines require satisfactory results of a confirmatory factor analysis (CFA) to satisfy this assessment parameter. To date there is no published CFA of the English-language version of the MDADI; of note is that CFA was not carried out in the initial validation of the MDADI by Chen etal.

There has however been a CFA performed on a translated version of the MDADI. In their analysis of the psychometric properties of the Dutch version of the MDADI, Speyer et al. (2011) performed a CFA on 64 MDADIs, which did not confirm that the subscales of the MDADI were homogenous, i.e., they found that the Dutch translation of the MDADI is not unidimensional. Given the small sample however this result does have to be viewed somewhat sceptically, in addition to the fact that translation into another language may also affect the validity of a tool, both in terms of quality of translation and cultural differences around eating and drinking in other countries and cultures. Speyer etal's results of non-unidimensionality contradict the findings of this study that the MDADI is unidimensional in terms of GRM model fit.

7.3.2 Internal consistency

The COSMIN checklist defines Internal Consistency as concerning item 'interrelatedness' (Mokkink et al., 2010a), and states that this parameter should only be considered for unidimensional subscales. As discussed in the previous section, the results of Speyer etal have cast some doubt on whether the subscales of the MDADI are truly unidimensional. However, as the analysis presented in this thesis following COSMIN guidance suggests the MDADI *is* unidimensional, internal consistency has been assessed and will now be discussed.

Of the previously published secondary analyses of the MDADI's psychometric properties, only Timmerman et al. (2014) assessed the domain of internal consistency, finding it to be 'indeterminate', which they define as "doubtful design or method … or only fulfilling part of the requested psychometric property" (p.192). It appears the authors gave the MDADI an indeterminate rating due to the small sample size used in the initial MDADI validation, but this is not made explicit.

Carrozzino et al. (2021) suggest that the traditional approach of assessing internal consistency, with its focus on item heterogeneity, is antithetic to clinical reality where some symptoms reported by a tool may carry more clinical weight or significance than others. They suggest that all items should not be 'weighted the same' and do suggest that the item-level approach of modern statistical models (such as IRT) may be more appropriate to use than CTT approaches such as Cronbach's alpha.

The COSMIN manual's guidance for IRT analysis of internal consistency is brief and suggests calculation of the standard error (SE) of θ; however, it does not provide guidance as to what an acceptable level of SE might be. There is no clear guidance in the wider literature as to what constitutes 'acceptable' SE. When considering SE, a major difference between CTT and IRT is that CTT assumes SE applies to a whole sample rather than analysing by level of latent trait, or on an item-by-item basis. IRT however can provide this more granular item and latent trait level detail.

TIFs for the MDADI at both timepoints presented in this thesis show a marked difference in range of θ covered by the MDADI at either timepoint, and SE data presented shows considerable variation across levels of both negative and positive θ for both timepoints. This thesis also demonstrates that SE in the MDADI increases for more extreme ends of the θ continuum. Nima et al. (2020) in their IRT analysis of a PROM designed to measure subjective wellbeing also replicated this finding. They suggest this information can inform decision making about item selection for future iterations of tools, where data on SE and range of θ covered will feed into whether items will be considered for inclusion or removal. A

tool should contain items that 'work' to provide cover across a full spectrum of range of θ with acceptable SE.

Olino et al. (2012) suggest that this level of SE and θ data provided by IRT analysis also has the potential to influence the choice of context for using an assessment. Their analysis of three depression measures using IRT indicated that different tools were more accurate for different levels of θ. This may be the case for the MDADI: in this thesis the range-of-θ analysis made possible by IRT shows the ceiling effect of the tool at the pre-treatment point, where many patients are asymptomatic and therefore have no impairment in their swallowing-related QoL. The wider range of θ coverage at the post-treatment data point suggests the tool's results may be more meaningful at this point, when a wider range of θ is also expected as patients present with dysphagia following HNC treatment.

IRT analysis of the MDADI's internal consistency has not previously been carried out in other published work, therefore no studies are available to compare directly with this thesis' results. IRT analysis produces detailed data which is not directly comparable with CTT analysis of the same parameter. The COSMIN guidelines for CTT analysis of internal consistency require the metric of Cronbach's α to be presented (Prinsen et al., 2018). The MDADI origin study reports an overall tool Cronbach's α of 0.96, which exceeds the value of 0.7 which is generally considered to be the cut-off of 'acceptable' for this parameter (Tavakol and Dennick, 2011). It is of note however that generally it is recommended that internal consistency analyses are carried out on a minimum data pool of x7 the number of items in the tool in question (Timmerman et al., 2007). For the MDADI, which has 20 items, this figure would be *n*=140, however the MDADI origin paper's analysis only included *n*=100 MDADIs. Lin et al. (2022) report an overall Cronbach's α of 0.939 for their dataset of 196 MDADI questionnaires; they propose this high value suggests redundancy of items in the original scale.

It is also important to observe however that Cronbach α values presented in the Chen etal paper for MDADI subscale analyses were significantly lower: 0.58 for the functional subscale and 0.69 for the emotional subscale, which are lower than the 0.7 'acceptable' cut-off. This suggests issues with the internal consistency of MDADI subscales. Lin et al. (2022) did not present subscale Cronbach's α values for their MDADI dataset but did report a Cronbach α value of 0.9 for their 'miniDADI' shortened scale. When a Cronbach's α is calculated for the MDADI data presented in this thesis, the following results are obtained:

	N	**Whole test**
Pre-treatment	274	0.9575
6m post treatment	146	0.9474

Table 63: Cronbach's α values for MDADI data

It has been suggested in the literature that a higher Cronbach's α is not always 'better' (Tavakol and Dennick, 2011), and that indeed a value >0.9 may indicate item redundancy within a scale, suggesting that it could be shortened. Cronbach's α values for the MDADI full scale reported in the literature, as well as those reported in this thesis, all exceed this threshold and could therefore be seen to give weight to the argument presented here that the MDADI contains redundant items.

It is also noteworthy that CTT and Cronbach's α/internal consistency does not differentiate between different levels of θ (O'Connor, 2018). The IRT analysis results presented in this thesis however indicate that internal consistency does in fact vary across different levels of θ for the MDADI, thus showing the CTT assumption to be an oversimplification.

The findings of this thesis suggest that internal consistency of the MDADI is more complex than previously appreciated by preceding studies. This is replicated in research reporting IRT analysis of other dysphagia outcomes tools. In analyses of the EAT-10 and SWAL-QOL scales, these tools, for which acceptable Cronbach's alpha scores had been published in previous psychometric evaluations, demonstrated considerably poorer and more variable internal consistency when assessed with IRT techniques (Cordier et al., 2017, Cordier et al., 2018).

In summary then this thesis demonstrates that IRT analysis of the MDADI's internal consistency produces a picture of variation of item accuracy within the tool by item and across datapoints, which is considerably more nuanced than any previously published analysis. However, use of IRT to assess this parameter of tool properties is complex and results of this study would need to be corroborated and expanded upon with future work.

7.3.3 Summary of Research Question 2 discussion

The results presented in this thesis add new data to the paucity of existing re-evaluation of the MDADI's psychometric properties. IRT analysis is a novel approach and has not previously been carried out with this tool, however this also means that there are no other extant IRT data with which to compare or corroborate this thesis' findings. Results of this thesis are equivocal in many respects with regards to both the structural validity and internal consistency of the MDADI and highlight the need for further in-depth analysis of the MDADI's properties with IRT. However, when taken in context of the significant content validity and clinical utility issues inherent in the MDADI presented here, potential issues with structural validity and internal consistency in the MDADI's current form may be moot.

7.4 Research Question 3

What are the potential factors which might result in Differential Item Functioning?

Qualitative data analysis of survey responses from UK clinicians working with patients with HNC indicated that age, sex and socioeconomic status might influence patients' responses to items within the MDADI. Following subsequent IRT analysis, DIF was confirmed for multiple items across all three of these variables at the pre-treatment datapoint, and for the variables of sex and socioeconomic status in one item each respectively at the post-treatment datapoint.

The presence of DIF implies that scores on a tool are not comparable across groups for the variable in question, even when the respondents' levels of θ are the same (Jones, 2019). DIF can therefore indicate item bias, allowing items identified as being subject to DIF by IRT analysis to be discounted or removed from tools.

This study is the first to explore the potential for DIF within the MDADI tool, however other dysphagia assessment tools have been evaluated for the presence of DIF. Cordier et al. (2017)'s IRT analysis of the EAT-10 dysphagia outcome tool found DIF for variables of diagnosis of dysphagia, language, gender and comorbidities across multiple items of the tool. The high prevalence of DIF within the tool contributed to the authors recommending that the EAT-10 requires redevelopment prior to continued clinical use. Similarly in their IRT analysis of the SWAL-QOL dysphagia PROM, Cordier et al. (2018) identified DIF across multiple items for multiple variables, in particular language and gender. The presence of DIF, in addition to other psychometric concerns highlighted by their investigation, again contributed to the team suggesting the use of the SWAL-QOL in clinical practice in its current form is not recommended.

These results, although considering other tools, show similarities with the current investigation of the MDADI: this thesis found the MDADI also to be subject to DIF across multiple variables and items. This therefore means that sex-based bias has now been shown across three different dysphagia PROMs: the MDADI, the EAT-10 and the SWAL-QOL.

In their European Society for Swallowing Disorders white paper on Screening and Non-instrumental Assessment for Dysphagia in Adults, Speyer et al. (2022) mention IRT including DIF analysis as a relevant future approach for psychometric assessment of tools. Certainly, the emerging evidence presented by this thesis and the work of Cordier etal suggest that DIF analysis adds an extra dimension to tool analysis and can uncover previously unacknowledged issues with tool bias.

7.4.1 Adverse versus benign DIF

It is noteworthy that the presence of DIF does not automatically indicate bias. DIF may be benign or adverse (Breslau et al., 2008): sometimes a variable causing DIF can be 'auxiliary' to the underlying trait under assessment and therefore be considered 'benign'. 'Adverse' DIF conversely is truly a source of bias. Quantitative techniques for distinguishing between adverse and benign DIF do not exist; rather a qualitative judgement must be made. If items with adverse DIF are identified they can be considered for removal from a tool (Teresi and Fleishman, 2007). I would argue that the variables identified in this thesis (age, socioeconomic status and gender) constitute adverse DIF for MDADI items, as the clinicians who identified them indicated concern regarding potential item response bias across the different groups that was not related to dysphagia per se. Therefore, the presence of adverse DIF across multiple variables and items of the MDADI further brings into question the tool's overall validity and opens up the possibility of item removal. DIF-informed item reduction for the MDADI will be discussed later in this chapter.

7.4.2 Summary of Research Question 3 discussion

MDADI items have been shown in this thesis to be subject to DIF, which adds an extra element to concerns about the robustness of the tool in addition to previously presented issues with content validity, clinical utility and internal consistency. However, DIF is a complex concept and judgement of what constitutes adverse or benign DIF is subjective. As DIF in the MDADI has not been previously investigated, and this thesis suggests DIF is present within the tool, this is a key area for future research to investigate further. The DIF identified in particular MDADI items in this thesis does however create the opportunity for this result to feed into consideration of item redundancy and creating a shorter version of the MDADI, which will be discussed in the next section of this chapter.

7.5 Synthesis of results from all Research Questions

This thesis has demonstrated how qualitative data and analysis can inform and enrich quantitative analysis of a PROM. In their review of conceptual models to guide analytical integration in mixed method studies, Moseholm and Fetters (2017) suggest that for 'convergent design' studies such as this thesis, a 'data diffraction' approach is taken which emphasises how the qualitative and quantitative data illuminate different aspects of a 'central phenomenon'. Qualitative data provided information on MDADI clinical validity that could not have been quantitatively generated; in addition, qualitative data helped to guide and inform quantitative assessment.

Combining qualitative and quantitative data analysis of diverse psychometric and clinical utility aspects of the MDADI has produced a complex and rich 'meta inference' of the

strengths and weaknesses of the tool. The mixed methods approach to analysis has illuminated many previously unexplored and unreported issues with the MDADI. This approach is innovative and has not previously been attempted for investigation of the properties of dysphagia-related PROMs.

A major opportunity provided by qualitative and quantitative data analysis synthesis in this study was the potential to make an in-depth proposal for MDADI item reduction, through mixed methods analysis of the tool at an item level.

7.5.1 Item reduction

The length of the MDADI was highlighted as a major issue in this thesis. Snyder et al. (2010) suggest that being able to select the most relevant items from a longer research tool, to make a shorter more useable tool for clinical practice, may be a relevant avenue for future PROM research to explore. The idea of a short-form tool as having more clinical relevance is echoed by Carrozzino et al. (2021) who suggest that tool length is a key feature of the clinical utility of a PROM. In their study considering the clinical utility of a HRQoL PROM used with people with cancer Benavent et al. (2021) reported that clinicians considered the length of the tool in question to negatively impact on its clinical utility, and that shortening the tool would improve its utility.

Item reduction analysis was not originally a separate research question singled out for this study. However, as the analysis of results developed it became clear that this was a significant topic and one worthy of detailed consideration, given the potential positive impact for practice that generation of a robust shortened version of the MDADI would have. This, coupled with the publishing of Lin et al. (2022)'s study during the undertaking of this thesis, which takes a different approach to MDADI item reduction, suggested that this was an important area to probe using the unique methodology of this thesis.

This thesis presents a more comprehensive analysis of MDADI item redundancy than is extant in the published literature. Whilst other studies have considered item redundancy in PROM tools from a purely mathematical perspective (Sekely et al., 2018, Lin et al., 2022), this thesis took an approach combining qualitative and quantitative data.

Sekely et al. (2018) used IRT and item information values to guide their choices regarding creating both a theoretically driven and empirically driven short form of a tool for assessing alexithymia. They did not consider DIF or incorporate any qualitative or content validity-related data into their synthesis.

With specific reference to the MDADI, only one other paper has considered item reduction: that of Lin et al. (2022) which detailed the team's quantitative analysis of MDADI item reduction.

Lin et al. (2022) took a factor analysis approach with a view to producing a shortform version of the MDADI which they have named the 'miniDADI'. This factor analysis was performed on 196 MDADI questionnaires completed at 3-months post-treatment. Their analysis and item reduction were purely empirical and did not also consider a 'theoretical' approach as suggested by Sekely et al. (2018). The final MiniDADI version proposed by Lin contains MDADI items 11, 12, 14, 18 and 20, comprising two functional, two emotional and one physical subscale items, so does not preserve the 'theoretical ratio' of subscales (1F:2E:2P). A comparison between Lin's MiniDADI and the synthesised MDADI shortforms proposed in this thesis is displayed in Table 64:

	MiniDADI (Lin et al., 2022)	**Pretreatment Shortform MDADI**	**6m post treatment Shortform MDADI**
Items	11, 12, 14, 18, 20	6, 11, 12, 14, 17	6, 7, 11, 12, 20

Table 64: Comparison of Lin's MiniDADI with shortform MDADIs proposed in this study

As demonstrated in the above table, there is 60% overlap between the pre-treatment MDADI shortform suggested by this study, and Lin's MiniDADI. At the 6m post-treatment point, there is also 60% overlap, but the item makeup of this overlap is slightly different.

A significant advantage of the current study's approach over Lin's approach is that this thesis incorporates content validity and clinical utility data; Lin's study does not. Of note is that in Lin's selected items of 11, 12, 14, 18 and 20 for the 'miniDADI', items 12, 14 and 18 are suggested by clinicians in this thesis as having problematic wording.

Lin et al. (2022) also did not consider potential DIF in their item reduction analysis. Given the results of this thesis, this has the potential to significantly weaken their results, as 40% of the MiniDADI items are subject to adverse DIF. This thesis has found both items 18 and 20 (both included in Lin's MiniDADI) to exhibit DIF at the pre-treatment datapoint, and item 18 also at 6 months post-treatment, which removed them from consideration for the shortform MDADIs presented here.

This thesis has taken a novel approach to item redundancy assessment, with its synthesis of both qualitative content validity and clinical utility factors, and quantitative IRT-led item information assessment. This provides a more complex and nuanced assessment than has previously been published in the literature and potentially provides a template for a mixed methods approach to PROM item reduction in future research.

7.6 Critique of study

When reflecting on the process of undertaking any project with the benefit of hindsight, there are always aspects that, on reflection, could have been executed differently. A critique of aspects of this thesis follows.

7.6.1 Study design

7.6.1.1 Lack of patient data

My biggest regret for this study was the necessary but intensely disappointing COVID-related change to the study design, which involved removal of collection of patient-generated data on MDADI content validity and clinical utility from the protocol.

Relevance of PROM items to patients' concerns is central to ensuring strong content validity (Francis et al., 2016). It is essential that patients are included when developing, and also it could be argued assessing the quality of, tools (Mokkink et al., 2010a). Detail on how patients' views were incorporated into the development of the MDADI is minimal, and certainly no research prior to this thesis has been undertaken to corroborate MDADI item content and choice from a patient's perspective.

Addario et al. (2020) also highlight the importance of including the patient voice when developing PROMs for use in cancer clinical trials. However, they also recommend the wider stakeholder group (including clinicians) be included when developing these tools. Asking clinicians about tool content remains a relatively novel approach; nevertheless, the depth of data generated from the clinicians in this study provides proof of concept that it is possible to produce relevant and useful qualitative data on PROM content validity and clinically utility by including clinicians in PROM evaluation.

The results of this thesis need to be considered with caution however, avoiding the assumption that clinicians' proxy reporting of patients' experiences is always reliable. In their mixed methods study investigating the opinions of patients and clinicians on QoL PROM tool content (tools aimed at assessing impact of cancer medicines on HRQoL) Dunlop et al. (2022) found only 60% overlap in themes generated by clinicians and patients. This highlights the facts that clinicians and patients' priorities may not align, and therefore that clinicians' comments on patients' views in this thesis should be viewed with an element of caution.

7.6.2 Data analysis

7.6.2.1 Ensuring rigour in qualitative analysis

Morse (2015) suggests that rigour of qualitative analysis is achieved through prioritising reliability, validity and generalisability during the data collection and analysis phases of a study. What follows is a critique of the rigour of the qualitative data collection and analysis performed in this thesis.

7.6.2.2 Validity and reliability

Some recommendations for ensuring rigour in qualitative research that are described in the literature were not possible to incorporate into this study due to the small scale and timeframes involved. Morse (2015) recommends 'prolonged engagement' and 'persistent observation' as ideal strategies to ensure validity. However, the nature of the study design of this project meant that prolonged engagement was not possible with the questionnaire method utilised, or within the timeframe of this study given the mixed methods approach: qualitative data collection was only half of the process. Likewise, persistent observation was not applicable to the questionnaire method.

When thinking about possible ways of verifying the validity of my qualitative data, both 'member validation' and 'outsider validation' were considered. Member validation is a facet of internal validity, looking at verification and substantiation of qualitative data (Ritchie et al., 2014). This technique reduces the potential for researcher bias in results/analysis by actively involving the participants in confirming and checking the results (Birt et al., 2016). However, not all authors and researchers agree that member validation adds value: Morse (2015) argues strongly against using this technique, suggesting that it is not practical, as individual members may not agree with the analysis as a whole, which is a synthesis of many opinions and not just theirs. Morse also conjects that there is no extant guidance for what to do if a participant doesn't agree with an analysis and proposes that "the researcher's background in theory and research methods must outrank the participant as a judge of the analysis" (p. 1216).

In this study, I considered this approach to data validation however ultimately decided against employing these techniques. Outsider validation was dismissed as a recruitment path to this was not clear: I had already tried to access all relevant clinicians for the initial survey and did not know who had already answered the survey as it was anonymous. Therefore, I had no way of ensuring accessing 'outsiders' only. Member validation was dismissed as the low response rate had confirmed that in the post-COVID climate significant demands on the NHS workforce may preclude engagement with research recruitment. As I

had already approached relevant clinicians once, I did not want to put additional pressure on their time to re-evaluate my data through a further process of member validation.

In my position as a clinician-researcher, I am embedded in HNC practice and the MDADI is part of my daily work. Through discussions with colleagues in my geographical area but also around the UK over my 20+ years of practice, I have a general insight into colleagues' impressions of the MDADI. This gave me some 'validity context' when analysing the data to influence my analysis of themes, patterns, and the potential for unexpected or 'outlying' data.

Reliability during the analysis phase of this study was established through developing a clear coding system and looking for data saturation within codes and themes. I also went through a careful process of 'disassembling and reassembling' the data whilst forming codes and themes: a process that Castleberry and Nolen (2018) suggest is a key aspect of a rigorous qualitative analysis.

A weakness of the validity and reliability of the clinician generated data in this study is that it was not triangulated with patient-generated data. This was due to the previously discussed COVID-related protocol change and was unfortunately unavoidable.

7.6.2.3 Generalisability

My findings represent the opinions and experience of UK clinicians practising in England and Scotland only. As the MDADI is an internationally used tool, and eating, drinking, swallowing and dysphagia are very culturally sensitive, it may be that the results of this thesis are not generalisable to other locations. Further triangulation with data from other countries, and with input from clinicians and patients around the world, would improve this aspect.

7.6.2.4 Reflexivity and potential for bias

In terms of inherent research bias, as a practising clinician who has used the MDADI with patients for many years, I acknowledge that I have developed opinions about the tool over this time. It could be argued therefore that I do not have a starting point of equipoise about the MDADI. Malterud (2001) describes reflexivity as 'the knower's mirror'. During the analysis process I was very aware of my potential lack of equipoise and took steps to incorporate reflexivity and monitor for bias, using a reflective diary as my 'mirror' to monitor and manage the personal thoughts and reactions I had to the data and the analysis process.

7.6.2.6 Representativeness and size of sample

Qualitative data

Although circulated around clinical networks accessible to all four UK nations of England, Scotland, Wales and Northern Ireland, a total of 31 questionnaire responses were recorded from SLTs practising in England and Scotland only. This number is comparable to other

practice-related survey-based research carried out with UK HNC SLTs: Roe et al. (2012) had 46 respondents to their questionnaire on UK HNC practice. However, although only 31 respondents completed the survey, they provided a wealth of information relevant to the research questions at hand. Variation in practice in use of the MDADI within this study's survey cohort mirrors the general variation in HNC SLT practice evidenced by Roe et al. (2012) and also that of US-based HNC SLTs surveyed by Krisciunas et al. (2012).

Quantitative data

MDADI data from patients used in this study was collected prospectively and consecutively and constitutes a 'real life' sample of patients attending the Edinburgh Cancer Centre for curative HNC treatment from 2016-2021. In terms of MDADI scores for this population, pre-treatment and 6-month post-treatment MDADI-G and -C mean scores are comparable with other sites around the UK, such as those published by Khan et al. (2015).

The MDADI was originally validated on 100 patients (Chen et al., 2001). More recent research considering properties of the tool has involved varying numbers: Khan et al. (2015)'s study comparing the MDADI-G and -C scores with the Performance Status Scale (PSS) included data from 86 patients pre-treatment, and 39 at 6 months post-treatment; Pedersen et al. (2016)'s comparison of the MDADI, PSS, Water Swallow Test and the Penetration-Aspiration scale involved MDADI-C score data from 173 patients at 3 months post-treatment. Lin et al. (2022)'s factor analysis of the MDADI was performed on data from 196 questionnaires. The current study included item-level analysis of 271 pre-treatment and 145 post-treatment questionnaires, which represents a significantly larger total number than previous published research into the properties of the MDADI.

7.6.2.7 Use of the COSMIN tool

The COSMIN tool was chosen for use in this thesis study as it went through a rigorous development process, it contained guidance on IRT analysis, and recent additions had been made with specific regard to content validity assessment.

Baijens et al. (2021) recommend the use of IRT for psychometric assessment of assessment tools used with HNC patients with dysphagia and cite the COSMIN tool as an appropriate way of evaluating the properties of existing measures. Carrozzino et al. (2021) however describe COSMIN as being "strongly rooted in the psychometric tradition" (p.224) with potential for insufficiency in the scope of its domains in terms of assessing PROMS for clinical use. In their comparison of the COSMIN tool with other outcomes tool assessments, Rosenkoetter and Tate (2018) found a lack of consistency of assessment definitions and standards across tools, meaning that comparison was confounded and a clear recommendation for the tool with the strongest qualities could not be made. These critiques

in the literature suggest that the COSMIN tool may need further development, and this needs to be acknowledged in regard to the analysis presented in this thesis.

Considering the lack of detail specifically regarding IRT, although COSMIN does provide some guidance with respect to IRT analysis of psychometric parameters, this is much less detailed than the CTT guidance provided for the same parameters: IRT guidance provided by COSMIN is at best brief. Results presented in this thesis suggest that COSMIN IRT guidance concerning structural validity and internal consistency is not sufficiently detailed, and does not consider the value and relevance of other IRT analyses that would provide additional valuable data when making judgements about the quality of the tool, for example DIF. This also adds to the suggestion that the COSMIN tool needs further development specifically regarding the guidance it provides on IRT assessment of tools.

The presence of DIF for multiple items within the MDADI demonstrated in this thesis could be seen to call into question the structural validity of the tool. The fact that, using COSMIN guidance, the 'model fit' of the tool was taken to be the method of choice for assessing structural validity, and the MDADI was shown to fulfil this criterion, but then later also shown to contain a significant proportion of items subject to potentially adverse DIF, calls into question the use of the 'model fit' method as the sole means of assessing structural validity by IRT means as suggested in the COSMIN tool. Perhaps it is not sufficient, and other analyses/factors such as DIF should also be considered when assessing this parameter via IRT. The current version of the COSMIN manual also only mentions DIF in the context of cross-cultural validity assessment. Given the potential impact of DIF, perhaps this should be expanded upon in future iterations of the COSMIN tool.

In addition, the COSMIN content validity guidance has much overlap with aspects of clinical utility as defined in this study. It may be that the definition of content validity has been cast too wide in the COSMIN guidance, and that this parameter should be kept narrow and specific, and additional clinical utility criteria added into the guidance that can cover other feasibility related aspects of tool use or performance.

Studies continue to be published following COSMIN guidelines, such as Manduchi et al. (2022) who describe a protocol for a systematic review of psychometric properties of PROMS for dysphagia in HNC which is currently underway. Unfortunately, any issues with the COSMIN assessment criteria, such as lack of reference to clinical utility identified here, will be transferred to any study that uses them.

7.6.2.8 Use of IRT

IRT is a modern, fast-developing approach to psychometric analysis, provides opportunities to produce novel insights into tool properties, and facilitates investigation of DIF. This is why IRT was chosen as the approach to mathematical analysis of aspects of the MDADI's psychometric properties in this study. However, as IRT is a relatively novel approach compared to CTT there are issues with its use: interpretation can be challenging and the metrics it produces are less universally well-understood than classic CTT analyses and metrics such as Cronbach's α. The advantages of IRT analyses have been previously outlined in this thesis. The fact that IRT analysis is recommended in recent multinational HNC guidance (Baijens et al., 2021), and that the MDADI has to date not been analysed using this technique, further justify its inclusion as an approach in this thesis.

7.6.2.9 Missed DIF analysis opportunities

As DIF is analysed at a specific variable and item level, it is possible that this study has missed additional potential DIF analysis opportunities. Taking the qualitative data as guidance for which variables and items to investigate for DIF opens up the possibility that data generated by different groups might have indicated other potential variables to investigate.

The data that were generated in this study also did indicate other potential sources of DIF that it was not possible to evaluate within the available dataset. For example, survey respondents highlighted that they had concerns that patients who are socially isolated or live alone may respond to the MDADI differently. However, as the study design did not allow for prospective collection of data, I did not have access to information on social circumstances to be able to analyse this. This would be worth exploring in future analysis of the MDADI however, as would other variables, for example dentition. Clough et al. (2018) highlight the impact of pre-HNC treatment dental extractions on HRQoL. Again, there is some indication in the qualitative data of this study that this may impact on MDADI item responses, however dentition information was not included in the patient demographic data collected for this study.

In this thesis, items subject to DIF were found to differ slightly between datapoints. Whilst this might be related to statistical power due to a larger *n* at the pre-treatment datapoint, it might instead be due to other characteristics of the two 'populations' before and after treatment. It is well documented in the literature that HNC treatment is lifechanging, and many other factors not investigated here might influence patients' responses to MDADI items following completion of treatment. At 6 months post treatment, subacute post-treatment side effects are often still in effect. In addition to this, patients are adapting to their 'new normal'

and navigating cancer survivorship. Phillips and Currow (2010) have described how cancer survival “means living with a chronic and complex condition” (p.47). In their qualitative study of patients’ food and eating-related coping strategies post HNC treatment, Einarsson et al. (2019) report a range of different coping, avoidance or management strategies that people develop in response to their EDS difficulties following treatment. Different people adapt or react differently to longer term treatment related comorbidities such as dysphagia, so it could even be that individual coping mechanisms might be a potential source of DIF in PROMs.

7.7 Discussion summary

This thesis is the first to present data on mixed methods IRT and qualitative assessment of a dysphagia PROM’s properties. It demonstrates the power of this approach and the detailed information it produces.

Since its inception more than twenty years ago, the MDADI has been subject to surprisingly little scrutiny of its properties, even though it is one of the most well-used tools in HNC clinical and research practice. Most evaluations of the tool purely re-present the development validation data without making any de novo evaluations of the MDADI’s strengths and weaknesses.

The results of this study paint a complex picture of the MDADI tool’s properties. The data suggest that for a UK-based population of clinicians and patients, there are not only issues with the tool at an item level, but also at a whole tool level, with both its content validity and clinical utility, in addition to the presence of DIF in specific items. This study also suggests that the properties of the tool differ if used at different timepoints within a patients’ journey, which may have implications for both clinical practice (e.g., should clinicians use different shortforms depending on when they use it?) and research practice (how can ‘before’ and ‘after’ scores on the tool be compared if it is now indicated that the tool does not perform comparably at different timepoints?).

With regards to use in clinical practice, results of this study indicate that many clinicians use the MDADI as more of a qualitative ‘conversation starter’ to facilitate difficult discussions around dysphagia-related quality of life, rather than a formal quantitative outcomes tool. Therefore, a change to the length of MDADI version with elimination of ‘problem’ items could be a welcome change and potentially easy to incorporate into practice.

Novel IRT analysis of PROMs provides much more detailed, nuanced mathematical data than CTT. However, analysis of established tools with IRT can uncover previously undocumented issues, bringing into question the validity of their ongoing use. This has been the case with other well-known dysphagia PROMS to the extent that researchers suggested

that tools were not fit for purpose in their current forms given their poor performance on IRT analysis (Cordier et al., 2018, Cordier et al., 2017). The initial results presented in this thesis suggest that this may also be the case for the MDADI.

This initial attempt at using IRT to assess some of the MDADI's psychometric properties has highlighted both the wealth and complexity of information that IRT produces, and perhaps also that current guidance on IRT use for psychometric assessment of PROMS lacks detail and requires further clarification.

From a research practice perspective, the results of this study present a significant challenge to the status quo in the field of HNC. The MDADI has been frequently used as a main outcome measure in large scale clinical trials of cancer interventions. It is also still commonly used as a 'gold standard' when validating other tools. The significant psychometric, validity and utility issues highlighted in this thesis suggest potential serious implications for judgements made on treatment efficacy and the validation of other dysphagia assessment tools that incorporate MDADI data.

The combination of detailed mathematical data, coupled with compelling qualitative data, suggests fundamental flaws with the MDADI, highlighting a pressing need for it to be modernised and updated.

In the following final chapter of this thesis, implications of this study are discussed, and recommendations made for future practice, policy and research with respect to the MDADI.

CHAPTER 8: STUDY IMPLICATIONS, RECOMMENDATIONS AND CONCLUSIONS

The proposed outcome of this study was as follows:

> *Qualitative and quantitative insight which will make clear the strengths and weaknesses of the MDADI tool, both in terms of psychometric analysis and the utility of the tool, that is the experience of using the tool in practice. These data will be useful to inform future tool development and use in both clinical and research practice providing guidance on how the validity of a future version of the tool could be improved, and how its use in practice could be enhanced.*

In this final chapter I will therefore present the potential practical outputs of this study and recommendations based on the results.

8.1 Practical outputs and recommendations of this study

This study has confirmed that UK HNC SLTs agree that a tool which looks specifically at the impact of eating, drinking and swallowing (EDS) difficulties on quality of life has clinical relevance. In HNC dysphagia practice we need a PROM that considers the impact of difficulty with EDS on quality of life that is robust, relevant and useable. This need, integrated with the results of this study, suggests the following recommendations for practice, policy and research.

8.1.1 Recommendations for practice

This in-depth analysis of the MDADI has produced results that can translate into practical action points for making evidence-based changes to clinical practice within my SLT team in NHS Lothian:

1. Reconsider current team policy regarding use of the MDADI as part of our suite of routine outcome measures given multiple issues highlighted by this study.
2. Explore other options for dysphagia related QoL measurement which are more acceptable to patients and clinicians until a better version of the MDADI is produced. We might move to trialling another tool, for example the HNSAM.

If we agreed to continue using the MDADI, the following recommendations are made:

3. Create a Microsoft Excel™ spreadsheet with a formula that includes a MDADI-C score calculator to minimise MDADI-C score calculation errors.

4. Create a database of languages the MDADI has been translated into, and who to contact to obtain the relevant version if necessary.
5. Consider using a shorter version of the tool as a clinical adjunct.

8.1.2 Recommendations for policy

1. Published guidance on selection of outcome tools for research and clinical practice gives much emphasis that the tools should be 'robust'. However, this thesis demonstrates just how complex that judgement can be, and how much an impact the clinical utility of a tool can have on its use in addition to its psychometric properties. Future guidance on PROM tool selection should be clearer, and tools for assessing PROMS need to be wider ranging in what aspects of a tool they assess, particularly with respect to clinical utility.
2. This thesis has demonstrated that the internal consistency, content validity, clinical utility and presence of DIF within the MDADI are problematic. This means that without future amendment caution is advised regarding its continued use as sole or primary dysphagia-related outcome tool in clinical trials, or as a gold standard in the validation of other dysphagia assessment tools. This is significant and somewhat controversial due to the current high profile and wide use of the MDADI in both clinical and research practice.
3. There has been much literature in recent years considering the development of a 'core set' of outcome measures in HNC (Nund et al., 2014b, Chera et al., 2014, Waters et al., 2014). Future work in this area should take the results and recommendations presented in this thesis into consideration when determining the content of this core set.

8.1.3 Recommendations for research

This study illustrates the different results that can be obtained by exploring the properties of a tool via a mixed methods approach. Further exploration of issues raised in this study are warranted as they have significant implications for the validity and reliability of MDADI data generated in clinical and research practice.

1. Patient-led assessment of the content validity and clinical utility of the MDADI

As opinions of people with HNC are only represented in this thesis by proxy, it is essential that the data in this thesis are triangulated with patient generated data about the content validity and clinical utility of the MDADI in future studies.

2. Reproduction of the analysis with other geographical populations

IRT assumes population invariance, so in one sense the same results would be expected in another similar analysis of the MDADI. However, EDS are culturally sensitive, so the potential for context to affect results would need to be explored on culturally differing populations. This issue is even more complex for translated versions of the MDADI.

In terms of the quantitative analysis and the difference in this thesis' results compared with Lin et al. (2022), further item-level analysis of the MDADI is recommended to investigate whether current results are replicated in terms of recommendations for item reduction.

3. Future modification of the MDADI?

On one level, the results of this study could be interpreted to suggest that the MDADI as a tool is fundamentally flawed and beyond 'fixing' as a tool, given the many statistical and qualitative issues presented in this thesis. However, the MDADI is unique currently as the only HNC-specific tool which measures dysphagia-related quality of life, and this study has shown that it does still have relevance to clinical practice in some form. To develop a new tool to replace the MDADI would be a mammoth undertaking, particularly due to the complexity of integrating multiple sources of data on tool performance in terms of psychometric profile, clinical utility and content validity for both clinicians and patients.

This study has suggested multiple changes to the existing form of the MDADI to strengthen its validity, clinical utility and psychometric properties. These include a decrease in length through item reduction, removal of 'reverse scored' items, exploration of alternatives to the Likert scale, and amendments increasing inclusivity and compassion of item wording. Guidance does exist for modifying existing tools (Snyder et al., 2007) and this along with the suggestions made in the current thesis and any future research exploring the properties of the MDADI would need to be synthesised to improve the tool as much as possible to a point where it could be re-validated.

4. Further in-depth investigation of existing tools that could replace MDADI using mixed methods study design to assess their content validity, DIF and clinical utility

Due to its high profile within clinical and research practice, much work would need to go into researching a tool that could replace the MDADI. Its closest alternative could be seen to be the HNSAM (Chan et al., 2019), however this has not yet been subject to independent psychometric, qualitative or clinical utility assessment, and does not claim to assess exactly the same domains as the MDADI. Other non-HNC specific tools such as the SWAL-QOL or the EAT-10 have also not undergone the thorough mixed-methods analysis that would be required; this therefore should be a future research priority.

5. Development of a new tool to assess dysphagia related QoL in people with HNC

If the MDADI is no longer considered fit for purpose and it is proven that there are no suitable existing alternatives, future research should focus on developing a new tool to take its place.

6. Increased use of IRT to investigate PROMS, including those assessing dysphagia-related QoL

Yang et al. (2022) suggest increased use of IRT vs CTT in future development of HRQoL PROMs for dysphagia, and that the item level analysis that IRT facilitates confers significant advantage over more traditional psychometric approaches. IRT analysis of all existing dysphagia specific QoL assessment tools is advised to investigate how they perform in terms of difficulty, information, and range of θ coverage. It may also be that synchronous CTT and IRT analysis on the same dataset, as suggested by Cordier et al. (2018) is a sensible approach, and this comparison of more traditional methods with more novel methods might also be a way of increasing understanding or IRT terminology and results amongst researchers and clinicians. For this to be realistically possible, improved guidance on how to carry out and interpret PROM IRT analysis is necessary and should be incorporated into tools designed to facilitate the assessment of PROMS.

7. Further development of the concept of tool ‘clinical utility’

In this study I have made a novel proposal for the definition of the concept of tool ‘clinical utility’ and presented a taxonomy for how this could be assessed. I have made the argument that considering tool ‘clinical utility’ is a key facet of overall tool strength, alongside the more traditional psychometric considerations of validity and reliability. The taxonomy presented here requires further validation and development.

8. Development of a formal tool for the assessment of PROM clinical utility

The taxonomy developed in this thesis, if developed as per point seven above, could lead to the production of a clinical utility assessment tool which could be used alongside psychometric assessment tools when assessing the robustness of PROMs.

8.2 Conclusion

International guidelines on HNC and dysphagia practice stipulate the use of psychometrically robust psychosocial assessment tools (Verdonck-de Leeuw et al., 2022, Speyer et al., 2022). In clinical practice, SLTs see every day the significant and lasting impact that HNC-related dysphagia has on our service users. It is therefore essential that we have tools that are not only psychometrically valid and reliable, but that are also practicably usable with strong relevance to patients and clinicians, and that produce meaningful data that can be used to drive and improve care at an individual patient level. This thesis confirms that a tool specific to assessing dysphagia related QoL has great relevance for clinical practice.

The results of this study have also shown however that the validity and clinical utility of 'classic' assessment tools cannot be assumed and has confirmed the importance of reflecting on and interrogating the tools we use in both clinical and research practice to ensure they are fit for purpose and add value to our patients' management. We cannot afford to be complacent about the quality of the tools that currently form part of 'established' HNC practice. We need to ask Kroenke et al. (2015)'s 'second generation questions' of the tools we routinely use both clinically and in research.

The MDADI is unique in its HNC-specific focus on dysphagia-related QoL assessment and is extremely widely used in both HNC research and clinical practice. However, this thesis suggests there is a pressing need for either further strengthening, development and testing of the MDADI, or development of a new tool that could take its place.

I hope this thesis can act as a clarion call for further investigation, not only of the MDADI but of all commonly used outcomes tools that have an impact on patient care in HNC dysphagia, both through their use in clinical practice and research trials. The recommendations made here can pave the way for the evolution of the MDADI so it can continue to be a key patient-centred and evidence-based adjunct to HNC dysphagia practice.

LIST OF REFERENCES

AARONSON, N., ALONSO, J., BURNAM, A., LOHR, K. N., PATRICK, D. L., PERRIN, E. & STEIN, R. E. 2002. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res,* 11**,** 193-205.

ABBOUD, W. A., HASSIN-BAER, S., ALON, E. E., GLUCK, I., DOBRIYAN, A., AMIT, U., YAHALOM, R. & YAROM, N. 2020. Restricted Mouth Opening in Head and Neck Cancer: Etiology, Prevention, and Treatment. *JCO Oncology Practice,* 16**,** 643-653.

ADDARIO, B., GEISSLER, J., HORN, M. K., KREBS, L. U., MASKENS, D., OLIVER, K., PLATE, A., SCHWARTZ, E. & WILLMARTH, N. 2020. Including the patient voice in the development and implementation of patient-reported outcomes in cancer clinical trials. *Health Expect,* 23**,** 41-51.

AIYEGBUSI, O. L., ISA, F., KYTE, D., PANKHURST, T., KERECUK, L., FERGUSON, J., LIPKIN, G. & CALVERT, M. 2020. Patient and clinician opinions of patient reported outcome measures (PROMs) in the management of patients with rare diseases: a qualitative study. *Health and Quality of Life Outcomes,* 18**,** 177.

AIYEGBUSI, O. L., KYTE, D., COCKWELL, P., MARSHALL, T., DUTTON, M., WALMSLEY-ALLEN, N., AUTI, R. & CALVERT, M. 2018. Development and usability testing of an electronic patient-reported outcome measure (ePROM) system for patients with advanced chronic kidney disease. *Comput Biol Med,* 101**,** 120-127.

BAIJENS, L. W. J., WALSHE, M., AALTONEN, L.-M., ARENS, C., CORDIER, R., CRAS, P., CREVIER-BUCHMAN, L., CURTIS, C., GOLUSINSKI, W., GOVENDER, R., ERIKSEN, J. G., HANSEN, K., HEATHCOTE, K., HESS, M. M., HOSAL, S., KLUSSMANN, J. P., LEEMANS, C. R., MACCARTHY, D., MANDUCHI, B., MARIE, J.-P., NOURAEI, R., PARKES, C., PFLUG, C., PILZ, W., REGAN, J., ROMMEL, N., SCHINDLER, A., SCHOLS, A. M. W. J., SPEYER, R., SUCCO, G., WESSEL, I., WILLEMSEN, A. C. H., YILMAZ, T. & CLAVÉ, P. 2021. European white paper: oropharyngeal dysphagia in head and neck cancer. *European Archives of Oto-Rhino-Laryngology,* 278**,** 577-616.

BEITLER, J. J., CHEN, A. Y., JACOBSON, K., OWENS, A., EDWARDS, M. & JOHNSTONE, P. A. 2010. Health literacy and health care in an inner-city, total laryngectomy population. *Am J Otolaryngol,* 31**,** 29-31.

BENAVENT, M., SASTRE, J., ESCOBAR, I. G., SEGURA, A., CAPDEVILA, J., CARMONA, A., SEVILLA, I., ALONSO, T., CRESPO, G., GARCÍA, L., CANAL, N., DE LA CRUZ, G. & GALLEGO, J. 2021. Physician-perceived utility of the EORTC QLQ-GINET21 questionnaire in the treatment of patients with gastrointestinal neuroendocrine tumours: a multicentre, cross-sectional survey (QUALINETS). *Health and Quality of Life Outcomes,* 19**,** 38.

BIRKS, M. 2014. Qualitative Methodology: A Practical Guide. London: SAGE Publications, Inc.

BIRT, L., SCOTT, S., CAVERS, D., CAMPBELL, C. & WALTER, F. 2016. Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research,* 26**,** 1802-1811.

BISHOP, P. A. & HERRON, R. L. 2015. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int J Exerc Sci,* 8**,** 297-302.

BOWDEN, A., FOX-RUSHBY, J. A., NYANDIEKA, L. & WANJAU, J. 2002. Methods for pre-testing and piloting survey questions: illustrations from the KENQOL survey of health-related quality of life. *Health Policy and Planning,* 17**,** 322-330.

BOYCE, M. B., BROWNE, J. P. & GREENHALGH, J. 2014. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf,* 23**,** 508-18.

BOYNTON, P. M. 2004. Administering, analysing, and reporting your questionnaire. *Bmj,* 328**,** 1372-5.

BOZEC, A., SCHULTZ, P., GAL, J., CHAMOREY, E., CHATEAU, Y., DASSONVILLE, O., POISSONNET, G., SANTINI, J., PEYRADE, F., SAADA, E., GUIGAY, J., BENEZERY, K., LEYSALLE, A., SANTINI, L., GIOVANNI, A., MESSAOUDI, L. & FAKHRY, N. 2016. Evaluation of the information given to

patients undergoing head and neck cancer surgery using the EORTC QLQ-INFO25 questionnaire: A prospective multicentric study. *Eur J Cancer,* 67**,** 73-82.
BRASIL, V., OLIVEIRA, G. & MORAES, K. L. 2018. Psychometric properties of health related quality of life measures in acute coronary syndrome patients: a systematic review protocol. *JBI Database System Rev Implement Rep,* 16**,** 316-323.
BRAUN, V. & CLARKE, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3**,** 77-101.
BRAUN, V. & CLARKE, V. 2021. *Thematic Analysis: A Practical Guide*, SAGE Publications.
BRESLAU, J., JAVARAS, K. N., BLACKER, D., MURPHY, J. M. & NORMAND, S. L. 2008. Differential item functioning between ethnic groups in the epidemiological assessment of depression. *J Nerv Ment Dis,* 196**,** 297-306.
BRESSAN, V., BAGNASCO, A., ALEO, G., CATANIA, G., ZANINI, M. P., TIMMINS, F. & SASSO, L. 2017. The life experience of nutrition impact symptoms during treatment for head and neck cancer patients: a systematic review and meta-synthesis. *Support Care Cancer,* 25**,** 1699-1712.
BRIERLEY, J., GOSPODAROWICZ, M. & WITTEKINDT, C. 2017. *TNM classification of malignant tumours,* Chichester, Wiley Blackwell.
BROCKBANK, S., MILLER, N., OWEN, S. & PATTERSON, J. M. 2015. Pretreatment information on dysphagia: exploring the views of head and neck cancer patients. *J Pain Symptom Manage,* 49**,** 89-97.
BROTHERTON, A. M. & JUDD, P. A. 2007. Quality of life in adult enteral tube feeding patients. *Journal of Human Nutrition and Dietetics,* 20**,** 513-522.
BRYMAN, A. 2016. *Social Research Methods,* Oxford, Oxford University Press.
BUCHANAN, E. M. & SCOFIELD, J. E. 2018. Methods to detect low quality data and its implication for psychological research.
BURGES WATSON, D. L., LEWIS, S., BRYANT, V., PATTERSON, J., KELLY, C., EDWARDS-STUART, R., MURTAGH, M. J. & DEARY, V. 2018. Altered eating: a definition and framework for assessment and intervention. *BMC Nutrition,* 4**,** 14.
BYRNE, D. 2021. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality & Quantity*.
CANCER RESEARCH UK. 2018. *Head and neck cancer statistics* [Online]. Cancer Research UK. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers#heading-Zero [Accessed 01/03/19].
CAPPELLERI, J. C., JASON LUNDY, J. & HAYS, R. D. 2014. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther,* 36**,** 648-62.
CARLSON, M., WILCOX, R., CHOU, C.-P., CHANG, M., YANG, F., BLANCHARD, J., MARTERELLA, A., KUO, A. & CLARK, F. 2011. Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological assessment,* 23**,** 558-562.
CARLSSON, S., RYDÉN, A., RUDBERG, I., BOVE, M., BERGQUIST, H. & FINIZIA, C. 2012. Validation of the Swedish M. D. Anderson Dysphagia Inventory (MDADI) in Patients with Head and Neck Cancer and Neurologic Swallowing Disturbances. *Dysphagia,* 27**,** 361-369.
CARROZZINO, D., PATIERNO, C., GUIDI, J., BERROCAL MONTIEL, C., CAO, J., CHARLSON, M. E., CHRISTENSEN, K. S., CONCATO, J., DE LAS CUEVAS, C., DE LEON, J., EÖRY, A., FLECK, M. P., FURUKAWA, T. A., HORWITZ, R. I., NIERENBERG, A. A., RAFANELLI, C., WANG, H., WISE, T. N., SONINO, N. & FAVA, G. A. 2021. Clinimetric Criteria for Patient-Reported Outcome Measures. *Psychother Psychosom,* 90**,** 222-232.
CASTELLANO, A. & SHARMA, A. 2019. Systematic Review of Validated Quality of Life and Swallow Outcomes after Transoral Robotic Surgery. *Otolaryngol Head Neck Surg,* 161**,** 561-567.
CASTLEBERRY, A. & NOLEN, A. 2018. Thematic analysis of qualitative research data: Is it as easy as it sounds? *Curr Pharm Teach Learn,* 10**,** 807-815.

CHAN, K. M. K., CHAN, H. K. W., SIU, J. Y. L., PU, D., NUND, R. L. & WARD, E. C. 2019. Impact of Head and Neck Cancer Treatment on Survivors' Mealtime Experience. *Laryngoscope,* 129**,** 1572-1578.

CHANG, C.-H. & REEVE, B. B. 2005. Item Response Theory and its Applications to Patient-Reported Outcomes Measurement. *Evaluation & the Health Professions,* 28**,** 264-282.

CHANG, E. M., GILLESPIE, E. F. & SHAVERDIAN, N. 2019. Truthfulness in patient-reported outcomes: factors affecting patients' responses and impact on data quality. *Patient Relat Outcome Meas,* 10**,** 171-186.

CHARTERS, E., WU, R., MILROSS, C., BOGAARDT, H., FREEMAN-SANDERSON, A., BALLARD, K., DAVIES, S., OATES, J. & CLARK, J. 2021. Swallowing and communication outcomes following primary transoral robotic surgery. *Head Neck,* 43**,** 2013-2023.

CHEN, A. Y., FRANKOWSKI, R., BISHOP-LEONE, J., HEBERT, T., LEYK, S., LEWIN, J. & GOEPFERT, H. 2001. The development and validation of a dysphagia-specific quality-of-life questionnaire for patients with head and neck cancer: the M. D. Anderson dysphagia inventory. *Arch Otolaryngol Head Neck Surg,* 127**,** 870-6.

CHERA, B. S., EISBRUCH, A., MURPHY, B. A., RIDGE, J. A., GAVIN, P., REEVE, B. B., BRUNER, D. W. & MOVSAS, B. 2014. Recommended patient-reported core set of symptoms to measure in head and neck cancer treatment trials. *J Natl Cancer Inst,* 106.

CHOW, L., Q. M. 2020. Head and Neck Cancer. *The New England Journal of Medicine,* 382**,** 60-72.

CLOUGH, S., BURKE, M., DALY, B. & SCAMBLER, S. 2018. The impact of pre-radiotherapy dental extractions on head and neck cancer patients: a qualitative study. *British Dental Journal,* 225**,** 28-32.

CONNELL, J., CARLTON, J., GRUNDY, A., TAYLOR BUCK, E., KEETHARUTH, A. D., RICKETTS, T., BARKHAM, M., ROBOTHAM, D., ROSE, D. & BRAZIER, J. 2018. The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation,* 27**,** 1893-1902.

COOPER, G. & MEADOWS, R. A. L. 2015. Conceptualising Social Life. *In:* STONEMAN, P., GILBERT, N., STONEMAN, P. & GILBERT, N. (eds.) *Researching Social Life.*

CORDIER, R., JOOSTEN, A., CLAVÉ, P., SCHINDLER, A., BÜLOW, M., DEMIR, N., ARSLAN, S. S. & SPEYER, R. 2017. Evaluating the Psychometric Properties of the Eating Assessment Tool (EAT-10) Using Rasch Analysis. *Dysphagia,* 32**,** 250-260.

CORDIER, R., SPEYER, R., SCHINDLER, A., MICHOU, E., HEIJNEN, B. J., BAIJENS, L., KARADUMAN, A., SWAN, K., CLAVÉ, P. & JOOSTEN, A. V. 2018. Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia,* 33**,** 441-456.

CRARY, M. A., MANN, G. D. & GROHER, M. E. 2005. Initial psychometric assessment of a functional oral intake scale for dysphagia in stroke patients. *Arch Phys Med Rehabil,* 86**,** 1516-20.

CRESWELL, J. W. 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches,* Thousand Oaks, CA, SAGE Publications.

D'ANDRÉA, G., BORDENAVE, L., NGUYEN, F., TAO, Y., PALERI, V., TEMAM, S., MOYA-PLANA, A. & GORPHE, P. 2022. A prospective longitudinal study of quality of life in robotic-assisted salvage surgery for oropharyngeal cancer. *Eur J Surg Oncol,* 48**,** 1243-1250.

D'SOUZA, V., SERRAO, M., WATSON, E., BLOUIN, E., ZEITOUNI, A. & ALLISON, P. J. 2017. Information service in head and neck cancer care-a qualitative study. *Support Care Cancer*.

DAWSON, C., ADAMS, J. & FENLON, D. 2019. The experiences of people who receive swallow therapy after surgical treatment of head and neck cancer. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology,* 128**,** 456-463.

DE KLERK, S., BUCHANAN, H. & JEROSCH-HEROLD, C. 2018. The validity and clinical utility of the Disabilities of the Arm Shoulder and Hand questionnaire for hand injuries in developing country contexts: A systematic review. *J Hand Ther,* 31**,** 80-90.e1.

DEPARTMENT OF HEALTH 2005. Research governance framework for health and social care. Second ed.: Department of Health.

DUNCAN, E. A. S. & MURRAY, J. 2012. The barriers and facilitators to routine outcome measurement by allied health professionals in practice: a systematic review. *BMC Health Services Research,* 12**,** 96.

DUNLOP, E., FERGUSON, A., MUELLER, T., BAILLIE, K., CLARKE, J., LASKEY, J., KURDI, A., WU, O., JONES, R., GLEN, H. & BENNIE, M. 2022. What matters to patients and clinicians when discussing the impact of cancer medicines on health-related quality of life? Consensus-based mixed methods approach in prostate cancer. *Support Care Cancer,* 30**,** 3141-3150.

DWIVEDI, R. C., ST ROSE, S., ROE, J. W., KHAN, A. S., PEPPER, C., NUTTING, C. M., CLARKE, P. M., KERAWALA, C. J., RHYS-EVANS, P. H., HARRINGTON, K. J. & KAZI, R. 2010. Validation of the Sydney Swallow Questionnaire (SSQ) in a cohort of head and neck cancer patients. *Oral Oncol,* 46**,** e10-4.

EDWARDS, T. C., FREDERICKSEN, R. J., CRANE, H. M., CRANE, P. K., KITAHATA, M. M., MATHEWS, W. C., MAYER, K. H., MORALES, L. S., MUGAVERO, M. J., SOLORIO, R., YANG, F. M. & PATRICK, D. L. 2016. Content validity of Patient-Reported Outcomes Measurement Information System (PROMIS) items in the context of HIV clinical care. *Qual Life Res,* 25**,** 293-302.

EINARSSON, S., LAURELL, G. & TIBLOM EHRSSON, Y. 2019. Experiences and coping strategies related to food and eating up to two years after the termination of treatment in patients with head and neck cancer. *European Journal of Cancer Care,* 28**,** e12964.

EMBRETSON, S. E. & REISE, S. P. 2013. *Item response theory,* New York, Psychology Press.

FEILZER, Y. 2009. Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. *Journal of Mixed Methods Research,* 4**,** 6-16.

FERRANS, C. E. 2007. Differences in what quality-of-life instruments measure. *J Natl Cancer Inst Monogr***,** 22-6.

FETTERS, M. D., CURRY, L. A. & CRESWELL, J. W. 2013. Achieving Integration in Mixed Methods Designs—Principles and Practices. *Health Services Research,* 48**,** 2134-2156.

FRANCIS, D., WEYMULLER, E. A., JR., PARVATHANENI, U., MERATI, A. L. & YUEH, B. 2010. Dysphagia, stricture, and pneumonia in head and neck cancer patients: does treatment modality matter? *Ann Otol Rhinol Laryngol,* 119**,** 391-7.

FRANCIS, D. O., MCPHEETERS, M. L., NOUD, M., PENSON, D. F. & FEURER, I. D. 2016. Checklist to operationalize measurement characteristics of patient-reported outcome measures. *Systematic reviews,* 5**,** 129-129.

FRANZ, A., WORRELL, M. & VÖGELE, C. 2013. Integrating Mixed Method Data in Psychological Research: Combining Q Methodology and Questionnaires in a Study Investigating Cultural and Psychological Influences on Adolescent Sexual Behavior. *Journal of Mixed Methods Research,* 7**,** 370-389.

FROST, M. H., REEVE, B. B., LIEPA, A. M., STAUFFER, J. W. & HAYS, R. D. 2007. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health,* 10 Suppl 2**,** S94-s105.

GALANTINO, M. L., EDEN, M. M., SPINELLI, B. A. & FLORES, A. M. 2015. EDGE Task Force on Head and Neck Cancer Outcomes: A Systematic Review of Outcome Measures for Temporomandibular-related Dysfunction. *Rehabilitation Oncology,* 33**,** 6-14.

GANZER, H., TOUGER-DECKER, R., BYHAM-GRAY, L., MURPHY, B. A. & EPSTEIN, J. B. 2015. The eating experience after treatment for head and neck cancer: A review of the literature. *Oral Oncol,* 51**,** 634-42.

GIBBONS, E. 2016. Patient-reported outcome measures and the evaluation of services. *Health Services and delivery research,* 4**,** 55-68.

GILBERT, N. & STONEMAN, P. 2016. *Researching Social Life,* Los Angeles, CA, SAGE.

GOEPFERT, R. P., FULLER, C. D., GUNN, G. B., HANNA, E. Y., LEWIN, J. S., ZAVERI, J. S., HUBBARD, R. M., BARROW, M. P. & HUTCHESON, K. A. 2017. Symptom burden as a driver of decisional regret in long-term oropharyngeal carcinoma survivors. *Head Neck,* 39**,** 2151-2158.

GOSAK, M., GRADIŠAR, K., ROTOVNIK KOZJEK, N. & STROJAN, P. 2020. Psychological distress and nutritional status in head and neck cancer patients: a pilot study. *European Archives of Oto-Rhino-Laryngology,* 277**,** 1211-1217.

GREENHALGH, J. 2009. The applications of PROs in clinical practice: what are they, do they work, and why? *Qual Life Res,* 18**,** 115-23.

GREENHALGH, J., DALKIN, S., GOODING, K., GIBBONS, E., WRIGHT, J., MEADS, D., BLACK, N., VALDERAS, J. M. & PAWSON, R. 2017. Health Services and Delivery Research. *Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care.* Southampton (UK): NIHR Journals Library.

GREENHALGH, T. 2014. *How to read a paper: the basics of evidence-based medicine,* Chichester, Wiley.

GUEDES, R. L. V., ANGELIS, E. C.-D., CHEN, A. Y., KOWALSKI, L. P. & VARTANIAN, J. G. 2013. Validation and Application of the M.D. Anderson Dysphagia Inventory in Patients Treated for Head and Neck Cancer in Brazil. *Dysphagia,* 28**,** 24-32.

GUETTERMAN, T. C., SAKAKIBARA, R. V., PLANO CLARK, V. L., LUBORSKY, M., MURRAY, S. M., CASTRO, F. G., CRESWELL, J. W., DEUTSCH, C. & GALLO, J. J. 2019. Mixed methods grant applications in the health sciences: An analysis of reviewer comments. *PLoS One,* 14**,** e0225308.

HAGELL, P., REIMER, J. & NYBERG, P. 2009. Whose quality of life? Ethical implications in patient-reported health outcome measurement. *Value Health,* 12**,** 613-7.

HAJDÚ, S. F., PLASCHKE, C. C., JOHANSEN, C., DALTON, S. O. & WESSEL, I. 2017. Cross-Cultural Translation, Adaptation and Reliability of the Danish M. D. Andeson Dysphagia Inventory (MDADI) in Patients with Head and Neck Cancer. *Dysphagia,* 32**,** 472-479.

HIGGINSON, I. J. & CARR, A. J. 2001. Using quality of life measures in the clinical setting. *BMJ,* 322**,** 1297.

HONG, M. K. H., YAO, H. H. I., PEDERSEN, J. S., PETERS, J. S., COSTELLO, A. J., MURPHY, D. G., HOVENS, C. M. & CORCORAN, N. M. 2013. Error rates in a clinical data repository: lessons from the transition to electronic data transfer--a descriptive study. *BMJ open,* 3**,** e002406.

HOWELL, D., MOLLOY, S., WILKINSON, K., GREEN, E., ORCHARD, K., WANG, K. & LIBERTY, J. 2015. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Ann Oncol,* 26**,** 1846-1858.

HRA 2020. UK Policy Framework for Health and Social Care Research. *In:* AUTHORITY, H. R. (ed.).

HSIEH, H.-F. & SHANNON, S. 2005. Three Approaches to Qualitative Content Analysis. *Qualitative health research,* 15**,** 1277-88.

HUNTER, K. U., SCHIPPER, M., FENG, F. Y., LYDEN, T., HAXER, M., MURDOCH-KINCH, C. A., CORNWALL, B., LEE, C. S., CHEPEHA, D. B. & EISBRUCH, A. 2013. Toxicities affecting quality of life after chemo-IMRT of oropharyngeal cancer: prospective study of patient-reported, observer-rated, and objective outcomes. *Int J Radiat Oncol Biol Phys,* 85**,** 935-40.

HUTCHESON, K. A., BARROW, M. P., BARRINGER, D. A., KNOTT, J. K., LIN, H. Y., WEBER, R. S., FULLER, C. D., LAI, S. Y., ALVAREZ, C. P., RAUT, J., LAZARUS, C. L., MAY, A., PATTERSON, J., ROE, J. W., STARMER, H. M. & LEWIN, J. S. 2017. Dynamic Imaging Grade of Swallowing Toxicity (DIGEST): Scale development and validation. *Cancer,* 123**,** 62-70.

HUTCHESON, K. A., BARROW, M. P., LISEC, A., BARRINGER, D. A., GRIES, K. & LEWIN, J. S. 2016. What is a clinically relevant difference in MDADI scores between groups of head and neck cancer patients? *Laryngoscope,* 126**,** 1108-13.

HUTCHESON, K. A., LEWIN, J. S., BARRINGER, D. A., LISEC, A., GUNN, G. B., MOORE, M. W. & HOLSINGER, F. C. 2012. Late dysphagia after radiotherapy-based treatment of head and neck cancer. *Cancer,* 118**,** 5793-9.

HUTCHESON, K. A., NURGALIEVA, Z., ZHAO, H., GUNN, G. B., GIORDANO, S. H., BHAYANI, M. K., LEWIN, J. S. & LEWIS, C. M. 2018. Two-year prevalence of dysphagia and related outcomes in head and neck cancer survivors: An updated SEER-Medicare analysis. *Head Neck*.

INVOLVE 2012. Briefing notes for researchers:public involvement in NHS, public health and social care research. National Institute for Health Research.

IRVINE, F. E., CLARK, M. T., EFSTATHIOU, N., HERBER, O. R., HOWROYD, F., GRATRIX, L., SAMMUT, D., TRUMM, A., HANSSEN, T. A., TAYLOR, J. & BRADBURY-JONES, C. 2020. The state of mixed methods research in nursing: A focused mapping review and synthesis. *Journal of Advanced Nursing,* 76**,** 2798-2809.

JABBOUR, J., MILROSS, C., SUNDARESAN, P., EBRAHIMI, A., SHEPHERD, H. L., DHILLON, H. M., MORGAN, G., ASHFORD, B., ABDUL-RAZAK, M., WONG, E., VENESS, M., PALME, C. E., FROGGATT, C., COHEN, R., EKMEJIAN, R., TAY, J., ROSHAN, D. & CLARK, J. R. 2017. Education and support needs in patients with head and neck cancer: A multi-institutional survey. *Cancer,* 123**,** 1949-1957.

JACOBSON, B. H., JOHNSON, A., GRYWALSKI, C., SILBERGLEIT, A., JACOBSON, G., M., B. & C., N. 1997. The Voice Handicap Index (VHI). *American Journal of Speech-Language Pathology,* 6**,** 66-70.

JARVIS, C. I., VAN ZANDVOORT, K., GIMMA, A., PREM, K., KLEPAC, P., RUBIN, G. J. & EDMUNDS, W. J. 2020. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC medicine,* 18**,** 1-10.

JEAN-PIERRE, P., CHENG, Y., PASKETT, E., SHAO, C., FISCELLA, K., WINTERS, P. & PATIENT NAVIGATION RESEARCH, P. 2014. Item response theory analysis of the patient satisfaction with cancer-related care measure: a psychometric investigation in a multicultural sample of 1,296 participants. *Supportive care in cancer : official journal of the Multinational Association of Supportive Care in Cancer,* 22**,** 2229-2240.

JONES, F. J. S., EZZEDDINE, F. L., HERMAN, S. T., BUCHHALTER, J., FUREMAN, B. & MOURA, L. 2020. A feasibility assessment of functioning and quality-of-life patient-reported outcome measures in adult epilepsy clinics: A systematic review. *Epilepsy Behav,* 102**,** 106704.

JONES, R. N. 2019. Differential item functioning and its relevance to epidemiology. *Curr Epidemiol Rep,* 6**,** 174-183.

JONES, T. L., BAXTER, M. A. & KHANDUJA, V. 2013. A quick guide to survey research. *Ann R Coll Surg Engl,* 95**,** 5-7.

KALKAN, Ö. K., KARA, Y. & KELECIOĞLU, H. 2018. Evaluating Performance of Missing Data Imputation Methods in IRT Analyses. *International Journal of Assessment Tools in Education*.

KEAN, J., BRODKE, D. S., BIBER, J. & GROSS, P. 2018. An introduction to Item Response Theory and Rasch Analysis of the Eating Assessment Tool (EAT-10). *Brain Impair,* 19**,** 91-102.

KELLEY, K., CLARK, B., BROWN, V. & SITZIA, J. 2003. Good practice in the conduct and reporting of survey research. *Int J Qual Health Care,* 15**,** 261-6.

KENDALL, K. A., TANNER, K. & KOSEK, S. R. 2014. Timing of events during deglutition after chemoradiation therapy for oropharyngeal carcinoma. *Head Neck,* 37**,** 1193-9.

KHAN, M. K., PATTERSON, J., OWEN, S., REES, S., GAMBERINI, L., PALERI, V. & NORTH OF ENGLAND CANCER NETWORK AUDIT, G. 2015. Comparing the Performance Status Scale and MD Anderson Dysphagia Inventory as swallowing outcome measures in head and neck cancer: a prospective cohort study. *Clinical Otolaryngology,* 40**,** 321-326.

KIRSH, E., NAUNHEIM, M., HOLMAN, A., KAMMER, R., VARVARES, M. & GOLDSMITH, T. 2019. Patient-reported versus physiologic swallowing outcomes in patients with head and neck cancer after chemoradiation. *Laryngoscope,* 129**,** 2059-2064.

KLUSSMANN, J. P. 2017. Head and Neck Cancer - New Insights into a Heterogeneous Disease. *Oncology Research and Treatment,* 40**,** 318-319.

KREBBERS, I., PILZ, W., VANBELLE, S., VERDONSCHOT, R. J. C. G. & BAIJENS, L. W. J. 2022. Affective Symptoms and Oropharyngeal Dysphagia in Head-and-Neck Cancer Patients: A Systematic Review. *Dysphagia,* 38**,** 127-144.

KRISCIUNAS, G. P., SOKOLOFF, W., STEPAS, K. & LANGMORE, S. E. 2012. Survey of usual practice: dysphagia therapy in head and neck cancer patients. *Dysphagia,* 27**,** 538-49.

KROENKE, K., MONAHAN, P. O. & KEAN, J. 2015. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol,* 68**,** 1085-92.

KWON, C. H., KIM, Y. H., PARK, J. H., OH, B. M. & HAN, T. R. 2013. Validity and reliability of the korean version of the MD anderson Dysphagia inventory for head and neck cancer patients. *Ann Rehabil Med,* 37**,** 479-87.

LAM, K. C., MARSHALL, A. N. & SNYDER VALIER, A. R. 2020. Patient-Reported Outcome Measures in Sports Medicine: A Concise Resource for Clinicians and Researchers. *Journal of Athletic Training,* 55**,** 390-408.

LANGERMAN, A., MACCRACKEN, E., KASZA, K., HARAF, D. J., VOKES, E. E. & STENSON, K. M. 2007. Aspiration in Chemoradiated Patients With Head and Neck Cancer. *Archives of Otolaryngology–Head & Neck Surgery,* 133**,** 1289-1295.

LAWLER, M., LE CHEVALIER, T., MURPHY, M. J., JR., BANKS, I., CONTE, P., DE LORENZO, F., MEUNIER, F., PINEDO, H. M., SELBY, P., ARMAND, J. P., BARBACID, M., BARZACH, M., BERGH, J., BODE, G., CAMERON, D. A., DE BRAUD, F., DE GRAMONT, A., DIEHL, V., DILER, S., ERDEM, S., FITZPATRICK, J. M., GEISSLER, J., HOLLYWOOD, D., HOJGAARD, L., HORGAN, D., JASSEM, J., JOHNSON, P. W., KAPITEIN, P., KELLY, J., KLOEZEN, S., LA VECCHIA, C., LOWENBERG, B., OLIVER, K., SULLIVAN, R., TABERNERO, J., VAN DE VELDE, C. J., WILKING, N., WILSON, R., ZIELINSKI, C., ZUR HAUSEN, H. & JOHNSTON, P. G. 2014. A catalyst for change: the European cancer Patient's Bill of Rights. *Oncologist,* 19**,** 217-24.

LECHIEN, J. R., CAVELIER, G., THILL, M.-P., BOUSARD, L., BLECIC, S., VANDERWEGEN, J., SAUSSEZ, S., RODRIGUEZ, A. & DEQUANTER, D. 2020. Validity and reliability of a French version of M.D. Anderson Dysphagia Inventory. *European Archives of Oto-Rhino-Laryngology,* 277**,** 3111-3119.

LEUNG, L. 2015. Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care,* 4**,** 324-327.

LIN, D. J., ALTAMIMI, J., PEARCE, K., WILSON, J. A. & PATTERSON, J. M. 2022. Psychometric Properties of the MDADI—A Preliminary Study of Whether Less is Truly More? *Dysphagia,* 37**,** 323-332.

LIST, M. A., D'ANTONIO, L. L., CELLA, D. F., SISTON, A., MUMBY, P., HARAF, D. & VOKES, E. 1996. The Performance Status Scale for Head and Neck Cancer Patients and the Functional Assessment of Cancer Therapy-Head and Neck Scale. A study of utility and validity. *Cancer,* 77**,** 2294-301.

LIST, M. A., RITTER-STERR, C. & LANSKY, S. B. 1990. A performance status scale for head and neck cancer patients. *Cancer,* 66**,** 564-9.

LOGEMANN, J. 1998. *Evaluation and Treatment of Swallowing Disorders,* Austin, PRO-ED.

LORENTE, S., VILADRICH, C., VIVES, J. & LOSILLA, J.-M. 2020. Tools to assess the measurement properties of quality of life instruments: a meta-review. *BMJ Open,* 10**,** e036038.

MAGASI, S., RYAN, G., REVICKI, D., LENDERKING, W., HAYS, R. D., BROD, M., SNYDER, C., BOERS, M. & CELLA, D. 2012. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. *Qual Life Res,* 21**,** 739-46.

MALTERUD, K. 2001. Qualitative research: standards, challenges, and guidelines. *Lancet,* 358**,** 483-8.

MANDUCHI, B., CHE, Z., FITCH, M. I., RINGASH, J., HOWELL, D. & MARTINO, R. 2022. Psychometric properties of patient-reported outcome measures for dysphagia in head and neck cancer: a systematic review protocol using COSMIN methodology. *Syst Rev,* 11**,** 27.

MANIKANTAN, K., KHODE, S., SAYED, S. I., ROE, J., NUTTING, C. M., RHYS-EVANS, P., HARRINGTON, K. J. & KAZI, R. 2009. Dysphagia in head and neck cancer. *Cancer Treat Rev,* 35**,** 724-32.

MARTINO, R., FITCH, M. I., FULLER, C. D., HOPE, A., KRISCIUNAS, G., LANGMORE, S. E., LAZARUS, C., MACDONALD, C. L., MCCULLOCH, T., MILLS, G., PALMA, D. A., PYTYNIA, K., RINGASH, J., SULTANEM, K., THEURER, J., THORPE, K. E. & HUTCHESON, K. 2021. The PRO-ACTIVE trial protocol: a randomized study comparing the effectiveness of PROphylACTic swallow InterVEntion for patients receiving radiotherapy for head and neck cancer. *BMC Cancer,* 21**,** 1100.

MATSUDA, Y., KANAZAWA, M., KOMAGAMINE, Y., YAMASHIRO, M., AKIFUSA, S. & MINAKUCHI, S. 2018. Reliability and Validity of the MD Anderson Dysphagia Inventory Among Japanese Patients. *Dysphagia,* 33**,** 123-132.

MAYS, J. A. & MATHIAS, P. C. 2019. Measuring the rate of manual transcription error in outpatient point-of-care testing. *Journal of the American Medical Informatics Association,* 26**,** 269-272.

MCGOWAN, J., STRAUS, S., MOHER, D., LANGLOIS, E. V., O'BRIEN, K. K., HORSLEY, T., ALDCROFT, A., ZARIN, W., GARITTY, C. M., HEMPEL, S., LILLIE, E., TUNÇALP, Ö. & TRICCO, A. C. 2020. Reporting scoping reviews—PRISMA ScR extension. *Journal of clinical epidemiology,* 123**,** 177-179.

MCHORNEY, C. A., BRICKER, D. E., KRAMER, A. E., ROSENBEK, J. C., ROBBINS, J., CHIGNELL, K. A., LOGEMANN, J. A. & CLARKE, C. 2000. The SWAL-QOL outcomes tool for oropharyngeal dysphagia in adults: I. Conceptual foundation and item development. *Dysphagia,* 15**,** 115-21.

MCQUESTION, M., FITCH, M. & HOWELL, D. 2011. The changed meaning of food: Physical, social and emotional loss for patients having received radiation treatment for head and neck cancer. *Eur J Oncol Nurs,* 15**,** 145-51.

MEHANNA, H., ROBINSON, M., HARTLEY, A., KONG, A., FORAN, B., FULTON-LIEUW, T., DALBY, M., MISTRY, P., SEN, M., O'TOOLE, L., AL BOOZ, H., DYKER, K., MOLERON, R., WHITAKER, S., BRENNAN, S., COOK, A., GRIFFIN, M., AYNSLEY, E., ROLLES, M., DE WINTON, E., CHAN, A., SRINIVASAN, D., NIXON, I., GRUMETT, J., LEEMANS, C. R., BUTER, J., HENDERSON, J., HARRINGTON, K., MCCONKEY, C., GRAY, A. & DUNN, J. 2019. Radiotherapy plus cisplatin or cetuximab in low-risk human papillomavirus-positive oropharyngeal cancer (De-ESCALaTE HPV): an open-label randomised controlled phase 3 trial. *Lancet,* 393**,** 51-60.

MEHANNA, H. M. & MORTON, R. P. 2006. Patients' views on the utility of quality of life questionnaires in head and neck cancer: a randomised trial. *Clinical Otolaryngology,* 31**,** 310-316.

MEHANNA, H. M., SEN, M., CHESTER, J. D., SANGHERA, P., PALERI, V., GAUNT, P., BABRAH, J., HARTLEY, A. G. J., KONG, A., AL-BOOZ, H., FORAN, B., MILES, E., ROBINSON, M., JEPSON, M., DONNOVAN, J., YAP, C., FIRTH, C., GAUNT, C., BOWDEN, S. J. & FORSTER, M. D. 2017. Phase III randomised controlled trial (RCT) comparing alternative regimens for escalating treatment of intermediate and high-risk oropharyngeal cancer (CompARE). *Journal of Clinical Oncology,* 35**,** TPS6091-TPS6091.

MENDEZ, A., SEIKALY, H., EURICH, D., DZIOBA, A., AALTO, D., OSSWALD, M., HARRIS, J. R., O'CONNELL, D. A., LAZARUS, C., URKEN, M., LIKHTEROV, I., CHAI, R. L., RAUSCHER, E., BUCHBINDER, D., OKAY, D., HAPPONEN, R. P., KINNUNEN, I., IRJALA, H., SOUKKA, T. & LAINE, J. 2020. Development of a Patient-Centered Functional Outcomes Questionnaire in Head and Neck Cancer. *JAMA Otolaryngol Head Neck Surg,* 146**,** 437-443.

MERCER, S. W., MAXWELL, M., HEANEY, D. & WATT, G. C. M. 2004. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Family Practice,* 21**,** 699-705.

MOHER, D., LIBERATI, A., TETZLAFF, J., ALTMAN, D. G. & THE, P. G. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine,* 6**,** e1000097.

MOKKINK, L. B., PRINSEN, C. A. C., PATRICK, D. L., ALONSO, J., BOUTER, L. M., DE VET, H. C. & TERWEE, C. B. 2019. COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam: Amsterdam Public Health Research Institute.

MOKKINK, L. B., PRINSEN, C. A. C., PATRICK, D. L., ALONSO, J., BOUTER, L. M., VET, H. C. W. D. & TERWEE, C. B. 2018. COSMIN methodology for systematic reviews of Patient - Reported Outcome Measures (PROMs) user manual. Amsterdam: COSMIN Initiative.

MOKKINK, L. B., TERWEE, C. B., KNOL, D. L., STRATFORD, P. W., ALONSO, J., PATRICK, D. L., BOUTER, L. M. & DE VET, H. C. 2010a. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC medical research methodology,* 10**,** 22-22.

MOKKINK, L. B., TERWEE, C. B., PATRICK, D. L., ALONSO, J., STRATFORD, P. W., KNOL, D. L., BOUTER, L. M. & DE VET, H. C. 2010b. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res,* 19**,** 539-49.

MOKKINK, L. B., TERWEE, C. B., PATRICK, D. L., ALONSO, J., STRATFORD, P. W., KNOL, D. L., BOUTER, L. M. & DE VET, H. C. 2010c. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol,* 63**,** 737-45.

MOLONEY, J. & WALSHE, M. 2019. Managing and supporting quality-of-life issues in dysphagia: A survey of clinical practice patterns and perspectives in the UK, Ireland and South Africa. *Int J Lang Commun Disord,* 54**,** 41-49.

MONTES-JOVELLAR, L., CARRILLO, A., MURIEL, A., BARBERA, R., SANCHEZ, F. & COBETA, I. 2019. Translation and validation of the MD Anderson Dysphagia Inventory (MDADI) for Spanish-speaking patients. *Head & Neck,* 41**,** 122-129.

MONTGOMERY, N., HOWELL, D., ISMAIL, Z., BARTLETT, S. J., BRUNDAGE, M., BRYANT-LUKOSIUS, D., KRZYZANOWSKA, M., MOODY, L., SNYDER, C., BARBERA, L., BARBERA, L., ISMAIL, Z., HOWELL, D., BRUNDAGE, M., SNYDER, C., KRZYZANOWSKA, M., BRYANT-LUKOSIUS, D., POTTIE, P., DUPRAS, L., MOODY, L., BARTLETT, S. J., STALEY, M., MARTELLI, L., MONTGOMERY, N. & THE CANCER CARE ONTARIO PATIENT REPORTED OUTCOME ADVISORY, C. 2020. Selecting, implementing and evaluating patient-reported outcome measures for routine clinical use in cancer: the Cancer Care Ontario approach. *Journal of Patient-Reported Outcomes,* 4**,** 101.

MORSE, J. M. 2015. Critical Analysis of Strategies for Determining Rigor in Qualitative Inquiry. *Qualitative Health Research,* 25**,** 1212-1222.

MORTENSEN, H. R., JENSEN, K. & GRAU, C. 2013. Aspiration pneumonia in patients treated with radiotherapy for head and neck cancer. *Acta Oncol,* 52**,** 270-6.

MOSEHOLM, E. & FETTERS, M. D. 2017. Conceptual models to guide integration during analysis in convergent mixed methods studies. *Methodological Innovations,* 10**,** 2059799117703118.

NGUYEN, H., BUTOW, P., DHILLON, H. & SUNDARESAN, P. 2021. A review of the barriers to using Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs) in routine cancer care. *Journal of Medical Radiation Sciences,* 68**,** 186-195.

NGUYEN, T. H., HAN, H.-R., KIM, M. T. & CHAN, K. S. 2014. An introduction to item response theory for patient-reported outcome measurement. *The patient,* 7**,** 23-35.

NIC GIOLLA EASPAIG, B., TRAN, Y., BIERBAUM, M., ARNOLDA, G., DELANEY, G. P., LIAUW, W., WARD, R. L., OLVER, I., CURROW, D., GIRGIS, A., DURCINOSKA, I. & BRAITHWAITE, J. 2020. What are the attitudes of health professionals regarding patient reported outcome measures (PROMs) in oncology practice? A mixed-method synthesis of the qualitative evidence. *BMC Health Services Research,* 20**,** 102.

NICE 2004. Improving Outcomes in Head and Neck Cancers. London: National Institute for Clinical Excellence.

NICE 2018. Research Governance Policy. London: National Institute for Clinical Excellence.

NIHR 2018. National Standards for Public Involvement in Research. National Institute for Health Research.

NIMA, A. A., CLONINGER, K. M., PERSSON, B. N., SIKSTRÖM, S. & GARCIA, D. 2020. Validation of Subjective Well-Being Measures Using Item Response Theory. *Frontiers in Psychology,* 10.

NUND, R., WARD, E. C., SCARINCI, N. A., CARTMILL, B., KUIPERS, P. & PORCEDDU, S. V. 2014a. The lived experience of dysphagia following non-surgical treatment for head and neck cancer. *Int J Speech Lang Pathol,* 16**,** 282-9.

NUND, R. L., SCARINCI, N. A., CARTMILL, B., WARD, E. C., KUIPERS, P. & PORCEDDU, S. V. 2014b. Application of the International Classification of Functioning, Disability and Health (ICF) to People with Dysphagia Following Non-surgical Head and Neck Cancer Management. *Dysphagia,* 29**,** 692-703.

O'CONNOR, B. P. 2018. An illustration of the effects of fluctuations in test information on measurement error, the attenuation of effect sizes, and diagnostic reliability. *Psychol Assess,* 30**,** 991-1003.

O'REILLY, G. M., GABBE, B., MOORE, L. & CAMERON, P. A. 2016. Classifying, measuring and improving the quality of data in trauma registries: A review of the literature. *Injury,* 47**,** 559-567.

OJO, B., GENDEN, E. M., TENG, M. S., MILBURY, K., MISIUKIEWICZ, K. J. & BADR, H. 2012. A systematic review of head and neck cancer quality of life assessment instruments. *Oral Oncol,* 48**,** 923-937.

OLINO, T. M., YU, L., KLEIN, D. N., ROHDE, P., SEELEY, J. R., PILKONIS, P. A. & LEWINSOHN, P. M. 2012. Measuring depression using item response theory: an examination of three measures of depressive symptomatology. *Int J Methods Psychiatr Res,* 21**,** 76-85.

OSBORN, H. A., GOLDSMITH, T. A. & VARVARES, M. A. 2019. Assessing functional outcomes in head and neck surgical oncology. *Head Neck,* 41**,** 2051-2057.

OTTOSSON, S., LAURELL, G. & OLSSON, C. 2013. The experience of food, eating and meals following radiotherapy for head and neck cancer: a qualitative study. *J Clin Nurs,* 22**,** 1034-43.

OWADALLY, W., HURT, C., TIMMINS, H., PARSONS, E., TOWNSEND, S., PATTERSON, J., HUTCHESON, K., POWELL, N., BEASLEY, M., PALANIAPPAN, N., ROBINSON, M., JONES, T. M. & EVANS, M. 2015. PATHOS: a phase II/III trial of risk-stratified, reduced intensity adjuvant treatment in patients undergoing transoral surgery for Human papillomavirus (HPV) positive oropharyngeal cancer. *BMC Cancer,* 15**,** 602.

PATEL, D. A., SHARDA, R., HOVIS, K. L., NICHOLS, E. E., SATHE, N., PENSON, D. F., FEURER, I. D., MCPHEETERS, M. L., VAEZI, M. F. & FRANCIS, D. O. 2017. Patient-reported outcome measures in dysphagia: a systematic review of instrument development and validation. *Dis Esophagus,* 30**,** 1-23.

PATRICK, D. L. 2019. Many ways to skin a cat: psychometric methods options illustrated. *J Patient Rep Outcomes,* 3**,** 48.

PATRICK, D. L., BURKE, L. B., GWALTNEY, C. J., LEIDY, N. K., MARTIN, M. L., MOLSEN, E. & RING, L. 2011. Content validity--establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2--assessing respondent understanding. *Value Health,* 14**,** 978-88.

PATTERSON, J. M., HILDRETH, A., MCCOLL, E., CARDING, P. N., HAMILTON, D. & WILSON, J. A. 2011. The clinical application of the 100mL water swallow test in head and neck cancer. *Oral Oncol,* 47**,** 180-4.

PATTERSON, J. M., MCCOLL, E., WILSON, J., CARDING, P. & RAPLEY, T. 2015. Head and neck cancer patients' perceptions of swallowing following chemoradiotherapy. *Support Care Cancer,* 23**,** 3531-8.

PATTERSON, J. M., RAPLEY, T., CARDING, P. N., WILSON, J. A. & MCCOLL, E. 2013. Head and neck cancer and dysphagia; caring for carers. *Psychooncology,* 22**,** 1815-20.

PAULSEN, A., OVERGAARD, S. & LAURITSEN, J. M. 2012. Quality of data entry using single entry, double entry and automated forms processing--an example based on a study of patient-reported outcomes. *PLoS One,* 7**,** e35087.

PEDERSEN, A., WILSON, J., MCCOLL, E., CARDING, P. & PATTERSON, J. 2016. Swallowing outcome measures in head and neck cancer--How do they compare? *Oral Oncol,* 52**,** 104-8.

PENG, J., CHEN, Y., SHEN, L., ZHU, Z., XING, W., JIN, G. & HU, Y. 2020. Psychometric properties of patient-reported outcome measures of self-management for cancer survivors: a systematic review protocol using COSMIN methodology. *BMJ Open,* 10**,** e038983.

PETKAR, I., ROONEY, K., ROE, J. W., PATTERSON, J. M., BERNSTEIN, D., TYLER, J. M., EMSON, M. A., MORDEN, J. P., MERTENS, K., MILES, E., BEASLEY, M., ROQUES, T., BHIDE, S. A., NEWBOLD, K. L., HARRINGTON, K. J., HALL, E. & NUTTING, C. M. 2016. DARS: a phase III randomised multicentre study of dysphagia- optimised intensity- modulated radiotherapy (Do-IMRT) versus standard intensity- modulated radiotherapy (S-IMRT) in head and neck cancer. *BMC Cancer,* 16**,** 770.

PHILLIPS, J. L. & CURROW, D. C. 2010. Cancer as a chronic disease. *Collegian,* 17**,** 47-50.

PRINSEN, C. A. C., MOKKINK, L. B., BOUTER, L. M., ALONSO, J., PATRICK, D. L., DE VET, H. C. W. & TERWEE, C. B. 2018. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research,* 27**,** 1147-1157.

PUBLIC HEALTH SCOTLAND. 2020. *Cancer of the Head and Neck statistics* [Online]. Public Health Scotland. Available: https://www.isdscotland.org/health-topics/cancer/cancer-statistics/head-and-neck/#head [Accessed 05/11/22].

RCSLT 2006. *Communicating Quality 3,* London, Royal College of Speech & Language Therapists.

REEVE, B. B., WYRWICH, K. W., WU, A. W., VELIKOVA, G., TERWEE, C. B., SNYDER, C. F., SCHWARTZ, C., REVICKI, D. A., MOINPOUR, C. M., MCLEOD, L. D., LYONS, J. C., LENDERKING, W. R., HINDS, P. S., HAYS, R. D., GREENHALGH, J., GERSHON, R., FEENY, D., FAYERS, P. M., CELLA, D., BRUNDAGE, M., AHMED, S., AARONSON, N. K. & BUTT, Z. 2013. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res,* 22**,** 1889-905.

REISE, S. P., AINSWORTH, A. T. & HAVILAND, M. G. 2005. Item Response Theory: Fundamentals, Applications, and Promise in Psychological Research. *Current Directions in Psychological Science,* 14**,** 95-101.

RICKARDS, G., MAGEE, C. & ARTINO JR, A. R. 2012. You can't fix by analysis what you've spoiled by design: developing survey instruments and collecting validity evidence. *Journal of graduate medical education,* 4**,** 407-410.

RIEGER, J. M., TANG, J. A., HARRIS, J., SEIKALY, H., WOLFAARDT, J., GLAUM, R., SCHMELZEISEN, R., BUCHBINDER, D., JACOBSON, A., LAZARUS, C., MARKOWITZ, E., OKAY, D., URKEN, M., AITASALO, K., HAPPONEN, R. P., KINNUNEN, I., LAINE, J. & SOUKKA, T. 2010. Survey of current functional outcomes assessment practices in patients with head and neck cancer: initial project of the head and neck research network. *J Otolaryngol Head Neck Surg,* 39**,** 523-31.

RITCHIE, J., LEWIS, J., MCNAUGHTON NICHOLLS, C. & ORMSTON, R. 2014. *Qualitative Research Practice: a guide for social science students and researchers,* London, SAGE Publications.

ROE, J. W., CARDING, P. N., RHYS-EVANS, P. H., NEWBOLD, K. L., HARRINGTON, K. J. & NUTTING, C. M. 2012. Assessment and management of dysphagia in patients with head and neck cancer who receive radiotherapy in the United Kingdom - a web-based survey. *Oral Oncol,* 48**,** 343-8.

ROGERS, S. N. & BARBER, B. 2017. Using PROMs to guide patients and practitioners through the head and neck cancer journey. *Patient Relat Outcome Meas,* 8**,** 133-142.

ROGERS, S. N., HOGG, E. S., CHEUNG, W. K., LAI, L. K., JASSAL, P. & LOWE, D. 2015a. The use of health related quality of life data to produce information sheets for patients with head and neck cancer. *Ann R Coll Surg Engl,* 97**,** 359-63.

ROGERS, S. N., HOGG, E. S., CHEUNG, W. K., LAI, L. K., JASSAL, P., LOWE, D. & KANATAS, A. 2015b. 'What will I be like' after my diagnosis of head and neck cancer? *Eur Arch Otorhinolaryngol,* 272**,** 2463-72.

ROGERS, S. N., SEMPLE, C., BABB, M. & HUMPHRIS, G. 2016. Quality of life considerations in head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *The Journal of Laryngology and Otology,* 130**,** S49-S52.

ROGUS-PULIA, N. M., PIERCE, M. C., MITTAL, B. B., ZECKER, S. G. & LOGEMANN, J. A. 2014. Changes in swallowing physiology and patient perception of swallowing function following chemoradiation for head and neck cancer. *Dysphagia,* 29**,** 223-33.

ROSEN, C. A., LEE, A. S., OSBORNE, J., ZULLO, T. & MURRY, T. 2004. Development and validation of the voice handicap index-10. *Laryngoscope,* 114**,** 1549-56.

ROSENBEK, J. C., ROBBINS, J. A., ROECKER, E. B., COYLE, J. L. & WOOD, J. L. 1996. A penetration-aspiration scale. *Dysphagia,* 11**,** 93-8.

ROSENKOETTER, U. & TATE, R. L. 2018. Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. *Brain Impairment,* 19**,** 103-118.

RUSSI, E. G., CORVO, R., MERLOTTI, A., ALTERIO, D., FRANCO, P., PERGOLIZZI, S., DE SANCTIS, V., RUO REDDA, M. G., RICARDI, U., PAIAR, F., BONOMO, P., MERLANO, M. C., ZURLO, V., CHIESA, F., SANGUINETI, G. & BERNIER, J. 2012. Swallowing dysfunction in head and neck cancer patients treated by radiotherapy: review and recommendations of the supportive task group of the Italian Association of Radiation Oncology. *Cancer Treat Rev,* 38**,** 1033-49.

SALKIND, N. 2010. Encyclopedia of Research Design. Thousand Oaks, California: SAGE Publications, Inc.

SAMEJIMA, F. 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometrika,* 34**,** 1-97.

SCHACHE, A., KERAWALA, C., AHMED, O., BRENNAN, P. A., COOK, F., GARRETT, M., HOMER, J., HUGHES, C., MAYLAND, C., MIHAI, R., NEWBOLD, K., O'HARA, J., ROE, J., SIBTAIN, A., SMITH, M., THAVARAJ, S., WELLER, A., WINTER, L., YOUNG, V. & WINTER, S. C. 2021. British Association of Head and Neck Oncologists (BAHNO) standards 2020. *Journal of Oral Pathology & Medicine,* 50**,** 262-273.

SCOTTISH GOVERNMENT. 2020. *Scottish Index of Multiple Deprivation 2020* [Online]. Scottish Government. Available: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/ [Accessed 01/04/2022].

SEKELY, A., TAYLOR, G. J. & BAGBY, R. M. 2018. Developing a short version of the Toronto Structured Interview for Alexithymia using item response theory. *Psychiatry Res,* 266**,** 218-227.

SEXTON-RADEK, K. & SIMMONS, L. 2018. The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation. Thousand Oaks, California: SAGE Publications, Inc.

SILVEIRA, A., GONÇALVES, J., SEQUEIRA, T., RIBEIRO, C., LOPES, C., MONTEIRO, E. & PIMENTEL, F. L. 2011. Geriatric oncology: comparing health related quality of life in head and neck cancer patients. *Head & neck oncology,* 3**,** 3-3.

SILVEIRA, A., MONTEIRO, E. & SEQUEIRA, T. 2018. Head and Neck Cancer: Improving Patient-Reported Outcome Measures for Clinical Practice. *Curr Treat Options Oncol,* 19**,** 59.

SIMPELAERE, I., WHITE, A., BEKKERING, G. E., GEURDEN, B., VAN NUFFELEN, G. & DE BODT, M. 2016. Patient-reported and proxy-reported outcome measures for the assessment of health-related quality of life among patients receiving enteral feeding: a systematic review protocol. *JBI Database System Rev Implement Rep,* 14**,** 45-75.

SMART, A. 2006. A multi-dimensional model of clinical utility. *International Journal for Quality in Health Care,* 18**,** 377-382.

SMITH, A. B., COCKS, K., PARRY, D. & TAYLOR, M. 2016. A Differential Item Functioning Analysis of the EQ-5D in Cancer. *Value Health,* 19**,** 1063-1067.

SMITH, K. W., AVIS, N. E. & ASSMANN, S. F. 1999. Distinguishing between quality of life and health status in quality of life research: a meta-analysis. *Qual Life Res,* 8**,** 447-59.
SNYDER, C. F., JENSEN, R. E., GELLER, G., CARDUCCI, M. A. & WU, A. W. 2010. Relevant content for a patient-reported outcomes questionnaire for use in oncology clinical practice: Putting doctors and patients on the same page. *Qual Life Res,* 19**,** 1045-55.
SNYDER, C. F., WATSON, M. E., JACKSON, J. D., CELLA, D. & HALYARD, M. Y. 2007. Patient-reported outcome instrument selection: designing a measurement strategy. *Value Health,* 10 Suppl 2**,** S76-85.
SPEYER, R., CORDIER, R., FARNETI, D., NASCIMENTO, W., PILZ, W., VERIN, E., WALSHE, M. & WOISARD, V. 2022. White Paper by the European Society for Swallowing Disorders: Screening and Non-instrumental Assessment for Dysphagia in Adults. *Dysphagia,* 37**,** 333-349.
SPEYER, R., CORDIER, R., KERTSCHER, B. & HEIJNEN, B. J. 2014. Psychometric properties of questionnaires on functional health status in oropharyngeal dysphagia: a systematic literature review. *Biomed Res Int,* 2014**,** 458678.
SPEYER, R., HEIJNEN, B. J., BAIJENS, L. W., VRIJENHOEF, F. H., OTTERS, E. F., ROODENBURG, N. & BOGAARDT, H. C. 2011. Quality of life in oncological patients with oropharyngeal dysphagia: validity and reliability of the Dutch version of the MD Anderson Dysphagia Inventory and the Deglutition Handicap Index. *Dysphagia,* 26**,** 407-14.
STARMER, H. M., TIPPETT, D., WEBSTER, K., QUON, H., JONES, B., HARDY, S. & GOURIN, C. G. 2014. Swallowing outcomes in patients with oropharyngeal cancer undergoing organ-preservation treatment. *Head Neck,* 36**,** 1392-7.
STATA PRESS 2021. Item Response Theory Reference Manual. College Station, TX: Stata Press.
STEWART, V. M., MENDIS, M. D. & LOW CHOY, N. 2018. A systematic review of patient-reported measures associated with vestibular dysfunction. *The Laryngoscope,* 128**,** 971-981.
STORY, D. A. & TAIT, A. R. 2019. Survey Research. *Anesthesiology,* 130**,** 192-202.
STOVER, A. M., MCLEOD, L. D., LANGER, M. M., CHEN, W.-H. & REEVE, B. B. 2019. State of the psychometric methods: patient-reported outcome measure development and refinement using item response theory. *Journal of Patient-Reported Outcomes,* 3**,** 50.
STREINER, D., NORMAN, G. R. & CAIRNEY, J. 2015. *Health measurement scales : a practical guide to their development and use,* New York, New York, Oxford University Press.
SULLIVAN, G. M. & ARTINO, A. R., JR 2017. How to Create a Bad Survey Instrument. *Journal of Graduate Medical Education,* 9**,** 411-415.
TAVAKOL, M. & DENNICK, R. 2011. Making sense of Cronbach's alpha. *Int J Med Educ,* 2**,** 53-55.
TEDDLIE, C. & TASHAKKORI, A. 2009. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences,* London, SAGE Publications.
TERESI, J. A. & FLEISHMAN, J. A. 2007. Differential item functioning and health assessment. *Qual Life Res,* 16 Suppl 1**,** 33-42.
TERWEE, C. B., BOT, S. D., DE BOER, M. R., VAN DER WINDT, D. A., KNOL, D. L., DEKKER, J., BOUTER, L. M. & DE VET, H. C. 2007. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol,* 60**,** 34-42.
TERWEE, C. B., PRINSEN, C. A. C., CHIAROTTO, A., WESTERMAN, M. J., PATRICK, D. L., ALONSO, J., BOUTER, L. M., DE VET, H. C. W. & MOKKINK, L. B. 2018. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res,* 27**,** 1159-1170.
THE SCOTTISH GOVERNMENT 2010. The Healthcare Quality Strategy for NHS Scotland. Edinburgh: The Scottish Government.
THOMPSON, D. R., SKI, C. F., GARSIDE, J. & ASTIN, F. 2016. A review of health-related quality of life patient-reported outcome measures in cardiovascular nursing. *European Journal of Cardiovascular Nursing,* 15**,** 114-125.

TIMMERMAN, A. A., MEESTERS, C. M. G., SPEYER, R. & ANTEUNIS, L. J. C. 2007. Psychometric qualities of questionnaires for the assessment of otitis media impact. *Clinical Otolaryngology,* 32**,** 429-439.

TIMMERMAN, A. A., SPEYER, R., HEIJNEN, B. J. & KLIJN-ZWIJNENBERG, I. R. 2014. Psychometric characteristics of health-related quality-of-life questionnaires in oropharyngeal dysphagia. *Dysphagia,* 29**,** 183-98.

TUOMI, L., FRANSSON, P., WENNERBERG, J. & FINIZIA, C. 2020. A longitudinal study of the Swedish MD Anderson Dysphagia Inventory in patients with oral cancer. *Laryngoscope Investigative Otolaryngology,* 5**,** 1125-1132.

TURNER, G. M., LITCHFIELD, I., FINNIKIN, S., AIYEGBUSI, O. L. & CALVERT, M. 2020. General practitioners' views on use of patient reported outcome measures in primary care: a cross-sectional survey and qualitative study. *BMC Family Practice,* 21**,** 14.

URSINO, S., CALISTRI, E., DE FELICE, F., BONOMO, P., DESIDERI, I., FRANCO, P., ARCADIPANE, F., COLOSIMO, C., MAZZOLA, R., MADDALO, M., GONNELLI, A., MALFATTI, G., MORGANTI, R., MUSIO, D. & PAIAR, F. 2022. Patient-Reported Outcomes After Swallowing (SWOARs)-Sparing IMRT in Head and Neck Cancers: Primary Results from a Prospective Study Endorsed by the Head and Neck Study Group (HNSG) of the Italian Association of Radiotherapy and Clinical Oncology (AIRO). *Dysphagia***,** 1-12.

VAISMORADI, M., TURUNEN, H. & BONDAS, T. 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences,* 15**,** 398-405.

VERDONCK-DE LEEUW, I., DAWSON, C., LICITRA, L., ERIKSEN, J. G., HOSAL, S., SINGER, S., LAVERTY, D. P., GOLUSINSKI, W., MACHCZYNSKI, P., VARGES GOMES, A., GIRVALAKI, C., SIMON, C. & LEEMANS, C. R. 2022. European Head and Neck Society recommendations for head and neck cancer survivorship care. *Oral Oncol,* 133**,** 106047.

VICKERS, A. J. & CHEN, L. Y. 2017. Manifesto: towards a clinically-oriented psychometrics. *Health and Quality of Life Outcomes,* 15**,** 83.

VOUTILAINEN, A., PITKÄAHO, T., KVIST, T. & VEHVILÄINEN-JULKUNEN, K. 2016. How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *J Adv Nurs,* 72**,** 946-57.

WALLER, N., JOHN, M. T., FEUERSTAHLER, L., BABA, K., LARSSON, P., PERŠIĆ, S., KENDE, D., REIßMANN, D. R. & RENER-SITAR, K. 2016. A 7-day recall period for a clinical application of the oral health impact profile questionnaire. *Clin Oral Investig,* 20**,** 91-9.

WATERS, A. M., TUDUR SMITH, C., YOUNG, B. & JONES, T. M. 2014. The CONSENSUS study: protocol for a mixed methods study to establish which outcomes should be included in a core outcome set for oropharyngeal cancer. *Trials,* 15**,** 168.

WELDRING, T. & SMITH, S. M. S. 2013. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health services insights,* 6**,** 61-68.

WILMSKOETTER, J., BONILHA, H., HONG, I., HAZELWOOD, R. J., MARTIN-HARRIS, B. & VELOZO, C. 2019. Construct validity of the Eating Assessment Tool (EAT-10). *Disabil Rehabil,* 41**,** 549-559.

WILSON, J. A., CARDING, P. N. & PATTERSON, J. M. 2011. Dysphagia after nonsurgical head and neck cancer treatment: patients' perspectives. *Otolaryngol Head Neck Surg,* 145**,** 767-71.

WISHART, L. R., HARRIS, G. B., CASSIM, N., ALIMIN, S., LIAO, T., BROWN, B., WARD, E. C. & NUND, R. L. 2022. Association Between Objective Ratings of Swallowing and Dysphagia-Specific Quality of Life in Patients Receiving (Chemo)radiotherapy for Oropharyngeal Cancer. *Dysphagia,* 37**,** 1014-1021.

WOLPE, J. 1969. *The practice of behavior therapy,* New York, NY, Pergamon Press.

YANG, L., ZHANG, Z., GAO, H., WU, Y., WEI, H., KONG, J., WANG, R., CHENG, J. & TIAN, J. 2022. Cultural Adaptation and Validation of Questionnaires for Evaluation of Health-Related Quality of Life with Dysphagia in Different Countries: A Systematic Review. *Dysphagia,* 37**,** 812-823.

YEE, K., WONG, S. M., TEO, I., LOY, J., ROCHE, E., TAN, Y. P., TAN, H. K., TAN, N. C. & IYER, N. G. 2020. Validity and reliability of the MD Anderson dysphagia inventory in English and Chinese in head and neck cancer patients. *Asia-Pacific Journal of Clinical Oncology,* 16**,** 372-379.

ZANON, C., HUTZ, C., YOO, H. & HAMBLETON, R. 2016. An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica,* 29**,** 18.

ZHANG, L.-J., JIANG, N., LI, Z., CHEN, X.-W., WANG, P.-G., WANG, X. & ZHAO, Y. 2017. Psychometric Properties of the Chinese Version of the M.D. Anderson Dysphagia Inventory for Head and Neck Cancer Patients. *Cancer nursing,* 40**,** E9-E16.

ZRAICK, R. I., ATCHERSON, S. R. & HAM, B. K. 2012. Readability of patient-reported outcome questionnaires for use with persons with swallowing disorders. *Dysphagia,* 27**,** 346-52.

APPENDICES

Appendix A: The MDADI

MD Anderson Dysphagia Inventory

This questionnaire asks for your views about your swallowing ability. This information will help us understand how you feel about swallowing. The following statements have been made by people who have problems with their swallowing. Some of the statements may apply to you. Please read each statement and circle the response which best reflects your experience in the past week.

1. My swallowing ability limits my day to day activities

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

2. I am embarrassed by my eating habits

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

3. People have difficulty cooking for me

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

4. Swallowing is more difficult at the end of the day

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

5. I do not feel self-conscious when I eat

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

6. I am upset by my swallowing problem

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

7. Swallowing takes great effort

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

8. I do not go out because of my swallowing problem

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

9. My swallowing difficulty has caused me to lose income

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

10. It takes me longer to eat because of my swallowing problem

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

11. People ask me 'why can't you eat that?'

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

12. Other people are irritated by my eating problem

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

13. I cough when I try to drink liquids

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

14. My swallowing problems limit my personal and social life

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

15. I feel free to go out to eat with my friends, neighbours, relatives

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

16. I limit my food intake because of my swallowing difficulty

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

17. I cannot maintain my weight because of my swallowing problem

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

18. I have low self esteem because of my swallowing problem

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

19. I feel that I am swallowing a huge amount of food

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

20. I feel excluded because of my eating habits

Strongly agree *agree* *no opinion* *disagree* *strongly disagree*

Appendix B: Clinical Utility database search strategy

CINAHL 25/03/21

Search	Search term	Results
S1	MM "Patient-Reported Outcomes”	1571
S2	“clinical utility” OR (MH “Outcomes Research”)	14083
S3	S1 AND S2	12

Limited to peer-reviewed academic journals reduced the total to 11. 11 abstracts were reviewed and **2** were found to be potentially relevant and the full text was accessed. 1 article was included in the review.

MEDLINE 1946-24th March 2021 – 25/03/21

Search	Search term	Results
S1	Patient reported outcome measure.mp. or Patient Reported Outcome Measures/	8761
S2	limit 1 to (English language and humans)	8213
S3	Clinical utility.mp.	27587
S4	limit 3 to (English language and humans)	20744
S5	2 and 4	33
S6	Usability.mp	14838
S7	limit 6 to (English language and humans)	8361
S8	2 and 7	46
S9	Implementation.mp.	271192
S10	limit 9 to (English language and humans)	161782
S11	2 and 10	316
S12	Practical issues.mp	3139
S13	limit 12 to (English language and humans)	2118
S14	2 and 13	4
S15	5 or 8 or 11 or 14	387

387 abstracts were studied and 357 were excluded. Of the remaining **30** articles, 1 duplicate was removed and full text was analysed for meeting inclusion criteria for the remaining **29** articles. Ultimately 7 articles met the inclusion/exclusion criteria and was included in the review.

EMBASE 1974 – 24th March 2021 – 25/03/21

Search	Search term	Results
S1	Patient reported outcome measure.mp. or Patient Reported Outcome Measures/	30571
S2	limit 1 to (English language and humans)	30141
S3	Clinical utility.mp.	41184
S4	limit 3 to (English language and humans)	35602
S5	2 and 4	127
S6	Usability.mp	19026
S7	limit 6 to (English language and humans)	15010
S8	2 and 7	224
S9	Implementation.mp.	361123
S10	limit 9 to (English language and humans)	267981
S11	2 and 10	1313
S12	Practical issues.mp	3960
S13	limit 12 to (English language and humans)	2949
S14	2 and 13	13
S15	5 or 8 or 11 or 14	1609
S16	5 or 8 or 14	360

Implementation removed from search terms as found to be less relevant from Medline search review of abstracts using that search term.

360 abstracts were studied and **319** were excluded. Of the remaining **41** articles, duplicates were removed and full text of **33**was analysed for meeting inclusion criteria. Ultimately 3 articles met these criteria.

PubMed 25/03/21

Using MeSH terms “assessment, patient outcomes”, “feasibility studies” (clinical utility not recognised as a search term)

Search	Search term	Results
S1	Search ("**assessment, patient outcomes** ") AND ("**feasibility studies** ")	249
S2	Filters: English, Humans, Medline results excluded (as Medline searched separately)	16

Of the 16 studies, full text of 1 article was accessed after screening of abstracts.

Of these, 0 were appropriate for inclusion in this literature review as none met all inclusion criteria.

Additional articles from ref lists:

Thompson etal 2016

Higginson &Carr 2001;

Patel etal 2017

Montgomery etal 2020

Howell etal 2014

Appendix C: MDADI database search strategy

CINAHL 30/04/21

Search	Search term	Results
S1	TI "MDADI"	6
S2	SU "MDADI"	0
S3	TI "MD Anderson Dysphagia Inventory"	10

After exclusions/duplicates: 11 articles accessed in full text

MEDLINE 1946-30th April 2021– 30/04/21

Search ID	Search term	Results
S1	MDADI.mp	150
S2	limit 1 to (English language and humans)	118
S3	MD Anderson Dysphagia Inventory.mp	155
S4	limit 3 to (English language and humans)	125
S5	2 or 4	151
S6	Dysphagia.mp	29814
S7	limit 6 to (English language and humans)	20425
S8	Patient reported outcome measures.mp or Patient Reported Outcome Measures/	12023
S9	Limit 8 to (English language and humans)	10065
S10	7 and 9	67

218 abstracts were studied and 195 were excluded. Of the remaining **23** articles, 13 duplicates were removed and full text was analysed for meeting inclusion criteria for the remaining **10** articles. Ultimately x articles met the inclusion/exclusion criteria and was included in the review.

EMBASE 1974 – 29th April 2021 – 30/04/21

Search ID	Search term	Results
S1	MDADI.mp.	288
S2	limit 1 to (English language and humans)	276
S3	MD Anderson dysphagia inventory.mp.	281
S4	limit 3 to (English language and humans)	272
S5	2 or 4	346
S6	Patient reported outcome measure.mp. or Patient Reported Outcome Measures/	31437
S7	limit 6 to (English language and humans)	31005
S8	Dysphagia.mp. or dysphagia/	84948
S9	limit 8 to (English language and humans)	72916
S10	7 and 9	382
S11	5 or 10	695

695 abstracts were studied and 662 were excluded. Of the remaining **33** articles, 22 duplicates were removed and full text of **11** were analysed for meeting inclusion criteria.

PubMed 25/03/21

Using MeSH terms "assessment, patient outcomes", "feasibility studies" (clinical utility not recognised as a search term)

Search ID	Search term	Results
S1	Search **MDADI** Filters: **Humans, English, Dental journals, Nursing journals (all fields)**	6
S2	**Search MD Anderson dysphagia inventory** Filters: **Humans, English, Dental journals, Nursing journals (all fields)**	11
S3	Search: **(dysphagia) AND (patient reported outcome measures)** Filters: **Humans, English, Dental journals, Nursing journals**	3

Of the 20 results, full text of 1 article was accessed after screening of abstracts however this was a duplicate of an article located in another search engine search.

Additional articles from ref lists: Kwon etal 2013

Appendix D: Online survey content

INTRODUCTION

Research Project Title

Analysing and enhancing the MD Anderson Dysphagia Inventory: a mixed methods analysis of the head and neck cancer swallowing-related quality of life patient-reported outcome tool

Background

This questionnaire forms part of a project called "Analysing and enhancing the MD Anderson Dysphagia Inventory: a mixed methods analysis of the head and neck cancer swallowing-related quality of life patient reported outcome tool" being carried out by Kate Toft, contributing towards a Clinical Doctorate at the University of Stirling. Dr Catherine Best and Professor Jayne Donaldson are supervising this project.

Topic

The aim of this questionnaire is to explore your use and opinion of the MD Anderson Dysphagia Inventory (MDADI). The MDADI is a self-administered written patient-reported outcome measure questionnaire. It quantifies swallowing-related quality of life, and is specifically designed for use with patients with head and neck cancer (Chen et al., 2001). The MDADI is one of the most frequently used dysphagia outcome assessment tools in research practice (Ojo et al., 2012) and is often used as a main outcome tool in multicentre trials (Hutcheson et al., 2016). To date however, no validation of the tool has taken place with UK patient data, and clinicians and patients have not been asked what they think of the content of the tool and its use in clinical practice.

This questionnaire is designed to be completed by Speech and Language Therapists (SLTs) who have experience of working with patients with head and neck cancer (HNC), whether or not you currently use the MDADI in your clinical or research practice. This will be the first time that UK SLTs have been asked about the MDADI. This data provides insight into the clinical utility of the MDADI, as well as informing a second stage of the research project which will be a statistical validation of the psychometric properties of the tool using UK patient data.

What is involved if you choose to take part?

The questionnaire starts by asking for some background information. Then you will be asked for your thoughts and comments about each of the 20 items in the MDADI. Following this, there are more general questions about using the tool.
There are no right or wrong answers when completing this questionnaire – the aim is to gather and understand SLTs' views on the tool and its use in their HNC practice.
Please consider providing anonymised illustrative examples from your own practice and your thoughts on the MDADI. An opportunity for general comments is given at the end of the survey.

Do I have to take part?

No. Your participation in this survey is voluntary. You may refuse to take part in the research or exit the survey at any time without penalty by pressing the 'Exit' button / closing the browser.
You can only take the survey once, but you can edit your responses until the survey is closed on [date].

Are there any potential risks in taking part?

There are no foreseeable risks involved in participating in this survey.

Are there any benefits in taking part?

The benefits of taking part are contributing to the evidence base in head and neck cancer research, and to the first UK validation of the MDADI tool.

Will my responses be made anonymous?

Personally identifiable information will not be requested in this survey. If giving examples from your own practice, please do not provide any patient-identifiable information. Verbatim quotes may be used in reports and publications and to support other research in the future. You will not be identifiable if quotes are used.

If I volunteer to take part, how will my consent be recorded?

By completing the consent form on the following page you are consenting to your anonymised data being included in the analysis for this project.

Will I be informed of the study results?

If you would like to be informed of the study results on its completion, please insert your email address into the questionnaire when requested and you will be added to the study dissemination mailing list. Providing your email address is entirely voluntary.

Legal basis for processing personal data

As part of the project we will be recording personal data relating to you. This will be processed in accordance with the General Data Protection Regulations (GDPR). Under GDPR the legal basis for processing your personal data will be public interest/the official authority of the University.

What happens to the data I provide?

Your answers will be completely anonymous, and we will use all reasonable endeavours to keep them confidential. Your data will be stored in a password-protected file [and may be used in academic publications]. Your IP address will not be stored.

Your personal data will be kept for 2 years on the X Research Drive – a secure data centre on the Stirling campus and then will be securely destroyed.

We ask all participants for their permission to use direct, anonymised quotes.

Will the research be published?

The research may be published in peer-reviewed scientific journals and/or presented at educational and scientific meetings.

Who is organising and funding the research?

NHS Lothian and the University of Stirling are sponsoring/funding this research.

Who has reviewed this research project?

The ethical approaches of this project have been approved via The University of Stirling [NHS, Invasive & Clinical Research Ethics Committee] and IRAS.

Your rights

You have the right to request to see a copy of the information we hold about you and to request corrections or deletions of the information that is no longer required. You have the right to withdraw from this project at any time without giving reasons and without consequences to you. You also have the right to object to us processing relevant personal data however, please note that once the data are being analysed and/or results published it may not be possible to remove your data from the study.

Whom do I contact if I have concerns about this study or I wish to complain?

If you would like to discuss the research with someone please contact Katherine.Toft1@stir.ac.uk or Jayne.Donaldson@stir.ac.uk . In case of complaint please contact Dr Angus Hunter, Associate Dean Research, Faculty of Health Sciences and Sport a.m.hunter1@stir.ac.uk .

You have the right to lodge a complaint against the University regarding data protection issues with the Information Commissioner's Office (https://ico.org.uk/concerns/). The University's Data Protection Officer is Joanna Morrow, Deputy Secretary. If you have any questions relating to data protection these can be addressed to data.protection@stir.ac.uk in the first instance.

THANK YOU FOR YOUR PARTICIPATION IN THIS SURVEY!

ELECTRONIC CONSENT PAGE

Research Project Title: Analysing and enhancing the MD Anderson Dysphagia Inventory: a mixed methods analysis of the head and neck cancer swallowing-related quality of life patient-reported outcome tool

(Participants will have to check all the following boxes before being able to progress to the rest of the questionnaire).

	Please check box
I confirm that I have read and understood the Participants Information Sheet Version 4 dated August 2021 explaining the above research project and I have had the opportunity to ask questions about the project	
I understand that my participation is voluntary and that I am free to withdraw at any time during the study and withdraw my data before October 2021 without giving a reason, and without any penalty. I understand that after October 2021 it will not be possible to remove my data from the study.	
I understand that my responses will be kept anonymous and I give permission for members of the research team to have access to my anonymised responses.	
I give permission to be quoted directly and anonymously in the research publication	
I agree for research data collected in the study to be given to researchers. I understand that any data that leaves the research group will be fully anonymised so that I cannot be identified.	
I agree to take part in this study	

Section 1: General background information

1. What is your UK country of work?
   Scotland England Wales N Ireland
2. How many years head & neck dysphagia experience do you have? (to nearest year)
3. Do you currently use the MDADI in your clinical HNC practice? Yes/No
   If yes go to 6 If no go to 4
4. Have you used the MDADI in your clinical HNC practice in the past? Yes / No
5. Why did you stop using the MDADI? (free text box)
6. At what time points do you use the MDADI? (select all that apply)

   Pre-treatment on-treatment post-treatment

7. If you use the MDADI post-tx, at what time points do you use it? (free text box)

Section 2 : MDADI Items - each individual item in the MDADI giving an opportunity for item-specific comments (as per COSMIN content validity checklist). Please consider relevance in terms of the HNC population and HNC-related dysphagia.

When responding to these questions please ensure that information that could potentially identify a patient or staff must not be included in any response.

8. **My swallowing ability limits my day to day activities**
   Relevant not relevant
   Comments:
9. **I am embarrassed by my eating habits**
   Relevant not relevant
   Comments:
10. **People have difficulty cooking for me**
    Relevant not relevant
    Comments:
11. **Swallowing is more difficult at the end of the day**
    Relevant not relevant
    Comments:
12. **I do not feel self-conscious when I eat**
    Relevant not relevant
    Comments:
13. **I am upset by my swallowing problem**
    Relevant not relevant
    Comments:
14. **Swallowing takes great effort**
    Relevant not relevant
    Comments:
15. **I do not go out because of my swallowing problem**
    Relevant not relevant
    Comments:
16. **My swallowing difficulty has caused me to lose income**
    Relevant not relevant
    Comments:
17. **It takes me longer to eat because of my swallowing problem**
    Relevant not relevant
    Comments:
18. **People ask me 'why can't you eat that?'**
    Relevant not relevant
    Comments:
19. **Other people are irritated by my eating problem**
    Relevant not relevant
    Comments:
20. **I cough when I try to drink liquids**
    Relevant not relevant
    Comments:
21. **My swallowing problems limit my personal and social life**

Relevant not relevant
Comments:

22. **I feel free to go out to eat with my friends, neighbours, relatives**
Relevant not relevant
Comments:
23. **I limit my food intake because of my swallowing difficulty**
Relevant not relevant
Comments:
24. **I cannot maintain my weight because of my swallowing problem**
Relevant not relevant
Comments:
25. **I have low self-esteem because of my swallowing problem**
Relevant not relevant
Comments:
26. **I feel that I am swallowing a huge amount of food**
Relevant not relevant
Comments:
27. **I feel excluded because of my eating habits**
Relevant not relevant
Comments:

Section 3: Further comments on MDADI contents and structure

When responding to these questions please ensure that information that could potentially identify a patient or staff must not be included in any response.

28. Are there any head and neck cancer patients with whom you would not use this tool, and if so, why? Yes/No (free text box)
29. Please comment on whether you think there is anything missing from the MDADI content (free text box)
30. Please comment on whether you think there is anything extraneous in the MDADI content (free text box)
31. Please comment on how you find using the MDADI (e.g. tool length, item clarity, response format, time taken to complete tool, scoring) (free text box)
32. What are the strengths of the MDADI? (free text box)
33. What are the weaknesses of the MDADI? (free text box)
34. What comments have you had from patients about the MDADI? (free text box)
35. Please comment about the possible emotional impact of completing the tool on patients, and/or whether patients have ever commented on this
36. Please note any other comments you have on the tool here (free text box)

Final page

Many thanks for taking the time to complete this questionnaire Your time and contribution is much appreciated. If you would like to be informed of the results of this study following its completion please contact Kate Toft at Katherine.toft1@stir.ac.uk to be added to the results dissemination mailing list.

Appendix E: Survey invite email

Dear dysphagia practitioner

As part of a Clinical Doctorate at Stirling University, I am carrying out a study: "Analysing and enhancing the MD Anderson Dysphagia Inventory: a mixed methods analysis of the head and neck cancer swallowing-related quality of life patient reported outcome tool".

I have contacted you, as an **SLT with experience working with patients with head and neck cancer and dysphagia,** to ask if you would complete a questionnaire on this topic. Questionnaire responses will form part of the data that I am collecting for the study.

I have attached a Participant Information Sheet, which gives more information about the study and what it involves.

The questionnaire is accessed here:

[URL]

The questionnaire will be open from XXXX until XXXXX and will take approximately 20-30 minutes to complete. Your participation and time taken to complete this questionnaire are very much appreciated.

Please feel free to circulate this link to other SLT colleagues who work, or have worked, in the field of adult head & neck cancer-related dysphagia.

If you have any queries please contact me on Katherine.Toft1@stir.ac.uk.

With many thanks

Kate Toft

Speech & Language Therapist, NHS Lothian

Clinical Doctorate candidate, University of Stirling

Appendix F: Survey PIS

Participant Information Sheet

Research Project Title

"Analysing and enhancing the MD Anderson Dysphagia Inventory"

Background, aims of project

You are invited to take part in this research study. Joining the study is entirely up to you. This information sheet is to help you understand the purpose of this project and what it would involve for you if you participate. Please contact Kate Toft, who is the lead investigator, if anything is unclear.

The aim of this study is to analyse the MD Anderson Dysphagia Inventory (MDADI) in detail. The MDADI is a self-administered written patient-reported outcome measure questionnaire. It quantifies swallowing-related quality of life, and is specifically designed for use with patients with head and neck cancer (Chen et al., 2001). The MDADI is one of the most frequently used dysphagia outcome assessment tools in research practice (Ojo et al., 2012) and is often used as a main outcome tool in multicentre trials (Hutcheson et al., 2016). To date however, no validation of the tool has taken place with UK patient data, and clinicians and patients have not been asked what they think of the content of the tool and its use in clinical practice.

Why have I been invited to take part?

You have been invited because you are a Speech & Language Therapist (SLT) who has experience of working with patients with head and neck cancer, and as someone who has experience of using the MDADI tool.

Do I have to take part?

No. You do not have to take part. If you do decide to take part, you can withdraw your participation at any time without needing to explain and without penalty by advising the researcher of this decision. If you withdraw we will not collect any more data from you. However, any data collected up until the point that you withdraw will be kept and used in the data analysis. You will be asked to complete an electronic consent form as part of completing the online questionnaire.

What will happen if I take part?

Taking part in this study will involve you in responding to an internet-based questionnaire looking in detail at the MDADI. Your opinion on the content of each of the 20 items in the questionnaire will be sought, in addition to information on your experiences of using the tool in your clinical practice. Opportunity will also be provided for you to describe what you feel the strengths and weaknesses of the tool are. The questionnaire should take approximately 20-30 minutes.

Are there any potential risks in taking part?

There are no foreseeable risks in taking part.

Are there any benefits in taking part?

The benefits of taking part are participating in expanding the research evidence base of clinical practice in head & neck cancer Speech & Language Therapist. This will be the first time that UK SLTs have been asked about the MDADI. This data will provide insight into the clinical utility of the MDADI, as well as informing a second stage of the research project which will be a statistical validation of the psychometric properties of the tool using UK patient data.

Legal basis for processing personal data

As part of the project we will be recording personal data relating to you (for example how long you have been working for as an SLT) but these will be anonymised within the research findings. This will be processed in accordance with the General Data Protection Regulation (GDPR). Under GDPR the legal basis for processing your personal data will be public interest/the official authority of the University.

What happens to the data I provide?

The research data will be kept anonymous as demographic data will not be linked to any quotes used. The research data will be kept anonymously, stored securely and will not be shared with a third party. The data will be kept for 2 years on the University of Stirling Research Drive – a secure data centre on the Stirling campus and then will be securely destroyed. We will ask all participants for their permission to use direct quotes in the online consent form.

Will the research be published?

The research may be published in relevant peer reviewed journals. You will not be identifiable in any report/publication. The University of Stirling is committed to making the outputs of research publicly accessible and supports this commitment through our online open access repository STORRE. Unless funder/publisher requirements prevent us, this research will be publicly disseminated through our open access repository.

Who is organising and funding the research?

NHS Lothian and the University of Stirling are sponsoring/funding this research.

Who has reviewed this research project?

The ethical approaches of this project have been approved via The University of Stirling NHS, Invasive & Clinical Research Ethics Committee.

Your rights

You have the right to request to see a copy of the information we hold about you and to request corrections or deletions of the information that is no longer required. You have the right to withdraw from this project at any time without giving reasons and without consequences to you. You also have the right to object to us processing relevant personal data however, please note that once the data are being analysed and/or results published it may not be possible to remove your data from the study.

Who do I contact if I have concerns about this study or I wish to complain?

If you would like to discuss the research with someone please contact Katherine.Toft1@stir.ac.uk or Federico.Andreis@stir.ac.uk . In case of complaint please contact Dr Angus Hunter, Associate Dean Research, Faculty of Health Sciences and Sport a.m.hunter1@stir.ac.uk. You have the right to lodge a complaint against the University regarding data protection issues with the Information Commissioner's Office (https://ico.org.uk/concerns/). The University's Data Protection Officer is Joanna Morrow, Deputy Secretary. If you have any questions relating to data protection these can be addressed to data.protection@stir.ac.uk in the first instance.

Thank you for your participation.

Appendix G: List of thematic analysis codes at each stage

- relevant or useful item for SLT
- Question wording comment or suggestion for revision
- ambiguous or confusing wording for patients
- Insensitive, unnecessarily distressing, emotive or negative
- irrelevant items or insufficiently specific or sensitive
- clear wording
- length of tool
- repetition or similar to another item or item redundancy
- issues with or alternative to Likert scale
- item not informative or relevant for SLT input with patient
- useful tool overall for SLT
- not sufficiently inclusive for all patients
- can lead to useful therapeutic conversation starter
- question wording or patient response not specific to swallowing
- Doesn't differentiate swallowing from other issues e.g. xerostomia
- not relevant to patients with no dysphagia yet
- scoring
- literacy or accessibility issues
- SLT has to make own interpretation of item
- COVID
- also captured in therapeutic conversations
- what about NBM patients
- use of word HABIT
- doesn't influence input or practice
- what about people who live alone or a socially isolated
- layout
- Likert is ok
- doesn't take into account nutritional support
- complementary link with other assessments
- subscales
- evidence of MDADI not being used as intended or modified
- item not sufficiently specific to HNC
- no differentiation between chewing and swallowing or oral vs pharyngeal
- issues with carrying out MDADI over phone or video
- doesn't take long
- important aspects of function missed
- insufficient oral dg coverage
- patients embarrassed by or reluctant to answer item
- emotionally difficult but necessary
- Sex differences in response
- unhelpful separation of eating from drinking
- useful for research
- ref to textures would be useful
- helpful for therapy
- item framed as a choice but may not be a choice
- frustrating
- potential for bias or external influence on response
- a way of acknowledging impact
- overlap with other outcome measures
- usefulness of global question
- potential for context to influence responses
- eating vs swallowing
- link between item subject and swallowing difficult for patients to understand
- positive spin of item is good

- Age
- ambiguous item requiring explanation
- clear wording
- Clinically relevant item
- complementary link with other assessments
- COVID
- Doesn't differentiate swallowing from other issues e.g. xerostomia
- doesn't take into account nutritional support
- eating vs swallowing
- emotionally difficult but necessary
- excludes NBM patients
- excludes patients who do not work
- excludes socially isolated pts
- excludes pts with low socioeconomic status
- Insensitive wording
- issues with or alternative to Likert scale
- item not relevant or helpful and could be excluded
- Item would be better if wording was changed
- layout CLINICAL UTILITY ADJUNCT
- Likert is ok
- Negative spin
- non literacy accessibility issues
- not patient centred
- not relevant to patients with no dysphagia yet
- oral vs pharyngeal
- palliative patients
- patient confusion with tool items
- patients find item embarrassing affecting response
- patients have a negative emotional response to item or using the tool
- potential for bias or external influence on response
- repetition or similar to another item or item redundancy
- Sex differences in response
- teeth
- trick questions
- unhelpful separation of eating from drinking
- use of word HABIT
- useful therapeutic conversation starter
- What is missing

- Age
- ambiguous item requiring explanation
- clear wording
- Clinically relevant item
- complementary link with other assessments
- COVID
- Doesn't differentiate swallowing from other issues eg xerostomia
- doesn't take into account nutritional support
- eating vs swallowing
- emotionally difficult but necessary
- excludes NBM patients
- excludes patients who do not work
- excludes socially isolated pts
- excludes pts with low socioeconomic status
- Insensitive wording
- issues with or alternative to Likert scale
- item not relevant or helpful and could be excluded
- Item would be better if wording was changed
- layout CLINICAL UTILITY ADJUNCT
- Likert is ok
- Negative spin
- non literacy accessibility issues
- not patient centred
- not relevant to patients with no dysphagia yet
- oral vs pharyngeal
- palliative patients
- patient confusion with tool items
- patients find item embarrassing affecting response
- patients have a negative emotional response to item or using the tool
- potential for bias or external influence on response
- R1 - impact on clinical decision making
- R2 - issues with use with subgroups
- R3 - SEM or MCID data available
- R4 - are questions distressing
- R5 - use of proxy
- R6 - is the tool acceptable to patients
- R7 - are the items relevant
- R8 - is the recall timeframe appropriate
- repetition or similar to another item or item redundancy
- Sex differences in response
- teeth
- trick questions
- U1 - appropriate literacy level
- U2 - is time taken to administer acceptable
- U3 - is scoring straightforward
- U4 - time taken to score
- unhelpful separation of eating from drinking
- use of word HABIT
- useful therapeutic conversation starter
- What is missing

Initial themes

Theme	Codes
Extra clinical utility issues	Non literacy accessibility issues Issues with or alternative to Likert scales Layout Potential for bias or external influence on response
Positive aspects of clinical utility	Clear wording Clinically relevant item
Helpful in clinical practice	Useful therapeutic conversation starter Complementary link with other assessments
Excluded groups	Excludes NBM patients Excludes socially isolated patients Excludes patients who do not work Not relevant to patients with no dysphagia yet Excludes patients with low socioeconomic status Palliative patients
Potential for DIF?	Age Sex differences in responses
The bigger picture of eating and drinking	Doesn't differentiate swallowing from other issues e.g. xerostomia Eating vs swallowing Teeth Doesn't take into account nutritional support Unhelpful separation of eating from drinking Oral vs pharyngeal
Not user-friendly	Negative spin Patients have a negative emotional response to item or using the tool Not patient centred Insensitive wording Ambiguous item requiring explanation Trick questions
Suggestions for change	Repetition or similar to another item or item redundancy Use of word HABIT Item not relevant or helpful and could be excluded What is missing Item would be better if wording was changed

Final themes and codes

Theme	Codes
"a conversation starter" – The MDADI can be helpful in clinical practice	Useful therapeutic conversation starter Complementary link with other assessments Clinically relevant item
Practical issues	Non literacy accessibility issues Issues with or alternative to Likert scales Layout Potential for bias or external influence on response
Potential for DIF?	Age Sex differences in responses Excludes patients with low socioeconomic status
The bigger picture of eating and drinking	Eating vs swallowing Teeth Eating vs drinking Oral dysphagia vs pharyngeal dysphagia Doesn't differentiate swallowing from other issues
"not user-friendly"	Trick questions Negative spin Not patient-centred Insensitive wording Ambiguous item requiring explanation Patients have a negative emotional response to the item/tool
Excluded groups	Low socioeconomic status Patients who are palliative Patients who are nil by mouth Patients with no dysphagia yet Patients who are socially isolated Patients who don't work
"not quite where we need it to be" – suggestions for change	What is missing Use of word HABIT Repetition/item redundancy Item would be better if wording was changed Item not relevant or helpful and could be excluded

Appendix H: Mind map of codes and themes

Appendix I: IRT GRM Model fit tables

Pre-treatment

```
Fitting fixed-effects model:
Fitting fixed-effects model:
Iteration 0:   log likelihood = -5878.2554
Iteration 1:   log likelihood = -5875.5245
Iteration 2:   log likelihood = -5875.3413
Iteration 3:   log likelihood = -5875.3404
Iteration 4:   log likelihood = -5875.3404
Fitting full model:
Iteration 0:   log likelihood = -4492.3518
Iteration 1:   log likelihood = -3902.4154
Iteration 2:   log likelihood = -3816.9104
Iteration 3:   log likelihood = -3795.4678
Iteration 4:   log likelihood = -3793.0436
Iteration 5:   log likelihood = -3792.9901
Iteration 6:   log likelihood = -3793.0095
Iteration 7:   log likelihood =  -3793.016
Iteration 8:   log likelihood = -3793.0183
Iteration 9:   log likelihood = -3793.0191
Iteration 10:  log likelihood = -3793.0193
Graded response model                 Number of obs     =        277
Log likelihood = -3793.0193
```

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
q1						
Discrim	4.556	0.575	7.930	0.000	3.430	5.682
Diff						
>=2	-1.781		0.155	-2.085	-1.476	
>=3	-1.315		0.103	-1.516	-1.114	
>=4	-1.181		0.093	-1.362	-0.999	
=5	-0.523		0.072	-0.665	-0.382	
q2						
Discrim	3.568	0.438	8.150	0.000	2.710	4.426
Diff						
>=2	-1.853		0.171	-2.187	-1.518	
>=3	-1.309		0.108	-1.521	-1.097	
>=4	-1.199		0.100	-1.394	-1.004	
=5	-0.483		0.076	-0.633	-0.334	
q3						
Discrim	2.848	0.362	7.860	0.000	2.137	3.558
Diff						
>=2	-2.463		0.298	-3.046	-1.880	
>=3	-1.609		0.149	-1.900	-1.318	
>=4	-1.360		0.122	-1.598	-1.121	
=5	-0.572		0.082	-0.733	-0.411	
q4						
Discrim	3.388	0.417	8.130	0.000	2.571	4.205
Diff						
>=2	-2.012		0.203	-2.410	-1.614	
>=3	-1.481		0.125	-1.726	-1.236	
>=4	-1.215		0.101	-1.414	-1.017	
=5	-0.503		0.077	-0.655	-0.351	
q5						
Discrim	1.876	0.229	8.210	0.000	1.429	2.324
Diff						
>=2	-2.181		0.244	-2.659	-1.704	
>=3	-1.206		0.127	-1.454	-0.958	
>=4	-1.084		0.118	-1.314	-0.854	
=5	-0.184		0.097	-0.373	0.006	
q6						

Discrim	3.861	0.451	8.560	0.000	2.977	4.746
Diff						
>=2	-1.877		0.169	-2.207	-1.546	
>=3	-1.224		0.099	-1.418	-1.031	
>=4	-0.962		0.083	-1.125	-0.800	
=5	-0.403		0.075	-0.549	-0.257	
q7						
Discrim	3.915	0.452	8.660	0.000	3.029	4.801
Diff						
>=2	-1.776		0.154	-2.078	-1.473	
>=3	-1.102		0.091	-1.280	-0.924	
>=4	-0.982		0.084	-1.146	-0.818	
=5	-0.323		0.074	-0.469	-0.178	
q8						
Discrim	6.337	0.882	7.180	0.000	4.608	8.066
Diff						
>=2	-1.890		0.176	-2.235	-1.546	
>=3	-1.446		0.108	-1.657	-1.234	
>=4	-1.230		0.090	-1.407	-1.054	
=5	-0.655		0.070	-0.792	-0.519	
q9						
Discrim	3.343	0.476	7.020	0.000	2.410	4.276
Diff						
>=2	-2.586		0.381	-3.334	-1.839	
>=3	-1.923		0.195	-2.305	-1.540	
>=4	-1.522		0.135	-1.787	-1.257	
=5	-0.772		0.082	-0.932	-0.611	
q10						
Discrim	3.277	0.365	8.970	0.000	2.561	3.993
Diff						
>=2	-1.582		0.137	-1.850	-1.314	
>=3	-0.720		0.080	-0.876	-0.563	
>=4	-0.665		0.079	-0.819	-0.511	
=5	-0.187		0.079	-0.342	-0.032	
q11						
Discrim	3.830	0.466	8.230	0.000	2.918	4.743
Diff						
>=2	-2.238		0.243	-2.713	-1.763	
>=3	-1.405		0.115	-1.630	-1.179	
>=4	-1.120		0.092	-1.300	-0.939	
=5	-0.502		0.075	-0.650	-0.355	
q12						
Discrim	5.166	0.699	7.390	0.000	3.796	6.535
Diff						
>=2	-2.162		0.245	-2.642	-1.682	
>=3	-1.660		0.140	-1.935	-1.385	
>=4	-1.273		0.097	-1.462	-1.084	
=5	-0.672		0.072	-0.813	-0.531	
q13						
Discrim	3.923	0.499	7.860	0.000	2.945	4.902
Diff						
>=2	-2.382		0.282	-2.934	-1.829	
>=3	-1.455		0.119	-1.690	-1.221	
>=4	-1.391		0.113	-1.612	-1.170	
=5	-0.569		0.075	-0.717	-0.421	
q14						
Discrim	6.813	0.942	7.230	0.000	4.966	8.660
Diff						
>=2	-1.872		0.171	-2.208	-1.536	
>=3	-1.277		0.093	-1.460	-1.095	

>=4	-1.145		0.084	-1.309	-0.981	
=5	-0.615		0.069	-0.750	-0.481	
q15						
Discrim	2.650	0.303	8.730	0.000	2.055	3.245
Diff						
>=2	-1.777		0.166	-2.103	-1.451	
>=3	-1.307		0.116	-1.534	-1.080	
>=4	-1.029		0.097	-1.219	-0.840	
=5	-0.229		0.084	-0.393	-0.065	
q16						
Discrim	3.911	0.448	8.720	0.000	3.032	4.790
Diff						
>=2	-1.556		0.129	-1.809	-1.302	
>=3	-0.954		0.083	-1.116	-0.792	
>=4	-0.901		0.080	-1.058	-0.743	
=5	-0.340		0.075	-0.487	-0.192	
q17						
Discrim	4.340	0.522	8.320	0.000	3.317	5.362
Diff						
>=2	-1.682		0.143	-1.961	-1.402	
>=3	-1.276		0.100	-1.472	-1.079	
>=4	-1.021		0.084	-1.185	-0.857	
=5	-0.453		0.073	-0.596	-0.309	
q18						
Discrim	9.227	1.420	6.500	0.000	6.444	12.011
Diff						
>=2	-1.828		0.159	-2.140	-1.515	
>=3	-1.352		0.095	-1.538	-1.165	
>=4	-1.149		0.082	-1.309	-0.989	
=5	-0.647		0.066	-0.777	-0.517	
q19						
Discrim	3.039	0.363	8.370	0.000	2.327	3.750
Diff						
>=2	-1.909		0.183	-2.268	-1.550	
>=3	-1.576		0.139	-1.847	-1.304	
>=4	-1.203		0.104	-1.407	-0.998	
=5	-0.381		0.079	-0.536	-0.225	
q20						
Discrim	6.931	1.011	6.860	0.000	4.949	8.912
Diff						
>=2	-1.787		0.155	-2.092	-1.483	
>=3	-1.479		0.110	-1.693	-1.264	
>=4	-1.318		0.095	-1.505	-1.131	
=5	-0.622		0.068	-0.756	-0.488	

6m post treatment

Fitting fixed-effects model:
Iteration 0: log likelihood = -3944.4477
Iteration 1: log likelihood = -3944.3841
Iteration 2: log likelihood = -3944.3841
Fitting full model:
Iteration 0: log likelihood = -3282.4632
Iteration 1: log likelihood = -2979.953
Iteration 2: log likelihood = -2948.034
Iteration 3: log likelihood = -2941.9055
Iteration 4: log likelihood = -2941.5697
Iteration 5: log likelihood = -2941.5686
Iteration 6: log likelihood = -2941.5686
Graded response model Number of obs = 149
Log likelihood = -2941.5686

	Coef.	Std.Err.	z	P>z	[95%Conf.	Interval]
q1_2						
Discrim	3.977	0.606	6.560	0.000	2.789	5.165
Diff						
>=2	-1.544		0.190	-1.916	-1.172	
>=3	-0.707		0.111	-0.924	-0.489	
>=4	-0.644		0.109	-0.858	-0.430	
=5	0.175		0.118	-0.056	0.405	
q2_2						
Discrim	3.639	0.528	6.890	0.000	2.604	4.675
Diff						
>=2	-1.540		0.190	-1.913	-1.167	
>=3	-0.694		0.113	-0.915	-0.473	
>=4	-0.450		0.106	-0.658	-0.242	
=5	0.260		0.123	0.018	0.502	
q3_2						
Discrim	2.632	0.393	6.690	0.000	1.861	3.403
Diff						
>=2	-1.936		0.272	-2.469	-1.403	
>=3	-0.879		0.134	-1.143	-0.616	
>=4	-0.631		0.122	-0.869	-0.393	
=5	0.235		0.133	-0.027	0.496	
q4_2						
Discrim	2.215	0.345	6.420	0.000	1.539	2.890
Diff						
>=2	-2.325		0.386	-3.082	-1.568	
>=3	-1.335		0.188	-1.705	-0.966	
>=4	-0.997		0.151	-1.293	-0.700	
=5	0.306		0.145	0.022	0.590	
q5_2						
Discrim	1.759	0.274	6.420	0.000	1.222	2.297
Diff						
>=2	-1.789		0.277	-2.331	-1.246	
>=3	-0.477		0.138	-0.748	-0.205	
>=4	-0.274		0.136	-0.540	-0.009	
=5	0.718		0.182	0.361	1.075	
q6_2						
Discrim	3.398	0.490	6.930	0.000	2.437	4.358
Diff						
>=2	-1.508		0.185	-1.870	-1.146	
>=3	-0.750		0.117	-0.978	-0.521	
>=4	-0.513		0.109	-0.726	-0.300	
=5	0.342		0.128	0.090	0.594	
q7_2						

Discrim	3.199	0.456	7.010	0.000	2.304	4.093
Diff						
>=2	-1.697		0.216	-2.120	-1.274	
>=3	-0.654		0.115	-0.879	-0.428	
>=4	-0.428		0.110	-0.644	-0.212	
=5	0.566		0.139	0.294	0.837	
q8_2						
Discrim	3.506	0.559	6.270	0.000	2.410	4.601
Diff						
>=2	-1.605		0.210	-2.017	-1.193	
>=3	-1.120		0.142	-1.398	-0.842	
>=4	-0.948		0.127	-1.196	-0.699	
=5	-0.048		0.115	-0.273	0.176	
q9 2						
Discrim	2.296	0.405	5.660	0.000	1.501	3.090
Diff						
>=2	-2.481		0.451	-3.366	-1.596	
>=3	-1.770		0.265	-2.289	-1.251	
>=4	-1.237		0.181	-1.591	-0.883	
=5	-0.275		0.125	-0.520	-0.030	
q10_2						
Discrim	1.777	0.261	6.810	0.000	1.266	2.288
Diff						
>=2	-0.808		0.157	-1.115	-0.500	
>=3	0.679		0.170	0.345	1.013	
>=4	0.811		0.180	0.459	1.163	
=5	1.329		0.217	0.904	1.753	
q11 2						
Discrim	2.651	0.390	6.800	0.000	1.888	3.415
Diff						
>=2	-2.269		0.368	-2.991	-1.547	
>=3	-0.913		0.136	-1.181	-0.646	
>=4	-0.467		0.116	-0.696	-0.239	
=5	0.398		0.139	0.125	0.671	
q12 2						
Discrim	3.274	0.534	6.130	0.000	2.228	4.320
Diff						
>=2	-2.029		0.315	-2.646	-1.411	
>=3	-1.493		0.195	-1.874	-1.112	
>=4	-0.985		0.135	-1.249	-0.721	
=5	-0.124		0.115	-0.349	0.101	
q13_2						
Discrim	1.618	0.279	5.790	0.000	1.071	2.166
Diff						
>=2	-3.577		0.812	-5.170	-1.985	
>=3	-1.025		0.186	-1.390	-0.659	
>=4	-0.850		0.169	-1.181	-0.519	
=5	0.160		0.158	-0.150	0.470	
q14 2						
Discrim	5.344	0.857	6.230	0.000	3.664	7.024
Diff						
>=2	-1.513		0.174	-1.853	-1.173	
>=3	-0.784		0.108	-0.996	-0.572	
>=4	-0.588		0.101	-0.786	-0.389	
=5	-0.049		0.106	-0.256	0.158	
q15 2						
Discrim	2.437	0.370	6.590	0.000	1.712	3.163
Diff						
>=2	-1.496		0.200	-1.887	-1.105	
>=3	-0.753		0.128	-1.004	-0.502	

>=4	-0.545		0.120	-0.781	-0.309	
=5	0.321		0.141	0.045	0.597	
q16_2						
Discrim	2.644	0.373	7.100	0.000	1.914	3.375
Diff						
>=2	-1.609		0.211	-2.022	-1.197	
>=3	-0.444		0.117	-0.673	-0.214	
>=4	-0.322		0.116	-0.550	-0.094	
=5	0.726		0.153	0.427	1.025	
q17_2						
Discrim	2.766	0.411	6.730	0.000	1.960	3.571
Diff						
>=2	-1.887		0.260	-2.397	-1.377	
>=3	-1.034		0.142	-1.313	-0.755	
>=4	-0.821		0.128	-1.073	-0.570	
=5	0.418		0.137	0.149	0.686	
q18_2						
Discrim	5.017	0.824	6.090	0.000	3.403	6.631
Diff						
>=2	-1.932		0.279	-2.479	-1.385	
>=3	-1.031		0.125	-1.276	-0.787	
>=4	-0.838		0.111	-1.056	-0.620	
=5	-0.055		0.107	-0.265	0.154	
q19_2						
Discrim	1.555	0.257	6.060	0.000	1.052	2.059
Diff						
>=2	-3.228		0.644	-4.491	-1.965	
>=3	-1.331		0.222	-1.766	-0.897	
>=4	-0.971		0.180	-1.323	-0.619	
=5	0.605		0.186	0.241	0.969	
q20_2						
Discrim	6.278	1.121	5.600	0.000	4.082	8.475
Diff						
>=2	-1.524		0.173	-1.864	-1.185	
>=3	-1.089		0.124	-1.332	-0.845	
>=4	-0.847		0.108	-1.059	-0.635	
=5	-0.147		0.101	-0.344	0.050	

Appendix J: Paper for publication

ABSTRACT

The MD Anderson Dysphagia Inventory (MDADI) is a widely used patient reported outcome measure (PROM) which assesses dysphagia-related quality of life (QoL) in head and neck cancer (HNC). Despite its common use in HNC research and clinical practice, few of its psychometric properties have been reappraised since its initial inception. The aim of this study was to perform a survey-based qualitative analysis of UK HNC clinicians' perceptions of the content validity of the MDADI, evaluating it across the parameters of Relevance, Comprehensiveness and Comprehensibility as per the COSMIN guideline for PROM assessment (Mokkink et al., 2019). Five themes were identified: 1. A PROM assessing dysphagia-related QoL has potential value in prompting discussion, 2. MDADI items lack clarity of definition of the terms 'swallowing', 'eating' and 'dysphagia', 3. The MDADI is perceived to be overly negative in tone including items that service users may find distressing or disempowering, 4. Items in the tool are exclusory to specific subgroups of patients, such as those who are Nil By Mouth or socially isolated, 5. Modifications to the MDADI were suggested and encouraged to make it more useful and patient-centred. In summary, this study indicates that MDADI's content validity could be improved across all three COSMIN parameters. Further re-evaluation of the content validity of the MDADI is warranted, with potential future amendment of items being indicated if the results of this study are corroborated in subsequent research.

INTRODUCTION

Difficulties with eating, drinking and swallowing are commonly reported by patients as one of the most impactful outcomes of their head and neck cancer (HNC) and its treatment (Mendez et al., 2020), and people with HNC-related dysphagia describe a complex interaction between it and other social, emotional and physical aspects of their lives (Dawson et al., 2019).Recent international guidance on HNC practice emphasises the importance of assessing, monitoring and managing the psychosocial impact of dysphagia (Verdonck-de Leeuw et al., 2022, Baijens et al., 2021).The impact of dysphagia on peoples' lives can be measured by assessing dysphagia-related quality of life (QoL), which is the patient's perception of the impact of swallowing difficulties across social, functional and psychological domains (Speyer et al., 2014).

Patient reported outcome measures (PROMs) are a commonly used way of conceptualising and quantifying health related QoL and are an important adjunct to clinician-led assessments (Ferrans, 2007). Silveira and colleagues describe PROM assessment in HNC as an 'essential' part of clinical practice. Several PROMs exist that can be used to assess the impact of difficulties with eating, drinking and swallowing on QoL. However, the only tool dedicated to assessing dysphagia-related QoL specifically in people with HNC is the MD Anderson Dysphagia Inventory (MDADI) (Chen et al., 2001).

The MDADI is one of the most frequently used dysphagia outcome assessment tools in HNC research practice internationally (Ojo et al., 2012) and is often used as a main outcome tool in multicentre trials (Hutcheson et al., 2016, Castellano and Sharma, 2019). In recent years it has been the primary or secondary endpoint in large multicentre trials such as De-ESCALaTE, PATHOS, CompARE, DARS and PRO-ACTIVE (Owadally et al., 2015, Petkar et al., 2016, Martino et al., 2021, Mehanna et al., 2017, Mehanna et al., 2019). The MDADI is also often used as a 'gold standard' in the validation of other dysphagia assessment tools for use with people with HNC, such as the DIGEST (Hutcheson et al., 2017) and the Sydney Swallow Questionnaire (Dwivedi et al., 2010).

The MDADI is a self-administered PROM which quantifies swallowing related QoL. It was originally validated on a cross-sectional sample of 100 English-speaking adult patients with HNC and dysphagia in the USA in the late 1990s (Chen et al., 2001). The tool consists of 20 items each rated on a 5-point Likert scale. Scoring the tool produces a global score (MDADI – G), scored from the first item ("my swallowing impacts my day-to-day life") and a composite score (MDADI –C) of the remaining 19 items. MDADI-G and MDADI-C scores range from 20 (low QoL) to 100 (high QoL).

High demands are placed on outcome measurement tools. They must be clinically meaningful and have strong psychometric properties ensuring that the data they capture is reliable and valid (Speyer et al., 2022).Several tools exist that facilitate the assessment of the psychometric properties of PROMs, however Lorente and Villadrich 's recent meta-review showed the most comprehensive and widely used tool to be the COSMIN tool (Mokkink et al., 2019).

The 'content validity' of a tool refers to whether it covers all of the important and relevant aspects of the subject under investigation (Connell et al., 2018), and can be established by asking patients and clinicians about the comprehensiveness, relevance and comprehensibility of the items in the tool (Terwee et al., 2018, Streiner et al., 2015). COSMIN suggests that this is the most important of all psychometric parameters (Terwee et al., 2018), with poor content validity having potential to impact all other aspects of tool psychometrics.

Since the MDADI's inception more than 20 years ago, studies have appraised its psychometric properties according to data provided in Chen's origin paper (Ojo et al., 2012, Timmerman et al., 2014, Patel et al., 2017), however surprisingly few aspects of the MDADI have been re-evaluated through original research given the tool's high standing within HNC practice. To date the concurrent validity (Pedersen et al., 2016, Khan et al., 2015), interpretability (Hutcheson et al., 2016) and construct validity (Lin et al., 2022) of the MDADI have been explored, however no research exists further examining the content validity of the tool.

This study therefore constitutes the first de novo exploration of the content validity of the MDADI since its inception. The aim of this study was to carry out an investigation of UK clinicians' perceptions of the content validity of the MDADI with the aim of providing novel information on this important aspect of the tool and potentially identifying areas for future investigation or development.

METHODS

This study took the form of a qualitative exploration of UK Speech & Language Therapists' (SLTs) perceptions of the content validity of the MDADI, via data generated by responses to an online questionnaire. Data presented and discussed in this study formed part of a larger Pragmatic mixed methods study which investigated the content validity, clinical utility and other psychometric properties of the MDADI carried out by the author.

Participants and procedure

Data were collected from UK-based SLTs via an online questionnaire tool. Participants were accessed via professional networks and social media, with the aim of convenience sampling progressing to snowball recruitment. Participants did not need to have experience of using the MDADI tool; however clinical experience of working with people with HNC was essential for them to be able to evaluate tool content validity with respect to this patient group.

A questionnaire was generated and piloted with both a group of SLTs working out with the field of HNC, and with the study Patient and Public Involvement (PPI) group. Greenhalgh (2014) suggests the goal of questionnaire design is clarity and ease of completion, therefore, length, layout, question wording and content were considered (Boynton, 2004, Kelley et al., 2003, Jones et al., 2013). To ensure participant informed consent for the questionnaire, a participant information sheet (PIS) was provided at time of recruitment. UK research approval was granted [IRAS REC no. 20/WM/0319] and R&D permission to carry out the study was sought and granted at both a local and national level (HRA, 2020, NICE, 2018).

In addition to a small number of demographic questions, questionnaire respondents were asked to comment on each MDADI item in detail. Questions were open-ended in nature to allow collection of qualitative, narrative data, with the response taking the format of a free-text comment box.

Demographic data gathered on questionnaire participants was limited to years of clinical experience, location and MDADI usage patterns to ensure anonymity. In total 31 UK clinicians responded to the questionnaire, representing the UK nations of England (n=19) and Scotland (n=12). The total number of SLTs with clinical experience in HNC in the UK is unknown. Respondents had wide range of years of clinical experience working with people with HNC (1-34 years, mean 12.42). Of the thirty-one questionnaire respondents, twenty were using the MDADI in their clinical practice at time of completion. The questionnaire was open to responses from December 2020 – February 2021 inclusive.

Data analysis

Questionnaire data were downloaded from the online questionnaire platform and the free-text narrative data underwent a process of reflexive thematic analysis (TA) using Braun and Clarke (2021)'s 6-phase guide to performing this technique. TA provides a theoretically flexible, accessible approach to qualitative data analysis (Braun and Clarke, 2006). The process of TA involves becoming familiar with the data before generating initial codes and searching for themes, which are ultimately named and defined; it is a process of disassembling, reassembling, interpreting and drawing conclusions on the data (Castleberry and Nolen, 2018).

RESULTS

Thematic analysis of the questionnaire responses provided a rich account of SLTs' impressions of the content validity of the MDADI. Five themes which had relevance to content validity were identified, as summarised in Table 1.

	Theme name	**Characteristics**
1	"a conversation starter" – The MDADI can be helpful in clinical practice	How the MDADI is useful in clinical practice
2	The bigger picture of eating and drinking	Lack of focus in the MDADI about whether it is assessing swallowing, eating and drinking, other issues, or everything!
3	"not user-friendly"	How the MDADI is perceived to be negative, emotive and non-patient-centered
4	Excluded groups	Patient subgroups excluded by MDADI item wording or content
5	"not quite where we need it to be" – suggestions for change	Practical suggestions for changes and improvements to the tool that would make it more useful and patient-centered

Table 1: Summary of themes generated from inductive Thematic analysis

Theme 1: “A conversation starter”

Despite issues with the tool identified in other themes, respondents felt using the MDADI still added overall value to therapist-patient consultations. Clinicians acknowledged “It's a useful QOL tool where few others exist. It was devised with HNC patients in mind” (Respondent 13). There was a common thread that the MDADI acted as a “conversation starter” (Respondent 4) when used during clinical sessions. Respondent 5 described how the MDADI “allows you to discuss a number of different topics that may not come up during informal discussion or a patient may not consider it to be an important part of their care”. Respondent 30 highlighted that the MDADI forms part of a thorough clinical evaluation, complementing other assessment tools: “It provides information which assists clinical decision-making alongside other tools/measures”.

Theme 2: The bigger picture of eating and drinking

The MDADI was designed to assess ‘dysphagia-specific’ impact on patients’ QoL, however the questionnaire responses richly spoke to the fact that the tool frequently mixes swallowing impairment with other, more overarching aspects of eating and drinking, to the point that it is often not clear which aspect is being assessed with respect to impact on QoL. Respondent 19 described this as “Poor wording-eating habits doesn't equate to swallowing”. Likewise Respondent 7 identified ambiguity in referring to ‘eating’ rather than swallowing: “The phrase 'eating habits' can be interpreted in many ways, and not necessarily relevant to swallowing. For example, some of my bariatric patients, or patients that wish to lose weight, will comment on this.”

A common theme throughout the dataset was the scope for issues other than oropharyngeal dysphagia to affect patients’ responses to items. Respondent 4 described it thus: “Most patients say my swallowing doesn't limit me- my issue is the pain/ saliva/ appetite/RT [radiotherapy] side effects. The impact of dental extractions, which are a common part of pre-HNC treatment workup, on MDADI responses, were highlighted by respondents: “If patients have had recent dental extractions this can influence their responses - sometimes need to guide them to think about oral intake prior to dental extractions” (Respondent 17).

Theme 3: “Not user-friendly”

Respondents found the tone of the MDADI to be overly negative and to lack patient-centredness, commenting how it often cast patients in a passive or disempowered position:

> *“Some aspects are negative and focused on the potential views of other people, this can be upsetting for patients, it also focuses on swallowing difficulty and not …. adaptations a patient may be effectively making” (Respondent 10).*

The use of the word swallowing ‘*problem’* throughout the MDADI was noted, with concerns that this negative framing had potential to “skew the question” (Respondent 19). Respondents commented on other examples of negative or insensitive wording used in MDADI items, some feeling it was unsuitable to the extent that they stopped using the tool in their clinical practice:

> *“If [people] are really struggling then the questions are insensitive - reminds [people] of everything they are missing and likely to damage clinician-client bond, as this adds insult to injury. Largely why I stopped using it - didn’t add to my clinical care and only those with mild-mod issues seemed well enough in themselves to complete” (Respondent 2).*

Items 5 (“I do not feel self-conscious when I eat“) and 15 (“I feel free to go out to eat with my friends, neighbours, relatives”) in the MDADI caused consternation amongst respondents.

They were commonly described as “the trick double-negative questions” (Respondent 29) that "catch patients out” (Respondent 19). This is because these are ‘reverse scored’ items, as explained by Respondent 17:

> *“I ask patients to read particularly carefully as they can often say disagree when they mean agree and vice versa due to trend in answers in the rest in the questionnaires i.e., in most questions "agree" indicates an issue whereas in this example "agree" is a positive thing”.*

Throughout the MDADI tool, respondents highlighted items that were felt to be ambiguous, requiring explanation, therefore being open to interpretation potentially affecting the validity of responses: “Some questions are worded in a confusing way and the answers then need to be checked” (Respondent 7). Many respondents described how they had to ‘step in’ to help patients complete the tool due to ambiguously worded items: “the wording of some questions is confusing therefore requires clarification from SLT” (Respondent 18). In addition, often it was not just concern about patients misunderstanding or being confused by items, but also SLTs themselves: “Patients don't understand the question and nor do I” (Respondent 19).

Theme 4: Excluded groups

Respondents identified subgroups of patients who would not be able to answer some, or any, of the MDADI items due to their health or social status at time of assessment.

Items that referred to ‘eating out’ were identified as potentially exclusory to patients of lower socioeconomic status who were not able to afford to do so, such as item 8 (“I do not go out because of my swallowing problem”) which “does not often highlight difficulties in our clinical caseload who often cannot afford to eat out” (Respondent 15).

Likewise, item 9 (“my swallowing difficulty has caused me to lose income”) was felt to exclude patients who either couldn’t work or who had retired. This was summarised by Respondent 12 who noted:

> *“I have had a number of respondents omitting this item or annotating it to say that they are not in employment. The predominance of over 60s or 65s in the HNC population tends to make this item slightly less relevant.”*

Several items were also highlighted as being exclusory to patients who lived alone or were socially isolated, with the items’ focus on friends, family and social interactions. These items were described as “open to misinterpretation/non-representative answers as so many of our patients live alone/ do all the cooking themselves/don't have anyone who cooks for them” (Respondent 17).

Significant issues throughout the tool were highlighted in terms of using the tool as a baseline measurement with patients prior to their HNC treatment, often before they have any symptoms of swallowing difficulty. Most items presume a swallowing problem and therefore clinicians have frequently experienced non-dysphagic patients not knowing how to answer an item, potentially skewing or invalidating their results; in addition to causing anxiety and trepidation about what might be on the horizon in terms of future treatment side effects. Respondent 17 summarised this concern: “It can also concern people who do not have swallowing difficulties, as they worry that the questions are an indication of what they will face in future e.g., they might not be able to eat out, enjoy a meal with friends, be embarrassed about their eating etc”.

Finally, patients whose dysphagia is so severe that it has been recommended they be ‘nil by mouth’ (NBM) were highlighted as a group for whom responding to the MDADI would be extremely challenging. This cohort of patients must rely on non-oral, enteral nutrition to meet

their nutritional requirements. A strong theme amongst respondents was that the MDADI was not appropriate to attempt in this situation and could not be ethically used with these patients: “I wouldn't use this with someone who is NBM as a result of cancer/treatment as I feel it would be pretty insensitive” (Respondent 17). This then means that the group who potentially have the greatest reduction in dysphagia related QoL, due to not being able to eat or drink at all, are excluded from having this impact measured, as summarised by Respondent 20:

> *“I don't use it with patients who are nil by mouth as I feel it's unfair - they can't answer many of the items and it is upsetting. This is a big issue though as they may be that patient group whose quality of life is most impacted by their dysphagia!!”.*

Theme 5: “Not quite where we need it to be” – suggestions for change

Throughout the survey, respondents made comments on how the MDADI could be changed, with practical suggestions for rewording, modification, elision or removal of items. Several items were described as “irrelevant” (Respondent 20), in addition many items were repetitious: “several questions are very similar and would be good if they could be reduced” (Respondent 8). The consensus was that shortening the tool by removing some similar items would make it more user-friendly, producing more valid data, as illustrated by Respondent 18: “I feel many questions are similar and could be removed as the questionnaire is extremely long and patients often rush through it. If it was shorter I feel patients would take more time to consider their answers”. Respondents also remarked on the Likert scale response modality, voicing concerns that the descriptors were inappropriate: “I don’t think anyone has ‘no opinion’ of the kind of questions that are being asked” (Respondent 14) and suggesting this response format potentially lead to an anchor effect: “I find patients don't vary their answers between e.g. strongly agree versus agree and will stick with the same whichever they go with, right through” (Respondent 6).

Respondents also made suggestions for topics or items that they felt were missing from the MDADI and that would make relevant additions as issues which had significant potential to impact patients’ QoL: “Think it would be good to ask specifically about safety of swallowing and efficiency. Any food avoidance. Impact of pain. Impact of taste changes. Impact of xerostomia. Impact of dental extractions/dental issues.” (Respondent 13).

An appetite for development and improvement of the MDADI was evident in the data: “the MDADI could be adapted/updated to better reflect patient experiences particularly patients having treatment for HNC” (Respondent 10). Respondent 20 summarised the MDADI thus: “Great potential but not quite where we need it to be to truly represent the impact of head and neck associated dysphagia”.

DISCUSSION

Despite the MDADI’s high profile within HNC research and practice, its content validity has been minimally appraised to date. This study provides the first qualitative data on the content validity of the MDADI from a clinician perspective, and highlights issues with all three aspects of content validity as defined by COSMIN: Relevance, Comprehensiveness and Comprehensibility(Prinsen et al., 2018, Terwee et al., 2018).

Relevance

COSMIN defines ‘relevance’ as whether tool items are relevant for the construct, context and target population of interest.

The HNC dysphagia literature highlights the existence of significant dysphagia-related QoL impact post-treatment (Patterson et al., 2015, Nund et al., 2014a) and therefore it is essential to have assessment tools that can assess this and provide meaningful data. Theme 1 indicates that a tool that facilitates clinicians instigating emotive and difficult conversations with patients about key aspects of cancer survivorship has huge clinical relevance and use; however, the items need to be relevant, and a major concern raised in this study is that the MDADI potentially lacks relevance in terms of its specific item content with respect to significant subgroups of people with HNC.

This is evidenced by Theme 4. A tool that aims to assess the impact of dysphagia on QoL should be relevant to all subgroups of patients with dysphagia. However, data from this study highlighted that clinicians feel the MDADI is exclusory to several patient groups, most notably patients with NBM status who are dependent on tube-feeding due to the severity of their dysphagia.

Patients with the most severe dysphagia also need assessment of their dysphagia related QoL so their support and survivorship needs can be understood and addressed; a tool capable of assessing this is a necessity for clinical practice. There is a bioethical concern with PROMs that fail to meet the needs of more vulnerable subgroups, with PROMs in other clinical areas having been shown to be more challenging for example for older people, or those with more severe symptoms (Hagell et al., 2009).

The subgroups identified in this study as being potentially excluded from validly completing the MDADI constitute important, common subsections of the HNC patient group, and the fact that the MDADI is inaccessible to them means a significant proportion of HNC patients are potentially excluded from measurement of their dysphagia-related QoL with this tool.

Theme 3 also relates to Relevance: when used with non-dysphagic patients at a pre-treatment point, MDADI items may not be relevant to their current circumstances, with subsequent potential for the tool to cause distress and scores to lack validity. This study also presents evidence that the tone of the MDADI is unhelpfully negative and therefore potentially unnecessarily distressing for patients to complete. This finding concurs with Greenhalgh and Dalkin 's assertion that there is potential for the act of completing a PROM to cause patient distress: items in the tool may remind patients of the ways their life has changed and the impact that their symptoms have on many aspects of their life. This finding was also corroborated by Duncan and Murray (2012) in their systematic review of barriers and facilitators to outcome measurement by AHPs which highlighted concerns that PROMs might cause distress.

However, clinicians involved in this study also suggest that a change in tone and rewording of items to be more inclusive, empowering and less negative might go some way to make the patient's experience of completing the MDADI less upsetting. Respondents suggested many practical suggestions for how MDADI items could be reworded, however these would need to be ratified and corroborated in further research involving patients.

Comprehensiveness

The COSMIN criterion of Comprehensiveness is that all key concepts should be included by a tool (Terwee et al., 2018). Whilst Theme 4could also be seen to be relevant here in that some items were identified as having limited relevance to patient subgroups, other themes also spoke to this parameter. Theme 2 highlighted issues with the comprehensiveness of the MDADI. Specifically, the concern that the MDADI is not sufficiently clear about what it is assessing, due to a lack of definition of concept in terms of dysphagia versus the wider process of eating and drinking. This study indicates that clinicians find the wording of the

MDADI inconsistent, interchanging the terms ‘swallowing’ and ‘eating’ without obvious rationale. It could be argued that the MDADI’s scope is *too* comprehensive and needs to be reined in to focus specifically on dysphagia, or conversely not comprehensive enough, in that it does not give explicit reference to other swallowing-adjacent issues such as xerostomia or reduced dentition, which may be inseparable from patients’ eating, drinking and swallowing experience (Bressan et al., 2017). Data in Theme 5 also supported this assertion with clinicians suggesting inclusion of additional swallowing-adjacent symptoms would improve tool validity.

Inconsistencies in definition and delineation of what ‘dysphagia’ as an entity includes are evident in the published literature. Swallow physiology adjacent issues common in HNC such as dysgeusia and xerostomia may be grouped under the term ‘dysphagia’ as per by Nund etal in their qualitative study investigating the lived experience of post-HNC dysphagia. Conversely, Ganzer etal use the term ‘eating experience’ in their literature review to encompass the complex, multifaceted physical, social and emotional impacts on eating drinking and swallowing post HNC, incorporating ‘dysphagia’ as one element of the eating experience. Likewise the recently developed Head and Neck Cancer Survivors’ Assessment of Mealtimes tool (Chan et al., 2019) refers to ‘mealtime experience’ rather than swallowing alone, and qualitative, co-produced research carried out with people with HNC suggests ‘altered eating’ as a more appropriate umbrella term (Burges Watson et al., 2018). It could be argued that the separation of swallow physiology from the bigger picture of eating and drinking is an artificial distinction; however, if this is the case terminology and focus should remain consistent within a tool, and the inconsistency in terminology in the MDADI at the very least needs to be addressed. This study has demonstrated that this inconsistency within the MDADI confuses users, disappoints clinicians and negatively impacts on its content validity.

Comprehensibility

The COSMIN criteria for comprehensibility are that tool instructions, items and response options should be understood by users. Themes 3 and 5 speak to this parameter.

In terms of being able to understand the MDADI, this study evidences clinicians’ concerns around ambiguity of wording in the tool, and that the literacy level of the MDADI may be too complex for many people with HNC, thus affecting their ability to complete it. This finding provides qualitative validation to Zraick etal ’s previous quantitative analysis of the ‘readability’ of the MDADI, which found the MDADI to be the ‘most difficult to read’ of all of the swallowing-related PROMs examined. Clinicians also voiced concerns that the Likert response modality of the tool had potential to impact on response validity. Clinicians highlighted their experience of ‘anchor effect’ in patients’ responses to the tool. ‘Anchor effect’ for Likert scales, where respondents are less likely to endorse the ‘extreme’ ends of the scale, thus affecting score validity, is a well-documented phenomenon (Bishop and Herron, 2015). Respondents also indicated that the MDADI is overlong, potentially containing redundant items; this is substantiated by Lin et al. (2022)’s recent factor analysis of the MDADI which resulted in item reduction, a decrease in tool length, and the formation of a 5-item ‘miniDADI’,

In summary, this study presents data that show UK clinicians have concerns about the content validity of the MDADI across all three COSMIN criteria of relevance, comprehensiveness and comprehensibility. This therefore supports a rating of the content validity of the MDADI as ‘insufficient’ using the COSMIN assessment criteria.

Study limitations

Lack of patient data

The depth of data generated from the clinicians produced relevant qualitative data on PROM content validity by including clinicians in tool evaluation. However, patient involvement in PROM development is strongly recommended in the literature (Addario et al., 2020) and the results of this study need to be considered with caution, as clinicians' proxy reporting of patients' experiences is not always reliable (Dunlop et al., 2022).

Generalisability

Although circulated around networks accessible across the UK, questionnaire responses were recorded from 31 SLTs practising in England and Scotland only. This number of respondents is comparable to other practice-related survey-based research carried out with UK HNC SLTs (Roe et al., 2012). As the MDADI is an internationally used tool, and eating, drinking, swallowing and dysphagia are very culturally sensitive, it may be however that the results of this study are not generalisable to other locations. Further triangulation with data from other countries is required to corroborate the data presented here.

Reflexivity and potential for bias

The author is a practising HNC clinician who has used the MDADI with patients for many years. It could be argued therefore that she does not have a starting point of equipoise about the MDADI. Malterud (2001) describes reflexivity as 'the knower's mirror'. During the analysis the author took steps to incorporate reflexivity and monitor for bias, using a reflective diary as a 'mirror' to monitor and manage the personal thoughts and reactions arising in response to the analysis process.

CONCLUSION

HNC dysphagia clinicians and researchers need to think long and hard about how and what we assess, and why: are the tools we commonly use fit for purpose, and do they generate meaningful data that will have an impact on our management of patients? We cannot afford to be complacent about the quality of the tools that currently form part of 'established' HNC practice. The results of this study have shown that the content validity of 'classic' assessment tools cannot be assumed. We must ask 'second generation questions' of these tools (Kroenke et al., 2015) and reflect on and interrogate the tools we use in both clinical and research practice to ensure they are fit for purpose and add value to our patients' management.

Since its inception more than twenty years ago, the MDADI has been subject to surprisingly little scrutiny of many of its psychometric properties, even though it is one of the most well-used tools in HNC clinical and research practice. Most evaluations of the tool purely re-present the development validation data without making any de novo evaluations of the MDADI's strengths and weaknesses. This study has highlighted issues with content validity from a clinician perspective that are salient and merit further investigation with service users. If they were to be corroborated this would constitute a significant challenge to the content validity of the MDADI, and thus the validity and reliability of MDADI data generated in clinical and research practice, indicating a pressing need for further development and amendment of the tool.

Possible future directions for research include patient-led assessment of MDADI content validity, and further exploration of clinicians' opinions with a larger and more geographically dispersed group to test if the results presented in this study are replicated.

This study confirms that a tool specific to assessing dysphagia related QoL has great relevance and use for clinical practice. The MDADI is unique in its HNC-specific slant on

dysphagia-related QoL assessment and has an established place in the world of HNC practice. However, this study suggests there is scope for further assessment and potentially amendment of the MDADI, to strengthen its content validity and maintain its position within the HNC firmament.