UNDERSTANDING THE FUNCTIONAL IMPLICATIONS OF COLORECTAL CANCER

UNDERSTANDING THE EFFECT OF COLORECTAL CANCER ON THE ABILITY TO PERFORM USUAL ACTIVITIES

By Vanessa Sum Kee Fan, H.B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree Master of Science

McMaster University Master of Science (2021) Hamilton, Ontario (Rehabilitation Science)

TITLE: Understanding the effect of colorectal cancer on the ability to perform usual activities

AUTHOR: Vanessa Sum Kee Fan, H.B.Sc. (University of Toronto)

SUPERVISOR: Dr. Jackie Bosch

COMMITTEE: Dr. Julie Richardson, Dr. Lauren E. Griffith

NUMBER OF PAGES: xi, 146

Lay abstract

More people are living with colorectal cancer (CRC), but may have problems performing their daily activities (i.e. functional problems) due to cancer-associated impairments. However, we do not understand the extent of these impairments and functional problems. We used a sample of people with newly-diagnosed gastrointestinal cancer (CRC being the most common type) to understand their type and extent of functional problems. People were found to participate less in functional activities and particularly have more difficulty walking after a cancer diagnosis. Fatigue is common among those with CRC and may primarily cause functional problems. However, it is not commonly measured, and it is unclear how to best measure fatigue among them. Therefore, we reviewed key qualities of 16 fatigue measures in a similar population (inflammatory bowel disease, IBD) and recommended the Functional Assessment of Cancer Therapy Instrument-Fatigue and the IBD-Fatigue scale (English) as the most promising measures for those with CRC.

Abstract

Introduction

Colorectal cancer (CRC) survivors may experience functional deficits due to cancer-associated impairments. However, we do not understand their type and extent of functional deficits and how we could measure the associated cause of functional deficits, such as fatigue. As the survival of CRC survivors improves, the burden of living with functional deficits can be high.

Purpose

My research program aims to understand (1) the functional changes and deficits that CRC survivors experience and (2) how to best measure fatigue in this population.

Methods

To address the first aim, we used the data from the International Study of the Risk Factors for Gastrointestinal Bleeding and Cardiovascular Events after Gastrointestinal Bleeding to examine individuals' functional abilities within 1 year of gastrointestinal cancer diagnosis (CRC being the most prevalent type).

For the second aim, we conducted a systematic review on fatigue measures in adults with inflammatory bowel disease (IBD) because the causes, severity, and impact of IBD and CRC-related fatigue might be similar. We identified fatigue measures in the IBD population, appraised their psychometric properties, and recommended the most psychometrically robust and feasible measures for clinical and research use, indicating the optimal measures for CRC survivors.

Results

After gastrointestinal cancer diagnosis, the majority (~70%) performed fewer functional tasks, mostly in the instrumental activities of daily living; and about 44% had more difficulty walking. Our review identified 16 measures, reviewed the content and psychometric properties, and recommended the Functional Assessment of Cancer Therapy Instrument-Fatigue and the IBD-Fatigue scale for research and clinical use in IBD and CRC populations.

Conclusion

We provided a novel understanding of the functional deficits that CRC survivors experience and recommended the optimal measures for assessing CRC-related fatigue. As CRC survivors commonly experience fatigue, fatigue should be measured to understand its role in the functional abilities of CRC survivors.

Acknowledgements

I would like to thank my supervisor Dr. Jackie Bosch, who has been helping me pursue what I am passionate about and provided guidance to produce high quality work. Her passion for research, content knowledge in occupational therapy and her perseverance has inspired me, and she is a role model for being a clinician-scientist. She has pushed me to go out of my comfort zone and grows academically. Her ongoing support helped me persevere and continued to do well.

I would also like to extend my gratitude to my committee members, Dr. Julie Richardson and Dr. Lauren Griffith. Dr. Julie Richardson provided me precious advice at key times during my academic journey and helped me in my personal and professional life. Her encouragement and trust in me during challenging times help me stay focused and build a direction for my work. Dr. Lauren Griffith rounded out the committee with her valuable expertise, by providing suggestions for developing my research design, carrying out statistical analysis and advancing my research work. I would sincerely like to thank you both for being understanding and supportive mentors. I have gained a valuable learning experience from this vastly knowledgeable and experienced committee.

I would like to thank the INTERBLEED team (Dr. Jackie Bosch, Dr. John Eikelboom, Leah Hamilton, Jacqueline Wilkinson) for providing me with access to their expertise and data. I would not be able to complete the thesis without their support. Also, I would like to thank Dr. Kristin Campbell for her encouragement and expertise in cancer rehabilitation.

I am very grateful for the life-long friendship that I have made during this journey. A special big thank you to Sheila Addanki and Sandra Abdel Malek for celebrating the wins with me and lifting me up during the lows. I would never forget all the surprise party, the heart-warming conversation, and the laughter that we had. The life-long friendship is definitely one of the things that I cherished the most.

Finally, I would like to thank my parents, my brothers, and friends who care and support me unconditionally and have fun with me during my most stressful time. I am grateful for this learning experience at the McMaster University that helped me grow both personally and academically.

Table of Contents

List of Tables, Figures and Supplemental information

Chapter 2

Tables

Figures

Supplemental file

Supplemental Tables

Supplemental Figures

Chapter 3

Tables

Figures

Supplementary Data Content

List of Abbreviations and Symbols

CRC: Colorectal cancer
INTERBLEED: The International Study of the Risk Factors for Gastrointestinal Bleeding and Cardiovascular Events after GI Bleeding
GI: Gastrointestinal
IBD: Inflammatory bowel disease
SAGEA: The Standard Assessment of Global Everyday Activities
CVD: Cardiovascular disease
cADLs: cognitive activities of daily living
iADLs: instrumental activities of daily living
bADLs: basic activities of daily living
ICC: Intraclass correlation
IQR: Interquartile Range
SD: Standard deviation
PROMs: Patient-reported outcome measures
PROSPERO: Prospective Register of Systematic Reviews
AMED: Allied and Complementary Medicine Database
COSMIN: COnsensus-based Standards for the selection of health status Measurement INstrument guideline
GRADE: Grading of Recommendations Assessment, Development and Evaluation
FSS: Fatigue Severity Scale
D-FIS: Daily Fatigue Impact Scale
FACIT-F: Functional Assessment of Chronic Illness Therapy- Fatigue
IBD-F: Inflammatory Bowel Disease Fatigue Scale
MFI-20: Multidimensional Fatigue Inventory-20
MAF: Multidimensional Assessment of Fatigue
M-FIS: Modified Fatigue Impact Scale
FQ: Fatigue Questionnaire
CD: Crohn's disease
UC: Ulcerative colitis
CDAI: Crohn's Disease Activity Index
SCCAI: Simple Clinical Colitis Activity Index
IBDQ-36: Inflammatory Bowel Disease Questionnaire
HBI: Harvey-Bradshaw Index
HADS: Hospital Anxiety and Depression Scale
SIBDQ: Short Inflammatory Bowel Disease Questionnaire
SHS: Short Health Scale
HRQoL: Health related quality of life
SF36: 36-Item Short Form Survey

Declaration of Academic Achievement

The details of all author contributions to each manuscript are summarized as follows:

Chapter 2: Vanessa Fan contributed to the manuscript concept and design and the literature search with the support of Dr. Jackie Bosch. Vanessa Fan, Ava Mediphour, Heather O'Grady performed the screening of articles for the systematic review. Vanessa Fan and Ava Mediphour evaluated the methodological quality of the studies and the psychometric properties of measures. Vanessa Fan analysed the results and drafted the manuscript in consultation with Dr. Jackie Bosch. Dr. Jackie Bosch, Dr. Julie Richardson, Dr. Lauren E. Griffith, and Dr. Kristin L. Campbell provided constructive feedback on the study design and the writing of the manuscript.

Chapter 3: Vanessa Fan identified the research question, performed all the statistical analyses, and wrote the manuscript in consultation with Dr. Jackie Bosch. Dr. Julie Richardson and Dr. Lauren E. Griffith provided editorial assistance.

Chapter 1- Introduction and Literature Review

Prevalence and survival rate of colorectal cancer

Colorectal cancer (CRC) is the second most commonly diagnosed cancer in Canada for both men and women and the risk increases with age (Canadian Cancer Society, 2019; Globocan, 2012). The introduction of the national CRC screening program in 2007 led to a significant increase in the number of persons diagnosed at a younger age which resulted in more than 26000 new cases every year and improved the five-year survival rate to about 65% (Canadian Cancer Society, 2019). Half of the CRC diagnoses are made at stage I-II (Bryan et al., 2018). The five-year survival rate of CRC patients with the diagnosis made at stage I-II (localised) is 90%, and for those with stage II-III (regional) is 71%, which is 80% higher than when the diagnosis is made at stage IV (17%) (American Cancer Society, 2017). While survival rates have improved, there are indications that CRC survivors may experience at least short term and likely long term difficulties performing usual activities.

The mechanism for cancer and treatments-associated impairments on function

Fatigue and weakness are common symptoms associated with CRC. Post-treatment CRC survivors (i.e. people who remain alive after cancer diagnosis [Marzorati et al., 2017]) may experience ongoing (1) physical (e.g. pain [Drury et al., 2017; Forsberg & Cedermark, 1996; Rauch et al., 2004; Rutherford et al., 2020], fatigue [Arndt et al., 2006; Drury et al., 2017; Forsberg & Cedermark, 1996; O'Gorman et al., 2018; Rutherford et al., 2020], insomnia [Arndt et al., 2006; Drury et al., 2017; Forsberg & Cedermark, 1996; O'Gorman et al., 2018; Rutherford et al., 2020], dyspnea [Arndt et al., 2006], bowel and sexual dysfunction [Arndt et al., 2006; Dean et al., 2007; Forsberg & Cedermark, 1996; O'Gorman et al., 2018; Rauch et al., 2004;

Rutherford et al., 2020], neuropathy [Den Bakker et al., 2018], stoma [O'Gorman et al., 2018]), (2) cognitive (e.g. chemotherapy-induced cognitive decline [Den Bakker et al., 2018]) and (3) psychological impairments (e.g. anxiety and depression [Abu-Helalah et al., 2014; Braamse et al., 2016; Foster et al., 2009; Ramsey et al., 2002]). These may in part be related to treatment, the disease or both.

In particular, fatigue, dyspnea, insomnia, constipation, and diarrhea are reported as the most severe impairments by CRC survivors one year after cancer diagnosis (Arndt et al., 2004). Fatigue has been associated with decreased activity in those with CRC (Eyl et al., 2020). Dyspnea can decrease function because the reduced aerobic capacity of muscle can cause muscle weakness and influence fatigue. The uncomfortable sensation, fear, and distress of dyspnea can reduce the motivation and the ability to participate in functional tasks (Victorson et al., 2009). Insomnia is found to be associated with limitation in household activities and reduced participation (Spira et al., 2012, 2014) because it can reduce physical capacity to perform activities (Spira et al., 2012) and cause fatigue (Ancoli-Israel et al., 2001). While constipation and diarrhea are not uncommon in the general population, those with CRC demonstrated worse constipation and diarrhea than the norm (Pucciarelli et al., 2010). CRC survivors experience more frequent bowel movements, a constant urge to defecate, and difficulty in emptying the bowel completely (Cancer Research UK, 2018; Pucciarelli et al., 2010; Yin et al., 2018), which negatively affect their physical and social functioning (Pucciarelli et al., 2010).

An estimated 18-35% of CRC survivors receive temporary or permanent ostomy (i.e. opening of the colon allows the passage of stools out of the body) after their surgical removal of the

colorectum, which could further impair their bowel and sexual dysfunction (Schmidt et al., 2005; Sun et al., 2013). CRC survivors with ostomy were found to reduce their working, socialising, leisure and sexual activities, compared to those without ostomy (Sprangers et al., 1995). A qualitative study by Sun et al. also found that CRC survivors have restricted activity participation even 5 years after ostomy formation (Sun et al., 2013). For example, CRC survivors cannot tolerate long car trips because the seatbelts aggravate the ostomy site or they are unable to do activities such as golf because of the limitation in bending associated with the ostomy location (Sun et al., 2013). The presence of an ostomy has also been associated with decreased social participation, potentially because of poorer body image (Sun et al., 2013). There are likely multiple possible causative factors that explain an immediate decline in function in those with CRC, which may persist even after active cancer treatment is completed.

In addition to the effects of cancer and treatment-associated impairments, cancer-related deconditioning and the associated comorbidities can lead to further impairments in CRC survivors. Although genetics plays a key role in the pathogenesis of CRC, its onset is predominantly influenced by the modifiable risk factors (e.g. sedentary lifestyle, calorie rich diets), which are also shared with other metabolic diseases (Type 2 diabetes, obesity) and cardiovascular disease (American Cancer Society, 2017). A cohort study by Van Leersum et al. showed an increase in the prevalence of comorbidity (i.e. comorbid disease in addition to CRC, from 47% to 62%) and multimorbidity (i.e. 2 or more coexisting conditions in addition to CRC, from 20% to 37%) among 27,339 people with CRC diagnoses in the Netherlands between 1995-2010 (Van Leersum et al., 2013). A recent population-based cohort study found that about one-third of CRC survivors (n=12,265) have at least one comorbidity, in which cardiovascular disease and diabetes are the

most common ones (Cuthbert et al., 2018). Therefore, CRC survivors may be at high risk of functional deficits resulting from cancer-associated deconditioning, commonly associated comorbidities, and ageing-related issues.

Data on functional outcomes of post-treatment CRC survivors

While it is likely that CRC survivors are at risk for long term functional deficits, this has not been well studied. Most studies have focused on the needs and the quality of life of CRC survivors during cancer treatments, and the transition from the active treatment phase to the survivorship phase, however, less is known about their long-term needs in the permanent survivorship phase. Previous studies have found that the inability to perform daily activities and to participate in regular activities are the most prevalent unmet needs among CRC survivors 1 year after treatment (Den Bakker et al., 2018; Sodergren et al., 2019), and these limitations could persist 1-10 years after diagnosis (Den Bakker et al., 2018; Schneider et al., 2007; Sodergren et al., 2019). A national survey (n=21802) found that 43% of CRC survivors have difficulties in usual activities, 37% have difficulties with mobility, and 20% have difficulties with self-care 1-3 years after cancer diagnosis (Downing et al., 2015). About 1 in every 5 CRC survivors were dissatisfied with their functional limitations even 7 years post-diagnosis (Breedveld-Peters et al., 2020). Currently, most studies have assessed the presence of functional needs among CRC survivors using quality of life measures, unmet needs surveys, or one question on activity restriction (Bailey et al., 2014; Downing et al., 2015; Engel et al., 2003; Schneider et al., 2007; Sodergren et al., 2019). Results of these studies demonstrated that functional limitations lower quality of life, contribute to difficulty in performing daily activities and activity restriction, which are the most common unmet needs reported by CRC survivors. While it is likely that CRC

survivors experience functional deficits, the type and the extent of functional limitations that CRC survivors experience is unclear. Understanding the common functional limitations experienced by CRC survivors can help both healthcare providers and CRC survivors be aware of the increased risk of functional limitation and potentially consider strategies to mitigate these issues. However, we do not have a clear picture of the breadth of the problem or the specific functional limitations that are experienced. The first objective of my research was to understand the type and extent of functional deficits experienced by CRC survivors.

First research objective

To better understand the effect of CRC on functional deficits, I used data from the INTERBLEED study (The International Study of the Risk Factors for Gastrointestinal [GI] Bleeding and Cardiovascular Events after GI Bleeding Study) to examine functional abilities after diagnosis.

The INTERBLEED study is an ongoing case-control study examining risk factors for GI bleeding in people with cardiovascular disease, and also a cohort study examining clinical (e.g., subsequent cardiovascular event) and functional outcomes (i.e., functional independence level, cognitive performance) one year after GI bleed. Those with GI bleed are 16.4 - 20 times more likely to have GI cancer and CRC accounted for most diagnoses (Eikelboom et al., 2019; Viborg et al., 2016). INTERBLEED participants completed a functional assessment at study entry, 3 months and 12 months, which provided an opportunity to study the effect of a GI cancer diagnosis on those without a previous GI cancer at baseline, to examine functional performance

within the first year after a diagnosis of GI cancer.

The first objective of my research was to conduct secondary data analysis in people with cardiovascular disease and a recent significant GI bleed that examined:

1. Functional performance within a year after GI cancer diagnosis
2. Change in function over about a year after diagnosis, including both increased difficulty in performing activities and decreased participation in activities.

This analysis was undertaken to better understand the functional outcomes and the types of functional deficits that CRC survivors had after cancer diagnosis.

Understanding the reasons for functional deficits

Understanding the functional deficits that CRC survivors experience is important if we intend to mitigate these deficits. However, to do so also requires an understanding of why the functional deficits exist. One possible reason is the substantial symptom burden among CRC survivors. O'Gorman et al.'s study examined the symptom burden among 496 CRC survivors at least 9 months post-diagnosis (O'Gorman et al., 2018). They found that 66.3% of the participants have at least 2 of the 11 studied symptoms, and 15.5% had more than 5 of the 11 studied symptoms (O'Gorman et al., 2018). These symptoms include fatigue, insomnia, flatulence, constipation, diarrhoea, bloating, appetite loss, weight worry, dry mouth, sore skin, and frequent urination (O'Gorman et al., 2018). The above findings suggest that CRC survivors can experience long-lasting symptoms and a substantial symptom load after 9 months post-CRC diagnosis, potentially contributing to functional decline.

Among all the above symptoms, fatigue is the most common and severe concern among CRC survivors, followed by insomnia, flatulence and dyspnea (Arndt et al., 2004; O'Gorman et al., 2018; Thong et al., 2013). A substantial proportion of short term CRC survivors (<5 years post-diagnosis, 24-78% [Arndt et al., 2004; O'Gorman et al., 2018; Thong et al., 2013]) and long term CRC survivors (≥5 years post-diagnosis, 35% [Thong et al., 2013]) experienced fatigue that is significantly higher compared to the age and gender-matched norms (Thong et al., 2013).

Fatigue also occurs with a number of post-treatment symptoms experienced by CRC survivors (O'Gorman et al., 2018). O'Gorman et al.'s study (n=475) examined which symptoms occur together and the correlations between symptom frequency scores (O'Gorman et al., 2018). Fatigue occurs coincidentally with 14 out of 17 symptoms, including insomnia, constipation, diarrhoea, flatulence, bloating, stool frequency, appetite loss, weight worry, taste, nausea/ vomiting, dry mouth, general, abdominal and buttock pain (O'Gorman et al., 2018). The above findings imply that CRC survivors with other post-treatment symptoms are likely to experience fatigue as well. Recognising the prevalence and severity of fatigue among CRC survivors, fatigue may be the key cause of functional deficits.

Currently, the association between fatigue and functional deficits has been described in a post-treatment mixed cancer population (Jones et al., 2016), but little is known specifically in the CRC population. Jones et al. (n=1294) examined the association between fatigue and disability among post-treatment mixed cancer survivors and found that over 90% of the participants with fatigue have a moderate to severe disability compared to those without fatigue (30.3%) (Jones et al., 2016). Among the 90% with significant fatigue, the majority (70%) reported significant

disability compared to the remaining 30% reporting mild or moderate disability (Jones et al., 2016). This suggests that fatigue can contribute to functional deficits and the burden of fatigue can be very high to cancer survivors from 6 months to 6 years post-treatments, and although similar observations are expected among post-treatment CRC survivors, it is possible that this is an underestimate of fatigue severity experienced by CRC survivors because GI inflammation and ongoing GI symptoms could further exacerbate fatigue.

Assessment of general cancer-related fatigue and the potential issue of its application on CRC survivors

Recognising the prevalence and impact of fatigue on post-treatment cancer survivors, the European Society for Medical Oncology Clinical Practice Guidelines suggests that all post-treatment cancer survivors should be regularly assessed for fatigue when there is a clinical indication (Fabi et al., 2020). They suggested performing a more detailed fatigue assessment if patients reported moderate (4-6) or severe fatigue (7-10) on a 10-point numerical screening tool (Fabi et al., 2020; Given et al., 2008). However, it is unclear which measure is the most appropriate (clinically feasible and psychometric robust) for further assessment of cancer-related fatigue, which is partly due to the lack of consensus on the definition of cancer-related fatigue (Fabi et al., 2020; Wang & Woodruff, 2015). One commonly used definition of cancer-related fatigue refers to 'a distressing, persistent, subjective sense of physical, emotional, and/or cognitive tiredness or exhaustion related to cancer and/or cancer treatment that is not proportional to recent activity and interferes with usual functioning'(Bower, 2014). Due to the subjective nature of fatigue, patient-reported outcome measures are considered the gold standard to assess cancer-related fatigue (Bower, 2014). Numerous patient-reported outcome measures of

fatigue have been developed, varying from single to multiple items and from unidimensional to multidimensional (Maqbali et al., 2019; Minton & Stone, 2009), however there are no specific measures for assessing CRC-related fatigue.

Previous systematic reviews have been conducted on the psychometric properties of the cancer-related fatigue measures (Maqbali et al., 2019; Minton & Stone, 2009; Seyidova-Khoshknabi et al., 2011). Yet, there is no consensus on the most psychometrically robust and clinically feasible measure to assess cancer-related fatigue. The most recent systematic review (2019) indicated the Brief Fatigue Inventory, the Functional Assessment of Chronic Therapy Instrument-Fatigue, the Multidimensional Fatigue Inventory-20, and the Piper Fatigue Scale-Revised as the most comprehensively validated measures and they are reliable and valid in their tested population (Maqbali et al., 2019). However, these measures were mostly validated in mixed or breast cancer populations (Maqbali et al., 2019). These results may not be wholly applicable for those with CRC because CRC-related fatigue can be distinctly due to GI inflammation and GI-related symptoms which may differ from other cancer-related fatigue. The associative factors, severity, and impact of fatigue can be different between cancer sites (Maqbali et al., 2019; Seyidova-khoshknabi et al., 2011; Stone et al., 2000).

The rationale of selecting inflammatory bowel disease as a comparable group to understand fatigue measurements

Generic cancer-related fatigue measures help compare the fatigue levels across cancer types. On the other hand, there could be more comprehensive, CRC-specific, and responsive measures that could better reflect the experience and impact of fatigue for CRC post-treatment survivors. In

general, cancer-related fatigue can be caused by tumour-related complications (e.g. liver failure), side effects of anticancer treatments, and comorbidity (Koornstra et al., 2014). While general cancer-related and CRC-specific fatigue might share similar contributors associated with comorbidity and anticancer treatments, some of the tumor-related factors might be more specific to CRC, such as inflammation in the gut, diarrhoea and abdominal pain. Gut inflammation can cause a unique type of fatigue because inflammation in the gut can affect the neurotransmitter signalling in the brain and cause fatigue (Borren et al., 2019). Therefore, CRC survivors may experience a severity and impact of fatigue that is similar to those with chronic gastrointestinal inflammatory disease such as inflammatory bowel disease (IBD).

Both CRC and IBD affect the GI tract through inflammation, and pro-inflammatory cytokines can cause fatigue (Borren et al., 2019; Bower, 2014). Both diseases share common GI-related symptoms, such as diarrhoea (O'Gorman et al., 2018; Singh et al., 2011) and abdominal pain (O'Gorman et al., 2018; Singh et al., 2011), that can contribute to fatigue. An imbalance of microbiome diversity in the GI tract is common in CRC and IBD (Jahani-Sherafat et al., 2018; Nocerino et al., 2019) which is hypothesized to cause fatigue through the changes in the gut-brain axis (Borren et al., 2019). Although it is unclear how microbiota changes lead to fatigue, some preliminary results suggest that imbalance of microbiome diversity can affect the cytokines release, neurotransmitter balance, and the hypothalamus-pituitary-adrenal system, contributing to fatigue (Borren et al., 2019). Therefore, it is likely that patients with IBD and CRC survivors share similar experiences, severity and impact of fatigue. Measurement of fatigue used in the IBD population may better quantify the fatigue experienced by the CRC post-treatment

survivors. Exploring the most appropriate measure of fatigue in those with IBD may help identify a more appropriate measure of fatigue in CRC survivors.

Second research objective

The second objective of my research was to conduct a systematic review to determine the most psychometrically robust and feasible fatigue measures for the IBD population. Understanding the measurement of fatigue in the IBD population could provide insights into better measures of fatigue in those with CRC. In the third thesis chapter, I evaluated the various patient-reported outcomes measures of fatigue used in studies of adults with IBD to determine the most responsive and clinically useful measures for research and clinical use. These findings may help identify fatigue measures that are the most sensitive and comprehensive to capture the fatigue experienced by CRC survivors. This will enable CRC survivors and health care providers to understand CRC-related fatigue better and determine the most effective interventions, as there are currently very few effective interventions for post-CRC fatigue (Aapro et al., 2017).

The following chapters will describe the results on the individuals' functional changes within the first year of GI cancer diagnosis (Chapter 2) and the results of the systematic review to determine the most psychometrically robust and clinically feasible measure to assess fatigue in the IBD population (Chapter 3) to better understand its potential use in assessing CRC-related fatigue.

Reference

Aapro, M., Scotte, F., Bouillet, T., Currow, D., & Vigano, A. (2017). A Practical Approach to Fatigue Management in Colorectal Cancer. In *Clinical Colorectal Cancer* (Vol. 16, Issue 4, pp. 275–285). https://doi.org/10.1016/j.clcc.2016.04.010

Abu-Helalah, M. A., Alshraideh, H. A., Mohammad, M. A.-H., & Arqoub, K. H. (2014). Quality of Life and Psychological Well-Being of Colorectal Cancer Survivors in Jordan. *Asian Pacific Journal of Cancer Prevention*, *15*(18), 7653–7664. http://journal.waocp.org/article_29817.html

American Cancer Society. (2017). Colorectal Cancer Facts & Figures 2017 - 2019. *Atlanta*, 1–40. https://doi.org/http://dx.doi.org/10.1016/S0140-6736(13)61649-9

Ancoli-Israel, S., Moore, P., & Jones, V. (2001). The relationship between fatigue and sleep in cancer patients: a review. *European Journal of Cancer Care*, *10*(4), 245–255. https://doi.org/10.1046/j.1365-2354.2001.00263.x

Arndt, V., Merx, H., Stegmaier, C., Ziegler, H., & Brenner, H. (2004). Quality of Life in Patients With Colorectal Cancer 1 Year After Diagnosis Compared With the General Population: A Population-Based Study. *Journal of Clinical Oncology*, *22*(23), 4829–4836. https://doi.org/10.1200/JCO.2004.02.018

Arndt, V., Merx, H., Stegmaier, C., Ziegler, H., & Brenner, H. (2006). Restrictions in quality of life in colorectal cancer patients over three years after diagnosis: A population based study. *European Journal of Cancer*, *42*(12), 1848–1857. https://doi.org/https://doi.org/10.1016/j.ejca.2006.01.059

Bailey, C. E., Cao, H. S. T., Hu, C. Y., Chang, G. J., Feig, B. W., Rodriguez-Bigas, M. A., Nguyen, S. T., Skibber, J. M., & You, Y. N. (2014). Functional Deficits and Symptoms of

Long-Term Survivors of Colorectal Cancer Treated by Multimodality Therapy Differ by Age at Diagnosis. *Journal of Gastrointestinal Surgery*, *19*(1), 180–188. https://doi.org/10.1007/s11605-014-2645-7

Borren, N. Z., Woude, C. J. van der, & Ananthakrishnan, A. N. (2019). Fatigue in IBD: epidemiology, pathophysiology and management. *Nature Reviews Gastroenterology and Hepatology*, *16*, 247–259.

Bower, J. E. (2014). Cancer-related fatigue--mechanisms, risk factors, and treatments. *Nature Reviews. Clinical Oncology*, *11*(10), 597–609. https://doi.org/10.1038/nrclinonc.2014.127

Braamse, A. M. J., van Turenhout, S. T., Terhaar sive Droste, J. S., de Groot, G. H., van der Hulst, R. W. M., Klemt-Kropp, M., Kuiken, S. D., Loffeld, R. J. L. F., Uiterwaal, M. T., Mulder, C. J. J., & Dekker, J. (2016). Factors associated with anxiety and depressive symptoms in colorectal cancer survivors. *European Journal of Gastroenterology & Hepatology*, *28*(7). https://journals.lww.com/eurojgh/Fulltext/2016/07000/Factors_associated_with_anxiety_and_depressive.17.aspx

Breedveld-Peters, J. J. L., Bours, M. J. L., Cords, C. I., Ditters, I. A. M., Habraken, V., Jongen, M. W. J., & Weijenberg, M. P. (2020). The impact of participation restrictions on everyday life in long-term colorectal cancer survivors in the EnCoRe study: A mixed-method study. *European Journal of Oncology Nursing : The Official Journal of European Oncology Nursing Society*, *45*, 101724. https://doi.org/10.1016/j.ejon.2020.101724

Bryan, S., Masoud, H., Weir, H. K., Woods, R., Lockwood, G., Smith, L., Brierley, J., Gospodarowicz, M., & Badets, N. (2018). *Cancer in Canada: Stage at diagnosis*. https://www150.statcan.gc.ca/n1/pub/82-003-x/2018012/article/00003-eng.htm

Canadian Cancer Society. (2019). *Canadian Cancer Statistics 2019*.

Cancer Research UK. (2018). *Bowel changes after treatment*. https://www.cancerresearchuk.org/about-cancer/bowel-cancer/living-with/bowel-changes-after-treatment

Cuthbert, C. A., Hemmelgarn, B. R., Xu, Y., & Cheung, W. Y. (2018). The effect of comorbidities on outcomes in colorectal cancer survivors: a population-based cohort study. *Journal of Cancer Survivorship*, *12*(6), 733–743. https://doi.org/10.1007/s11764-018-0710-z

Dean, G., Ph, D., Nelson, R., Ph, D., & Chu, D. (2007). Quality of Life Outcomes in 599 Cancer and Non-Cancer Patients with Colostomies. *The Journal of Surgical Research*, *138*(1), 79–87. https://doi.org/10.1016/j.jss.2006.04.033

Den Bakker, C. M., Schaafsma, F. G., Huirne, J. A. F., Consten, E. C. J., Stockmann, H. B. A. C., Rodenburg, C. J., De Klerk, G. J., Bonjer, H. J., & Anema, J. R. (2018). Cancer survivors' needs during various treatment phases after multimodal treatment for colon cancer - Is there a role for eHealth? *BMC Cancer*, *18*(1), 1–12. https://doi.org/10.1186/s12885-018-5105-z

Downing, A., Morris, E. J. A., Richards, M., Corner, J., Wright, P., Sebag-Montefiore, D., Finan, P., Kind, P., Wood, C., Lawton, S., Feltbower, R., Wagland, R., Vernon, S., Thomas, J., & Glaser, A. W. (2015). Health-related quality of life after colorectal cancer in England: A patient-reported outcomes study of individuals 12 to 36 months after diagnosis. *Journal of Clinical Oncology*, *33*(6), 616–624. https://doi.org/10.1200/JCO.2014.56.6539

Drury, A., Payne, S., & Brady, A.-M. (2017). The cost of survival: an exploration of colorectal cancer survivors' experiences of pain. *Acta Oncologica*, *56*(2), 205–211.

https://doi.org/10.1080/0284186X.2016.1266084

Eikelboom, J. W., Connolly, S. J., Bosch, J., Shestakovska, O., Aboyans, V., Alings, M., Anand, S. S., Avezum, A., Berkowitz, S. D., Bhatt, D. L., Cook-Bruns, N., Felix, C., Fox, K. A. A., Hart, R. G., Maggioni, A. P., Moayyedi, P., O'Donnell, M., Rydén, L., Verhamme, P., … Yusuf, S. (2019). Bleeding and New Cancer Diagnosis in Patients With Atherosclerosis. *Circulation*, *140*(18), 1451–1459. https://doi.org/10.1161/CIRCULATIONAHA.119.041949

Engel, J., Kerr, J., Schlesinger-Raab, A., Eckel, R., Sauer, H., & Hölzel, D. (2003). Quality of life in rectal cancer patients: a four-year prospective study. *Annals of Surgery*, *238*(2), 203–213. https://doi.org/10.1097/01.sla.0000080823.38569.b0

Eyl, R. E., Koch-Gallenkamp, L., Jansen, L., Walter, V., Carr, P. R., Hoffmeister, M., Chang-Claude, J., Brenner, H., & Arndt, V. (2020). Physical Activity and Long-term Quality of Life among Colorectal Cancer Survivors—A Population-based Prospective Study. *Cancer Prevention Research*, *13*(7), 611 LP – 622. https://doi.org/10.1158/1940-6207.CAPR-19-0377

Fabi, A., Bhargava, R., Fatigoni, S., Guglielmo, M., Horneber, M., Roila, F., Weis, J., Jordan, K., & Ripamonti, C. I. (2020). Cancer-related fatigue: ESMO Clinical Practice Guidelines for diagnosis and treatment. *Annals of Oncology*, *31*(6), 713–723. https://doi.org/10.1016/j.annonc.2020.02.016

Forsberg, C., & Cedermark, B. (1996). Well-being, general health and coping ability: 1-year follow-up of patients treated for colorectal and gastric cancer. *European Journal of Cancer Care*, *5*(4), 209–216. https://doi.org/10.1111/j.1365-2354.1996.tb00237.x

Foster, C., Wright, D., Hill, H., Hopkinson, J., & Roffe, L. (2009). Psychosocial implications of

living 5 years or more following a cancer diagnosis: A systematic review of the research evidence. *European Journal of Cancer Care*, *18*(3), 223–247. https://doi.org/10.1111/j.1365-2354.2008.01001.x

Given, B., Given, C. W., Sikorskii, A., Jeon, S., McCorkle, R., Champion, V., & Decker, D. (2008). Establishing mild, moderate, and severe scores for cancer-related symptoms: how consistent and clinically meaningful are interference-based severity cut-points? *Journal of Pain and Symptom Management*, *35*(2), 126–135. https://doi.org/10.1016/j.jpainsymman.2007.03.012

Globocan. (2012). *Cancers Fact Sheets : Colorectal Cancer*. 1–10. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx

Jahani-Sherafat, S., Alebouyeh, M., Moghim, S., Ahmadi Amoli, H., & Ghasemian-Safaei, H. (2018). Role of gut microbiota in the pathogenesis of colorectal cancer; a review article. *Gastroenterology and Hepatology from Bed to Bench*, *11*(2), 101–109. https://pubmed.ncbi.nlm.nih.gov/29910850

Jones, J. M., Olson, K., Catton, P., Catton, C. N., Fleshner, N. E., Krzyzanowska, M. K., McCready, D. R., Wong, R. K. S., Jiang, H., & Howell, D. (2016). Cancer-related fatigue and associated disability in post-treatment cancer survivors. *Journal of Cancer Survivorship : Research and Practice*, *10*(1), 51–61. https://doi.org/10.1007/s11764-015-0450-2

Koornstra, R. H. T., Peters, M., Donofrio, S., van den Borne, B., & de Jong, F. A. (2014). Management of fatigue in patients with cancer -- a practical overview. *Cancer Treatment Reviews*, *40*(6), 791–799. https://doi.org/10.1016/j.ctrv.2014.01.004

Maqbali, M. Al, Hughes, C., Gracey, J., Rankin, J., Hacker, E., Al, M., Hughes, C., Gracey, J., &

Rankin, J. (2019). Quality assessment criteria : psychometric properties of measurement tools for cancer related fatigue. *Acta Oncologica*, *58*(9), 1286–1297. https://doi.org/10.1080/0284186X.2019.1622773

Marzorati, C., Riva, S., & Pravettoni, G. (2017). Who Is a Cancer Survivor? A Systematic Review of Published Definitions. *Journal of Cancer Education*, *32*(2), 228–237. https://doi.org/10.1007/s13187-016-0997-2

Minton, O., & Stone, P. (2009). A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). *Annals of Oncology*, *20*(1), 17–25. https://doi.org/10.1093/annonc/mdn537

Nocerino, A., Nguyen, A., Agrawal, M., Mone, A., Lakhani, K., & Swaminath, A. (2019). Fatigue in Inflammatory Bowel Diseases: Etiologies and Management. *Advances in Therapy*. https://doi.org/https://dx.doi.org/10.1007/s12325-019-01151-w

O'Gorman, C., Stack, J., O'Ceilleachair, A., Denieffe, S., Gooney, M., McKnight, M., & Sharp, L. (2018). Colorectal cancer survivors: an investigation of symptom burden and influencing factors. *BMC Cancer*, *18*(1), 1022. https://doi.org/10.1186/s12885-018-4923-3

Pucciarelli, S., Del Bianco, P., Efficace, F., Toppan, P., Serpentini, S., Friso, M. L., Lonardi, S., De Salvo, G. L., & Nitti, D. (2010). Health-related quality of life, faecal continence and bowel function in rectal cancer patients after chemoradiotherapy followed by radical surgery. *Supportive Care in Cancer*, *18*(5), 601–608. https://doi.org/10.1007/s00520-009-0699-y

Ramsey, S. D., Berry, K., Moinpour, C., Giedzinska, A., & Andersen, M. R. (2002). Quality of life in long term survivors of colorectal cancer. *The American Journal of Gastroenterology*, *97*(5), 1228–1234. https://doi.org/10.1111/j.1572-0241.2002.05694.x

Rauch, P., Miny, J., Conroy, T., Neyton, L., & Guillemin, F. (2004). Quality of life among disease-free survivors of rectal cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *22*(2), 354–360. https://doi.org/10.1200/JCO.2004.03.137

Rutherford, C., Müller, F., Faiz, N., King, M. T., & White, K. (2020). Patient-reported outcomes and experiences from the perspective of colorectal cancer survivors: meta-synthesis of qualitative studies. *Journal of Patient-Reported Outcomes*, *4*(1), 27. https://doi.org/10.1186/s41687-020-00195-9

Schmidt, C. E., Bestmann, B., Küchler, T., Longo, W. E., & Kremer, B. (2005). Prospective Evaluation of Quality of Life of Patients Receiving Either Abdominoperineal Resection or Sphincter-Preserving Procedure for Rectal Cancer. *Annals of Surgical Oncology*, *12*(2), 117–123. https://doi.org/10.1245/ASO.2005.12.036

Schneider, E. C., Malin, J. L., Kahn, K. L., Ko, C. Y., Adams, J., & Epstein, A. M. (2007). Surviving colorectal cancer. *Cancer*, *110*(9), 2075–2082. https://doi.org/https://doi.org/10.1002/cncr.23021

Seyidova-khoshknabi, D., Davis, M. P., & Walsh, D. (2011). *A Systematic Review of Cancer-Related Fatigue Measurement Questionnaires*. *28*(2), 119–129. https://doi.org/10.1177/1049909110381590

Seyidova-Khoshknabi, D., Davis, M. P., & Walsh, D. (2011). Review Article: A Systematic Review of Cancer-Related Fatigue Measurement Questionnaires. *American Journal of Hospice and Palliative Medicine*, *28*(2), 119–129. https://doi.org/10.1177/1049909110381590

Singh, S., Blanchard, A., Walker, J. R., Graff, L. A., Miller, N., & Bernstein, C. N. (2011).

Common Symptoms and Stressors Among Individuals With Inflammatory Bowel Diseases. *Clinical Gastroenterology and Hepatology*, *9*(9), 769–775. https://doi.org/10.1016/j.cgh.2011.05.016

Sodergren, S. C., Wheelwright, S. J., Permyakova, N. V., Patel, M., Calman, L., Smith, P. W. F., Din, A., Richardson, A., Fenlon, D., Winter, J., Corner, J., & Foster, C. (2019). Supportive care needs of patients following treatment for colorectal cancer: risk factors for unmet needs and the association between unmet needs and health-related quality of life—results from the ColoREctal Wellbeing (CREW) study. *Journal of Cancer Survivorship*. https://doi.org/10.1007/s11764-019-00805-6

Spira, A. P., Covinsky, K., Rebok, G. W., Punjabi, N. M., Stone, K. L., Hillier, T. A., Ensrud, K. E., & Yaffe, K. (2012). Poor sleep quality and functional decline in older women. *Journal of the American Geriatrics Society*, *60*(6), 1092–1098. https://doi.org/10.1111/j.1532-5415.2012.03968.x

Spira, A. P., Kaufmann, C. N., Kasper, J. D., Ohayon, M. M., Rebok, G. W., Skidmore, E., Parisi, J. M., & Reynolds Charles F, 3rd. (2014). Association between insomnia symptoms and functional status in U.S. older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *69 Suppl 1*(Suppl 1), S35–S41. https://doi.org/10.1093/geronb/gbu116

Sprangers, M. A., Taal, B. G., Aaronson, N. K., & te Velde, A. (1995). Quality of life in colorectal cancer. Stoma vs. nonstoma patients. *Diseases of the Colon & Rectum*, *38*(4), 361–369. https://doi.org/10.1007/BF02054222

Stone, P., Richards, M., A'Hern, R., & Hardy, J. (2000). A study to investigate the prevalence, severity and correlates of fatigue among patients with cancer in comparison with a control

group of volunteers without cancer. *Annals of Oncology : Official Journal of the European Society for Medical Oncology*, *11*(5), 561–567. https://doi.org/10.1023/a:1008331230608

Sun, V., Grant, M., McMullen, C. K., Altschuler, A., Mohler, M. J., Hornbrook, M. C., Herrinton, L. J., Baldwin, C. M., & Krouse, R. S. (2013). Surviving colorectal cancer: Long-term, persistent ostomy-specific concerns and adaptations. *Journal of Wound, Ostomy and Continence Nursing*, *40*(1), 61–72. https://doi.org/10.1097/WON.0b013e3182750143

Thong, M. S. Y., Mols, F., Wang, X. S., Lemmens, V. E. P. P., Smilde, T. J., & Van De Poll-Franse, L. V. (2013). Quantifying fatigue in (long-term) colorectal cancer survivors: A study from the population-based Patient Reported Outcomes Following Initial treatment and Long term Evaluation of Survivorship registry. *European Journal of Cancer*, *49*(8), 1957–1966. https://doi.org/10.1016/j.ejca.2013.01.012

Van Leersum, N. J., Janssen-Heijnen, M. L. G., Wouters, M. W. J. M., Rutten, H. J. T., Coebergh, J. W., Tollenaar, R. A. E. M., & Lemmens, V. E. P. P. (2013). Increasing prevalence of comorbidity in patients with colorectal cancer in the South of the Netherlands 1995-2010. *International Journal of Cancer*, *132*(9), 2157–2163. https://doi.org/10.1002/ijc.27871

Viborg, S., Søgaard, K. K., Farkas, D. K., Nørrelund, H., Pedersen, L., & Sørensen, H. T. (2016). Lower Gastrointestinal Bleeding And Risk of Gastrointestinal Cancer. *Clinical and Translational Gastroenterology*, *7*(4), e162–e162. https://doi.org/10.1038/ctg.2016.16

Victorson, D. E., Anton, S., Hamilton, A., Yount, S., & Cella, D. (2009). A conceptual model of the experience of dyspnea and functional limitations in chronic obstructive pulmonary disease. *Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, *12*(6), 1018–1025. https://doi.org/10.1111/j.1524-

4733.2009.00547.x

Wang, X. S., & Woodruff, J. F. (2015). Cancer-related and treatment-related fatigue. *Gynecologic Oncology*, *136*(3), 446–452. https://doi.org/10.1016/j.ygyno.2014.10.013

Yin, L., Fan, L., Tan, R., Yang, G., Jiang, F., Zhang, C., Ma, J., Yan, Y., Zou, Y., Zhang, Y., Wang, Y., & Zhang, G. (2018). Bowel symptoms and self-care strategies of survivors in the process of restoration after low anterior resection of rectal cancer. *BMC Surgery*, *18*(1), 35. https://doi.org/10.1186/s12893-018-0368-5

Chapter 2

Functional outcomes of individuals with cardiovascular disease after gastrointestinal cancer

Functional outcomes of individuals with cardiovascular disease after gastrointestinal cancer

Vanessa S.K. Fan [1], Julie Richardson [1,2], Lauren E. Griffith[2], John Eikelboom[3,4], Jackie Bosch[1,4]

[1] School of Rehabilitation Science, McMaster University

[2] Department of Health Research Methods, Evidence, and Impact, McMaster University

[3] Department of Medicine, McMaster University

[4] Population Health Research Institute

Corresponding author: Jackie Bosch

This manuscript is prepared for the submission to the Canadian Journal of Occupational Therapy.

Functional outcomes of individuals with cardiovascular disease after gastrointestinal cancer

Abstract

Background. Gastrointestinal (GI) cancer likely increases the risk of functional impairment.

Purpose. To examine change in functional ability after GI cancer diagnoses.

Methods. We identified INTERBLEED (The International Study of the Risk Factors for GI Bleeding and Cardiovascular Events after GI Bleeding) participants who were enrolled in the cohort portion of the study, were diagnosed with GI cancer after baseline and completed the Standard Assessment of Global Everyday Activities (SAGEA) pre and post-GI cancer diagnoses. We then used the Wilcoxon Sign Rank test to analyse the difference in SAGEA scores and task participation.

Findings. Twenty-six participants had a mean age of 79, mostly men, and reported some baseline functional impairments. Their SAGEA scores were not significantly different from pre- and post-GI cancer. However, participants performed fewer tasks after GI cancer.

Implication. Individuals experience some functional decline after GI cancer. Further research is needed to understand the longer-term prognosis.

Keywords: Activities of daily living, Independent living, Occupational participation, Occupational performance, Impairments

Introduction

Gastrointestinal (GI) cancer is one of the commonly diagnosed cancers in the world, accounting for 26% of new cancer cases each year (Arnold et al., 2020; International Agency for Research on Cancer, n.d.) and has a 5-year survival rate of about 65% for colorectal cancer, and 8-28% for other GI cancer types (Canadian Cancer Society, 2019). GI cancer can impair one's ability to perform usual daily activities (functional deficits) due to (1) physical impairments (e.g. pain [Drury et al., 2017; Forsberg & Cedermark, 1996; Rauch et al., 2004; Rutherford et al., 2020], fatigue [Arndt et al., 2006; Drury et al., 2017; Forsberg & Cedermark, 1996; O'Gorman et al., 2018; Rutherford et al., 2020], insomnia [Arndt et al., 2006; Drury et al., 2017; Forsberg & Cedermark, 1996; O'Gorman et al., 2018; Rutherford et al., 2020], dyspnea [Arndt et al., 2006], bowel and sexual dysfunction [Arndt et al., 2006; Dean et al., 2007; Forsberg & Cedermark, 1996; O'Gorman et al., 2018; Rauch et al., 2004; Rutherford et al., 2020], neuropathy [Den Bakker et al., 2018], and stoma [O'Gorman et al., 2018]), (2) cognitive impairments (e.g. chemotherapy-induced cognitive decline [Den Bakker et al., 2018]) and (3) psychological sequelae (e.g. anxiety and depression [Abu-Helalah et al., 2014; Braamse et al., 2016; Foster et al., 2009; Ramsey et al., 2002]) from both cancer and its treatments. Early diagnosis of GI cancer is difficult as symptoms often do not occur until a more advanced stage and these symptoms often mimic those of the ulcer-related diseases (e.g. loss of appetite, ingestion, abdominal pain)(Yale Medicine, n.d.), therefore it is possible that the effects of cancer on function may begin well before the GI cancer is diagnosed. While over 100,000 people are living with GI cancer in Canada, we know very little about the functional abilities of this group at high risk for functional deficits (Government of Canada, 2019).

GI cancer is 20 times more commonly diagnosed among individuals with cardiovascular disease (CVD) after GI bleeding (Eikelboom et al., 2019). Although the pathological mechanism between GI bleeding and GI cancer is unclear, GI bleeding might be the first indicator of GI cancer (Viborg et al., 2016). Individuals with CVD often take anti-thrombotic medications which increase the risk of GI bleeding (Bhatt et al., 2006; CAPRIE Steering committee, 1996). Investigation of the GI bleed may unmask GI cancer, resulting in early diagnosis and treatment (Eikelboom et al., 2019). It has been reported that people with CVD had lower functional level by 38% than the age, race and sex comparable group without CVD (Kucharska-Newton et al., 2017). GI bleeding can also lead to further functional deficits in patients with acute ischemic stroke, as 20-40% more people with GI bleed experienced functional deficits than those without GI bleed (Donnell et al., 2008; Ogata et al., 2014; Rumalla & Mittal, 2016). Yet no studies have prospectively examined the functional abilities of people, with CVD and GI bleed, before and after their GI cancer diagnosis, despite the functional risk. Even in the early stages of GI cancer it is possible that altered bowel function and fatigue affect the ability to perform usual activities. While it may not be possible to address symptoms related to disease progression, addressing functional deficits may improve or at least slow the deterioration in function and quality of life reported by those with GI cancer (Chau et al., 2019).

In this secondary data analysis, we examined whether people with CVD and a recent significant GI bleed, diagnosed with a new GI cancer experienced:

1. Functional deficits after the GI cancer diagnosis
2. A change in functional ability from prior to their diagnosis
3. A change in participation in functional tasks

Methods

Data sources and setting

We used data from the INTERBLEED study (The International Study of the Risk Factors for Gastrointestinal (GI) Bleeding and Cardiovascular Events after GI Bleeding). INTERBLEED is both a case-control study, examining risk factors for GI bleeding in people with CVD (cases) compared to those with CVD alone (controls), and a prospective cohort study following these same participants at 3 months and 1 year after their GI bleed (cases) or enrolment (controls) to determine the rate of cardiovascular events and functional outcomes. The study is conducted in 26 sites across 10 countries and recruitment began in 2014, with 2,216 cases and 1,126 controls recruited to date out of an expected 2500 of each. Ethics approval was obtained from each participating centre, and informed consent was obtained from all participants.

Inclusion criteria of participants for this analysis

INTERBLEED participants were included in this analysis if they: (1) were ≥ 18 years of age, (2) had previously established CVD (Coronary artery disease: myocardial infarction, stable/ unstable angina, coronary revascularization; Cerebrovascular disease: ischemic stroke, transient ischemic attack, heart failure, atrial fibrillation or flutter, venous thromboembolism; Peripheral arterial disease: Peripheral arterial disease in lower/ upper limb, carotid stenosis, aortic aneurysm, or peripheral revascularisation, (3) had a significant GI bleeding (melena, hematochezia, or hematemesis), (4) did not have a history of GI cancer at the time of the GI bleed, (5) reported a new GI cancer (including cancers developed along the GI tract, pancreas, and liver) on or before their final study visit, and (6) completed a baseline functional assessment, the Standard Assessment of Global Everyday Activities, before or within 7 days of new GI cancer diagnoses

(participants were asked to report their performance on SAGEA over the past month and therefore SAGEA completed within 7 days of GI diagnosis should reflect a pre-diagnosis functional status).

Data Collection

Sociodemographic characteristics (sex, age, ethnicity, education, and living conditions), anthropometrics, and medical history were collected at study entry. Details on the clinical diagnoses of GI bleeding (types, site, severity, and pathology) were collected after study entry. At hospital discharge and each follow up, participants were asked about events that occurred since the GI bleed, including GI cancer, and the date of diagnosis was collected.

Primary outcome

Functional ability was assessed using the patient-reported outcome measure, the Standard Assessment of Global Everyday Activities (SAGEA), at baseline (time of GI bleed), 3 months and 12 months after GI bleed. The SAGEA is a 15-item measure which assesses the difficulty performing tasks in 4 domains: cognitive activities of daily living (cADLs) (3 items), instrumental ADLs (iADLs) (7 items), mobility (2 items), and basic ADLs (bADLs)(3 items)(Marzona Irene et al., 2011). Participants are asked if they performed the tasks in the previous month and if they did, whether they had difficulty performing the task (rated using a four-point Likert scale ranging from‘0 - No difficulty’ to ‘3 – Severe difficulty’) (Marzona Irene et al., 2011). Total SAGEA scores range from 0 to 24, with a *higher score indicating more functional limitation*. The total score is the sum of 4 subscores. The cognitive and basic ADL subscores are the sum of the item scores within each domain. For mobility and iADL domain, we

compute the maximum score among the items within each domain as the subscore. Details on scoring can be found in Supplemental file 1. The SAGEA was conducted in person within 7 days of the study entry and at telephone follow-up visits at 3 and 12 months after GI bleed. The SAGEA has been validated in patients with cardiac surgery and demonstrated excellent reliability (Intraclass correlation= 0.99 for in-person and telephone administrations of the SAGEA) and moderate-to-strong construct validity (correlation coefficient = 0.54-0.8 across subscales with their comparator measures)(Spence et al., 2021).

Estimates of function pre and post GI cancer

Pre GI cancer function was estimated using the SAGEA completed before the date of GI cancer diagnosis. Post GI cancer function was estimated using the SAGEA completed farthest from the diagnosis date of the GI cancer.

The pre GI cancer SAGEA was completed 8 days (median) before GI cancer diagnoses (IQR: 56 days, the earliest day: 306 days before GI cancer diagnosis, the latest day: 7 day post GI cancer diagnosis), with the majority (61%) completed within 1 month of their diagnoses. The post GI cancer SAGEA was completed 266 days (median) after GI cancer diagnoses (IQR: 262 days, Range: 23 – 394 days after GI cancer diagnosis), with the majority (58%) completed 8-14 months after GI cancer diagnoses. The average number of days between pre and post-GI cancer SAGEA was 358 days (median) (IQR: 275 days, Range: 78-481 days).

Statistical analyses

The normality of continuous variables was tested using the Shapiro-Wilk test and visually examined by the histogram. Means and standard deviations (SD) are presented for normally distributed data, otherwise medians and interquartile ranges (IQR) are presented. For categorical variables, frequency and percentages are presented. The SAGEA scores and the number of performed tasks were compared between pre and post GI cancer, using the Wilcoxon Signed Rank test because the differences were not normally distributed.

The proportion of participants with a change in function was calculated as follows: the number of participants whose function improved (i.e. change score <0), worsened (i.e. change score >0), or did not change (i.e. change score=0) divided by the total number of participants. The number of performed tasks were compared between pre and post GI cancer using the Wilcoxon Signed Rank test. The proportion of participants with a change in number of performed tasks was calculated as follows: the number of participants who performed fewer functional tasks (i.e. change in number of performed tasks: <0), performed the same number of functional tasks (i.e. change in number of performed tasks = 0), or performed more functional tasks (i.e. change in number of performed tasks >0) divided by the total number of participants.

The proportion of participants who no longer continued the task was calculated as follows: the number of participants who performed before GI cancer and did not continue after GI cancer divided by the number of participants who performed the task at pre GI cancer. The proportion of participants who continued to perform the task after GI cancer and reported changes in difficulty of each functional task was calculated as follows: the number of participants who had

more difficulty (i.e. at least 1 point increase in the item score), less difficulty (i.e. at least 1 point decrease in the item score), or no change in difficulty (i.e. change of item score=0) divided by the number of participants who performed at the tasks before GI cancer diagnosis. Two-sample Wilcoxon rank-sum test was used to test whether the difference in the baseline SAGEA score and its subscores were statistically different between those who completed both pre and post GI cancer SAGEA and those who only completed baseline SAGEA. STATA 15.1 and Excel 16.46 were used for the statistical analyses (StataCorp, 2017).

Results

From the 2216 cases recruited in INTERBLEED, 46 participants were identified who were at least 18 years old, had established CVD and significant GI bleed, and reported newly diagnosed GI cancer after GI bleed. Of these, 3 had a history of GI cancer, 8 did not have a baseline SAGEA, and 2 did not complete baseline SAGEA before the newly diagnosed GI cancer (no data on pre-GI cancer SAGEA was available), therefore 13 were excluded (Figure 1). Of the remaining 33 patients, 7 did not complete a follow-up SAGEA, 5 of which died prior to follow up, leaving a sample of 26 participants for this analysis.

[Insert Figure 1 here]

Baseline characteristics

Participants had a mean age of 79 (SD ± 9.2), were primarily men (65%), white (96%), had either high school or college education (79%), and were living at home (88%) prior to their GI cancer diagnosis (Table 1). The majority were identified at the inpatient ward or endoscopy/ colonoscopy (77%) and had melena (69%) which was primarily due to malignancy (46.2%)

(Table 1). Most had hypertension (65%), anemia (62%), dyslipidemia (58%) and atrial fibrillation (58%) (Table 2).

Seven participants who only completed the pre-GI cancer SAGEA were excluded from our analysis. Their baseline median overall SAGEA score (pre only: 3 vs pre and post: 3.5), cADL (pre only: 1 vs pre and post: 0), IADL (pre only: 0 vs pre and post: 0), mobility (pre only: 1 vs pre and post: 1), and bADL (pre only: 0 vs pre and post: 0) subscores were not significantly different from the participants who completed both pre- and post-GI cancer SAGEA. The characteristics of those excluded patients can be found in Table 1.

[Insert Table 1,2 here]

SAGEA scores Pre and Post GI Cancer Diagnosis

Pre GI cancer diagnosis, participants had an overall SAGEA score of 3.5 (median, IQR: 6), a cADL score of 0 (median, IQR: 2), an iADL score of 0 (median, IQR: 1), a mobility score of 1 (median, IQR: 2), and a bADL score of 0 (median, IQR: 2) (Table 3).

Post GI cancer diagnosis, the overall SAGEA score was 3 (median, IQR: 7), the cADL score was 0 (median, IQR: 2), the iADL score was 0 (median, IQR: 1), the mobility score was 1 (median, IQR: 2), and the bADL score was 0.5 (median, IQR: 3) (Table 3). The overall SAGEA score and the subscores were not significantly different between pre and post GI cancer diagnosis (Table 3).

[Insert Table 3 here]

Proportion of participants with functional changes after GI cancer diagnosis

After GI cancer diagnosis, half of the participants (50%) had worsened function (i.e., had at least a 1-point increase in SAGEA score), while 42% improved and 8% did not change. More people improved in cADL and iADL (23% and 27%) then worsened (15% and 23%), but the majority did not change (62% and 50%) (Figure 2). Similar numbers of people improved, stayed the same, or worsened in their mobility (Figure 2). Participants mostly experienced no change (46%) or worse function (35%) in bADL, and 19% of them improved (Figure 2).

[Insert Figure 2 here]

Change in performing functional tasks

Of the list of 12 tasks on the SAGEA for which one can answer that they do not perform, on average participants performed 10 tasks (median; IQR:1) pre GI cancer diagnosis. The number of performed tasks were 6 (median, IQR:1) in iADL, 2 (median, IQR: 0) in mobility, and 3 (median, IQR:0) in bADL before GI cancer diagnosis.

Post GI cancer diagnosis, participants performed 9 tasks (median, IQR:3). The number of performed tasks were 4.5 (median, IQR: 3) in iADL, 2 (median, IQR: 1) in mobility, and 3 (median, IQR: 0) in bADL after GI cancer diagnosis. Compared to pre diagnosis, participants performed fewer tasks in general ($p=0.0028$), iADL ($p=0.0082$) and mobility ($p=0.025$), but not for bADL($p=0.083$) after GI cancer diagnosis.

Among the 26 participants, 18 (69%) performed fewer functional tasks, 4 (15%) performed the same number of functional tasks, and 4 (15%) performed more functional tasks at post GI cancer than pre GI cancer state. Among the 18 participants who decreased their functional tasks, 8 (44.4%) reduced 1-2 tasks, 5 (27.8%) reduced 3-4 tasks, and 5 (27.8%) reduced 5-6 tasks.

The most common tasks that participants no longer continued after GI cancer diagnoses were: iADLs ('organising a trip/social activities' [61.5%], 'finding your way around a new building' [61.1%], and 'preparing a meal and/or doing laundry' [50%]), mobility (using stairs [18.2%]), and bADLs (transferring from bed to chair [13%] (Table 4).

[Insert Table 4 here]

Participants who performed fewer tasks (n=18) had an increased SAGEA score from pre to post GI cancer (median of the difference: +2, IQR: 5). Participants who performed the same number of tasks (n=4) or more tasks (n=4) had a decreased SAGEA score from pre to post GI cancer (median of the difference: -1, IQR:4.5) (Supplemental table 1).

Change in difficulty in performing functional tasks

Pre GI cancer diagnosis about 19% of the participants had no difficulty in performing tasks, this was reduced to 15% post GI cancer diagnosis. Same proportion of participants (69%) had no difficulty in cADL pre and post GI cancer diagnosis. For those that continued to perform tasks post GI cancer diagnosis, 42.3% reported increased difficulty walking, 24% had difficulty with bathing/toileting and 17.4% transferring from bed to chair (Table 4).

Post GI cancer diagnosis participants reported the most difficulty with walking (Supplemental figure 1.1), followed by transferring from bed to chair and bathing/toileting (Supplemental figure 1.2).

Discussion

This is the first prospective cohort study to describe functional changes for people diagnosed with GI cancer, in a population with CVD who have had a significant GI bleed. We found that participants had pre-existing functional deficits before their GI cancer diagnosis (median SAGEA score 3.5) and did not experience a significant change in functional deficits after GI cancer diagnosis (median SAGEA score 3). Participants performed fewer functional tasks after GI cancer diagnosis (10 tasks compared to 9 post GI cancer diagnosis), particularly iADLs. Most participants no longer 'organised a trip/social activities'(61.5%) and 'found ways around a new building' (61.1%). Participants also reported more difficulty in performing tasks that they had been doing prior to their diagnosis, particularly walking (42.3%).

Some functional limitations are not unexpected in a sample that is older (79 years), has CVD, and experienced GI bleeding. The normative data from the World Health Organisation Disability Assessment Schedule 2.0 (WHODAS 2.0) demonstrated some functional deficits (WHODAS median: 3) in a sample of 796 individuals within the age of 75-85 (Andrews et al., 2009). Based on the functional outcome score, our sample appeared to have a similar functional level than age-matched norm (Norm: WHODAS 2.0 median of 3 < Our sample: Pre GI cancer SAGEA: 3.5, Post GI cancer SAGEA: 3)(Andrews et al., 2009). However there was no significant change in SAGEA score after GI cancer diagnosis. Thus, we further analysed the participation and the

change in difficulty among the functional tasks to understand the patterns of functional deficits.

Our findings also showed that participants perform fewer functional tasks after GI cancer diagnosis. On average participants performed 10 tasks before GI cancer and 9 tasks after. About 70% of participants (n=18) reduced participation in at least 1 functional task, mostly related to iADLs. This observation is consistent with Sodergren et al.'s findings where they identified "Not being able to do the things you used to" as the most common moderate-to-severe unmet needs among colorectal cancer patients at 15 and 24 months after treatments (Sodergren et al., 2019). The decrease in task participation could imply poorer function, opposed to fewer opportunities to perform tasks, as we found that participants who reduced participation had worse function (a median increase in SAGEA score of 2) after GI cancer diagnosis. In contrast, participants who continued their tasks seemed to have improved functional abilities (median decrease in SAGEA score of 1). Further follow up is needed to understand whether the decrease in function continues. If so, it means that asking a simple question to understand the reason behind the lack of task participation could identify an area of concern and the need to review strategies for maintaining ability.

Our study found that participants continued tasks, such as walking, but with more difficulty. While participants did not change their difficulty level for most tasks (11 out of 13), more participants (43%) reported increased difficulty in walking than those with no change (35%) or less difficulty (19%). It would be helpful to understand the issues that are causing more difficulty. It is likely that pain and fatigue, the most commonly identified symptoms after GI cancer (Kobayashi et al., 2011; Lv et al., 2014; Schneider et al., 2007; Stauder et al., 2013; Sun

& Sarna, 2008), are making it more difficult to walk. Strategies in minimizing these deficits may be warranted, thus prolonging the ability to walk without increased difficulty.

We did not see any difficulty in performing cADLs (median cADL score=0) and cognitive decline after GI cancer. The SAGEA is a measure of global function rather than cognitive assessment. Therefore, future research needs to confirm our findings with more robust cognitive assessment such as Montreal Cognitive Assessment or neuropsychological tests.

Our findings need to be interpreted within the context of the limitations of our data and analyses. First, the major limitation of this dataset is the size of the sample. It is a small convenience sample and as such the results should be considered hypothesis-generating. Our results should not be considered conclusive and replication of the analyses in a larger sample is required. Second, we did not collect data on the specific site, stage of GI cancer at diagnosis, and types of treatment received, therefore we cannot explain the functional changes with respect to the above factors. This information could help to explore whether functional limitations differ by these factors, which warrants future research. Third, our analyses did not include age and gender-matched comparator groups with CVD. Thus we could not justify whether the functional changes were attributed to GI cancer alone, ageing, or a combination of both. Fourth, our findings may not be generalizable as the data only apply to those with CVD, GI bleeding and GI cancer. Fifth, our analyses excluded those who died which may underestimate the disability.

This study is the first step towards enhancing our understanding of the functional changes among patients, with CVD and GI bleed history, before and after newly diagnosed GI cancer. Our results showed that individuals who developed GI cancer are at risk for decreased task participation and more difficulties with walking. Reduced occupational participation may result from decreased functional ability after GI cancer, highlighting the importance of assessing both task participation and difficulty in task performance in practice. Clinicians (oncologists, physiatrists, primary health care providers, nurses) should get a baseline understanding of functional performance at the time of GI cancer diagnosis and reassess regularly to understand if there are new functional limitations. Walking is the most affected area that clinicians and patients need to be aware of after GI cancer and have strategies to mitigate these impairments. Future research is needed to validate our findings and further understand patients' functional needs specific to their site, stage of GI cancer, and treatment type.

Conclusion

Our study suggests that those diagnosed with GI cancer may experience greater loss of function, specifically in walking. Since the survival rate of GI cancer improves, it is important to have a dialogue about issues that may be affecting functional independence.

Key messages

- GI cancer impairs individuals' ability to participate in meaningful occupations across the area of instrumental and basic activities of daily livings and mobility.
- The role of occupational therapy in cancer rehabilitation is important but often under-recognized.

- Occupational therapists are well-positioned to re-able and support patients' occupational participation beyond their ageing and cancer-associated impairments, improving their clinical and functional prognoses and life quality.

Reference

Abu-Helalah, M. A., Alshraideh, H. A., Mohammad, M. A.-H., & Arqoub, K. H. (2014). Quality of Life and Psychological Well-Being of Colorectal Cancer Survivors in Jordan. *Asian Pacific Journal of Cancer Prevention*, *15*(18), 7653–7664. http://journal.waocp.org/article_29817.html

Andrews, G., Kemp, A., Sunderland, M., Von Korff, M., & Ustun, T. B. (2009). Normative Data for the 12 Item WHO Disability Assessment Schedule 2.0. *PLOS ONE*, *4*(12), e8343. https://doi.org/10.1371/journal.pone.0008343

Arndt, V., Merx, H., Stegmaier, C., Ziegler, H., & Brenner, H. (2006). Restrictions in quality of life in colorectal cancer patients over three years after diagnosis: A population based study. *European Journal of Cancer*, *42*(12), 1848–1857. https://doi.org/https://doi.org/10.1016/j.ejca.2006.01.059

Arnold, M., Abnet, C. C., Neale, R. E., Vignat, J., Giovannucci, E. L., McGlynn, K. A., & Bray, F. (2020). Global Burden of 5 Major Types of Gastrointestinal Cancer. *Gastroenterology*, *159*(1), 335-349.e15. https://doi.org/https://doi.org/10.1053/j.gastro.2020.02.068

Bhatt, D. L., Fox, K. A. A., Hacke, W., Berger, P. B., Black, H. R., Boden, W. E., Cacoub, P., Cohen, E. A., Creager, M. A., Easton, J. D., Flather, M. D., Haffner, S. M., Hamm, C. W., Hankey, G. J., Johnston, S. C., Mak, K.-H., Mas, J.-L., Montalescot, G., Pearson, T. A., … Topol, E. J. (2006). Clopidogrel and Aspirin versus Aspirin Alone for the Prevention of Atherothrombotic Events. *New England Journal of Medicine*, *354*(16), 1706–1717. https://doi.org/10.1056/NEJMoa060989

Braamse, A. M. J., van Turenhout, S. T., Terhaar sive Droste, J. S., de Groot, G. H., van der Hulst, R. W. M., Klemt-Kropp, M., Kuiken, S. D., Loffeld, R. J. L. F., Uiterwaal, M. T.,

Mulder, C. J. J., & Dekker, J. (2016). Factors associated with anxiety and depressive symptoms in colorectal cancer survivors. *European Journal of Gastroenterology & Hepatology*, *28*(7). https://journals.lww.com/eurojgh/Fulltext/2016/07000/Factors_associated_with_anxiety_and_depressive.17.aspx

Canadian Cancer Society. (2019). *Canadian Cancer Statistics 2019*.

CAPRIE Steering committee. (1996). A randomised, blinded, trial of clopidogrel versus aspirin in patients at risk of ischaemic events. *Lancet*, *348*(9038), 1329–1339. https://doi.org/10.1016/s0140-6736(96)09457-3

Chau, I., Fuchs, C. S., Ohtsu, A., Barzi, A., Liepa, A. M., Cui, Z. L., Hsu, Y., & Al-Batran, S.-E. (2019). Association of quality of life with disease characteristics and treatment outcomes in patients with advanced gastric cancer: Exploratory analysis of RAINBOW and REGARD phase III trials. *European Journal of Cancer*, *107*, 115–123. https://doi.org/10.1016/j.ejca.2018.11.013

Dean, G., Ph, D., Nelson, R., Ph, D., & Chu, D. (2007). Quality of Life Outcomes in 599 Cancer and Non-Cancer Patients with Colostomies. *The Journal of Surgical Research*, *138*(1), 79–87. https://doi.org/10.1016/j.jss.2006.04.033

Den Bakker, C. M., Schaafsma, F. G., Huirne, J. A. F., Consten, E. C. J., Stockmann, H. B. A. C., Rodenburg, C. J., De Klerk, G. J., Bonjer, H. J., & Anema, J. R. (2018). Cancer survivors' needs during various treatment phases after multimodal treatment for colon cancer - Is there a role for eHealth? *BMC Cancer*, *18*(1), 1–12. https://doi.org/10.1186/s12885-018-5105-z

Donnell, M. J., Kapral, M. K., Fang, J., Saposnik, G., Eikelboom, J. W., Oczkowski, W., Silva,

J., Gould, L., D'Uva, C., & Silver, F. L. (2008). Gastrointestinal bleeding after acute ischemic stroke. *Neurology*, *71*(9), 650 LP – 655. https://doi.org/10.1212/01.wnl.0000319689.48946.25

Drury, A., Payne, S., & Brady, A.-M. (2017). The cost of survival: an exploration of colorectal cancer survivors' experiences of pain. *Acta Oncologica*, *56*(2), 205–211. https://doi.org/10.1080/0284186X.2016.1266084

Eikelboom, J. W., Connolly, S. J., Bosch, J., Shestakovska, O., Aboyans, V., Alings, M., Anand, S. S., Avezum, A., Berkowitz, S. D., Bhatt, D. L., Cook-Bruns, N., Felix, C., Fox, K. A. A., Hart, R. G., Maggioni, A. P., Moayyedi, P., O'Donnell, M., Rydén, L., Verhamme, P., … Yusuf, S. (2019). Bleeding and New Cancer Diagnosis in Patients With Atherosclerosis. *Circulation*, *140*(18), 1451–1459. https://doi.org/10.1161/CIRCULATIONAHA.119.041949

Forsberg, C., & Cedermark, B. (1996). Well-being, general health and coping ability: 1-year follow-up of patients treated for colorectal and gastric cancer. *European Journal of Cancer Care*, *5*(4), 209–216. https://doi.org/10.1111/j.1365-2354.1996.tb00237.x

Foster, C., Wright, D., Hill, H., Hopkinson, J., & Roffe, L. (2009). Psychosocial implications of living 5 years or more following a cancer diagnosis: A systematic review of the research evidence. *European Journal of Cancer Care*, *18*(3), 223–247. https://doi.org/10.1111/j.1365-2354.2008.01001.x

Government of Canada. (2019). *Colorectal cancer*. https://www.canada.ca/en/public-health/services/chronic-diseases/cancer/colorectal-cancer.html

International Agency for Research on Cancer. (n.d.). *Global Cancer Observatory: cancer today*. Retrieved January 15, 2020, from https://gco.iarc.fr/ today

Kobayashi, D., Kodera, Y., Fujiwara, M., Koike, M., Nakayama, G., & Nakao, A. (2011). Assessment of Quality of Life After Gastrectomy Using EORTC QLQ-C30 and STO22. *World Journal of Surgery*, *35*(2), 357–364. https://doi.org/10.1007/s00268-010-0860-2

Kucharska-Newton, A., Griswold, M., Yao, Z. H., Foraker, R., Rose, K., Rosamond, W., Wagenknecht, L., Koton, S., Pompeii, L., & Windham, B. G. (2017). Cardiovascular Disease and Patterns of Change in Functional Status Over 15 Years: Findings From the Atherosclerosis Risk in Communities (ARIC) Study. *Journal of the American Heart Association*, *6*, e004144. https://doi.org/10.1161/JAHA.116.004144

Lv, Y., Zhang, J., & Qiao, L. (2014). Quality of life in patients with esophageal cancer receiving definitive chemoradiotherapy or esophagectomy. *Molecular and Clinical Oncology*, *2*(5), 870–874. https://doi.org/10.3892/mco.2014.313

Marzona Irene, Yusuf, S., Koon, T., Bosch, J., & O'Donnell, M. (2011). *The Standard Assessment of Global Activities in the Elderly (SAGE) Scale: Validation Process of A New Tool for the Assessment of Disability in Older Adults* [McMaster University]. https://macsphere.mcmaster.ca/bitstream/11375/11287/1/fulltext.pdf

O'Gorman, C., Stack, J., O'Ceilleachair, A., Denieffe, S., Gooney, M., McKnight, M., & Sharp, L. (2018). Colorectal cancer survivors: an investigation of symptom burden and influencing factors. *BMC Cancer*, *18*(1), 1022. https://doi.org/10.1186/s12885-018-4923-3

Ogata, T., Kamouchi, M., Matsuo, R., Hata, J., Kuroda, J., Ago, T., Sugimori, H., Inoue, T., & Kitazono, T. (2014). Gastrointestinal Bleeding in Acute Ischemic Stroke: Recent Trends from the Fukuoka Stroke Registry. *Cerebrovascular Diseases Extra*, *4*(2), 156–164. https://doi.org/10.1159/000365245

Ramsey, S. D., Berry, K., Moinpour, C., Giedzinska, A., & Andersen, M. R. (2002). Quality of

life in long term survivors of colorectal cancer. *The American Journal of Gastroenterology*, *97*(5), 1228–1234. https://doi.org/10.1111/j.1572-0241.2002.05694.x

Rauch, P., Miny, J., Conroy, T., Neyton, L., & Guillemin, F. (2004). Quality of life among disease-free survivors of rectal cancer. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *22*(2), 354–360. https://doi.org/10.1200/JCO.2004.03.137

Rumalla, K., & Mittal, M. K. (2016). Gastrointestinal Bleeding in Acute Ischemic Stroke: A Population-Based Analysis of Hospitalizations in the United States. *Journal of Stroke and Cerebrovascular Diseases*, *25*(7), 1728–1735. https://doi.org/10.1016/j.jstrokecerebrovasdis.2016.03.044

Rutherford, C., Müller, F., Faiz, N., King, M. T., & White, K. (2020). Patient-reported outcomes and experiences from the perspective of colorectal cancer survivors: meta-synthesis of qualitative studies. *Journal of Patient-Reported Outcomes*, *4*(1), 27. https://doi.org/10.1186/s41687-020-00195-9

Schneider, E. C., Malin, J. L., Kahn, K. L., Ko, C. Y., Adams, J., & Epstein, A. M. (2007). Surviving colorectal cancer. *Cancer*, *110*(9), 2075–2082. https://doi.org/https://doi.org/10.1002/cncr.23021

Sodergren, S. C., Wheelwright, S. J., Permyakova, N. V., Patel, M., Calman, L., Smith, P. W. F., Din, A., Richardson, A., Fenlon, D., Winter, J., Corner, J., Foster, C., Armes, J., Baird, J., Bateman, A., Beck, N., Moon, G., Hall, P., Poole, K., … Williams, F. (2019). Supportive care needs of patients following treatment for colorectal cancer: risk factors for unmet needs and the association between unmet needs and health-related quality of life—results from the ColoREctal Wellbeing (CREW) study. *Journal of Cancer Survivorship*, *13*(6), 899–909.

https://doi.org/10.1007/s11764-019-00805-6

Spence, J., Bosch, J., Chongsi, E., Lee, S. F., Thabane, L., Mendoza, P., Belley-Côté, E., Whitlock, R., Brady, K., McIntyre, W. F., Lamy, A., & Devereaux, P. J. (2021). Standardized Assessment of Global activities in the Elderly scale in adult cardiac surgery patients. *British Journal of Anaesthesia*. https://doi.org/10.1016/j.bja.2021.05.037

StataCorp. (2017). *Stata Statistical Software: Release 15* (p. College Station, TX: StataCorp LLC.). StataCorp LLC.

Stauder, M. C., Romero, Y., Kabat, B., Atherton, P. J., Geno, D., Deschamps, C., Jatoi, A., Sloan, J. A., Botros, M., Jung, K. W., Arora, A. S., & Miller, R. C. (2013). Overall survival and self-reported fatigue in patients with esophageal cancer. *Supportive Care in Cancer*, *21*(2), 511–519. https://doi.org/10.1007/s00520-012-1537-1

Sun, V. C.-Y., & Sarna, L. (2008). Symptom management in hepatocellular carcinoma. *Clinical Journal of Oncology Nursing*, *12*(5), 759–766. https://doi.org/10.1188/08.CJON.759-766

Viborg, S., Søgaard, K. K., Farkas, D. K., Nørrelund, H., Pedersen, L., & Sørensen, H. T. (2016). Lower Gastrointestinal Bleeding And Risk of Gastrointestinal Cancer. *Clinical and Translational Gastroenterology*, *7*(4), e162–e162. https://doi.org/10.1038/ctg.2016.16

Yale Medicine. (n.d.). *Fact Sheets - Gastrointestinal Cancers: Symptoms, Diagnosis and Treatment*. Retrieved August 4, 2021, from https://www.yalemedicine.org/conditions/gastrointestinal-cancers

Tables and figures

Table 1.

Baseline sample characteristics at the study entry (N=33)

Characteristics	With pre and post GI cancer SAGEA completed; N=26	With baseline SAGEA completed only; N=7
Sociodemographic characteristics		
Years of age, Mean (Standard deviation)	79 (9.2)	73 (9.7)
Male sex, n (%)	17 (65.4)	6 (85.7)
Ethnicity, n (%)		
Asian	0 (0)	1 (14.3)
Indigenous	1 (3.9)	0 (0)
White	25 (96.2)	6 (85.7)
Education, n (%)		
1-8 years	5 (19.2)	1 (14.3)
9-12 years	8 (30.8)	5 (71.4)
Trade school	3 (11.5)	0 (0)
College/ University	10 (38.5)	1 (14.3)
Living condition, n (%)		
At home (With others)	20 (76.9)	6 (85.7)
At home (Alone)	3 (11.5)	0 (0)
In retirement home	3 (11.5)	1 (14.3)
Details of the GI bleeding experienced at the study entry (Data collected after the baseline visit)		
Types of GI bleeding, n (%)		
Hematemesis	1 (3.8)	1 (14.3)
Melena	18 (69.2)	6 (85.7)
Hematochezia	12 (46.2)	1 (14.3)
Site of GI bleeding, n (%)		
Esophageal	2 (7.7)	2 (28.6)
Gastric	12 (46.2)	3 (42.9)
Duodenal	4 (15.4)	1 (14.3)
Large intestine/ Colon	10 (38.5)	3 (42.9)
Rectal – non-hemorrhoidal	4 (15.4)	0 (0)
Rectal – Hemorrhoidal	1 (3.9)	0 (0)
Unknown	1 (3.9)	1 (14.3)
Pathology of GI bleed, n (%)		
Esophagitis	1 (3.9)	1 (14.3)
Ulcer	5 (19.2)	2 (28.6)
Polyp	3 (11.5)	1 (14.3)
Malignancy	12 (46.2)	2 (28.6)
Diverticulosis	1 (3.9)	0 (0)
Esophageal varices	0 (0)	1 (14.3)
Unknown cause	1 (3.9)	0 (0)
Other pathology of the bleed	8 (30.8)[a]	1 (14.3)[b]
Severity of GI bleeding, n (%)		
Inpatient stay required	21 (80.8)	6 (85.7)
Stay in the intensive care unit/ critical care unit required	2 (7.7)	2 (28.6)
Intravenous inotropic support required	1 (3.9)	1 (14.3)
Syncope experienced	3 (11.5)	0 (0)

[a] Other pathology of the bleed: Lesion - Esophageal and gastric lesion, Cameron lesions; Inflammation - Duodenitis, Ischemic colitis, Erosive gastritis; Vascular abnormalities - Hemorrhoids, Arteriovenous malformation; Others - Gastric mass, Large mass with friable surface, hiatus hernia. [b] Other pathology of the bleed: Ulcerated firm mass with fixable surfaces

Table 2.

Participants' medical history of cardiovascular disease, its risk factors and other diseases at study entry

	With pre and post GI cancer SAGEA completed; N=26	With baseline SAGEA completed only; N=7
Cardiovascular		
Coronary artery disease		
Myocardial infarction	10 (38.5)	3 (42.9)
Angina	2 (7.8)	2 (28.6)
Coronary revascularisation [a]	11 (42.3)	2 (28.6)
Total N conditions	23	7
Cerebrovascular disease		
Stroke/ Transient ischemic attack	7 (26.9)	0 (0)
Heart failure	7 (26.9)	3 (42.9)
Atrial fibrillation/flutter	15 (57.7)	3 (42.9)
Venous thromboembolism	2 (7.7)	1 (14.3)
Total N conditions	31	7
Peripheral arterial diseases		
Valve replacement	2 (7.7)	1 (14.3)
Asymptomatic carotid artery stenosis	4 (15.4)	1 (14.3)
Peripheral artery bypass surgery	1 (3.9)	0 (0)
Intermittent claudication	2 (7.7)	2 (14.3)
Total N conditions	9	4
Metabolic		
Hypertension	17 (65.4)	3 (42.9)
Diabetes	10 (38.5)	3 (42.9)
Dyslipidemia	15 (57.7)	2 (28.6)
Non-cardiovascular		
Cancer	8 (30.8)[b]	1 (14.3)[c]
Renal dysfunction	4 (15.4)	2 (28.6)
Dementia	0 (0)	0 (0)
Anemia	16 (61.5)	2 (28.6)
Gastrointestinal		
Abdominal or pelvic radiation	2 (7.7)	1 (14.3)
Abdominal surgery	13 (50)	3 (42.9)
Ulcers	3 (11.5)	1 (14.3)
Diverticular disease	2 (7.7)	1 (14.3)
Helicobacter pylori	2 (7.7)	0
Hemorrhoids	11 (42.3)	2 (28.6)
Liver disease	0 (0)	1 (14.3)
Varices	0 (0)	1 (14.3)

[a] Coronary revascularisation includes coronary percutaneous transluminal coronary angioplasty, Atherectomy, Percutaneous coronary intervention, and Coronary Artery Bypass Graft Surgery.
[b] Cancer types: Brain cancer (n=4), lung cancer (n=1), bladder cancer (n=1), hematologic cancer (n=2); Year of cancer onset: 1975-2004, missing data for 3 participants.
[c] Cancer type: Prostate cancer (n=1); Year of onset: 1980

Table 3.

The Standard Assessment of Global Activities in the Elderly (SAGEA) Scores pre- and post- GI cancer and paired comparison between timepoints (N=26)

	Pre GI cancer		Post GI cancer		P- value
	Range	Median (IQR)	Range	Median (IQR)	
Overall	0-14	3.5 (6)	0-18	3 (7)	0.47
cADLs	0-5	0 (2)	0-6	0 (2)	0.59
iADLs	0-2	0 (1)	0-3	0 (1)	0.88
Mobility	0-3	1 (2)	0-3	1 (2)	0.98
bADLs	0-9	0 (2)	0-9	0.5 (3)	0.18

Note. Abbreviation: IQR – Interquartile range, cADLs – Cognitive activities of daily living (Out of 9), iADLs – Instrumental activities of daily living (Out of 3), mobility (Out of 3), bADLs – Basic activities of daily living (Out of 9)

Table 4.

Level of Difficulty Performing Tasks After GI Cancer Diagnosis

Tasks	N participants												
	Pre GI cancer	At Post GI cancer											
	Performed the task	No longer performed the task		Continued the task		Continued with less difficulty		Continued with no change in difficulty		Continued with more difficulty		No longer performed/ Continued with more difficulty	
	N	N	%	N	%	N	%	N	%	N	%	N	%
IADLs													
Playing games/reading that requires concentration	23	9	39.1	14	60.9	1	4.3	12	52.2	1	4.3	10	43.5
Finding your way around a new building	18	11	61.1	7	38.9	2	11.1	4	22.2	1	5.6	12	66.7
Organising a trip/ social activities	13	8	61.5	5	38.5	0	0	5	38.5	0	0	8	61.5
Doing your own finances/shopping	22	6	27.3	16	72.7	1	4.5	13	59.1	2	9.1	8	36.4
Organising medications	23	1	4.3	22	95.7	2	8.7	17	73.9	3	13	4	17.4
Preparing a meal and/or doing laundry	18	9	50	9	50	2	11.1	7	38.9	0	0	9	50
Driving	15	5	33.3	10	66.7	1	6.7	9	60	0	0	5	33.3
Using public transportation	7	2	28.6	5	71.4	0	0	5	71.4	0	0	2	28.6
Mobility													
Using stairs	22	4	18.2	18	81.8	8	36.4	7	31.8	3	13.6	7	31.8
Walking	26	1	3.8	25	96.2	5	19.2	9	34.6	11	42.3	12	46.2
BADLs													
Dressing	25	0	0	25	100	2	8	19	76	4	16	4	16
Transferring from bed to chair	23	3	13	20	87	0	0	16	69.6	4	17.4	7	30.4
Bath/toileting	25	0	0	25	100	5	20	14	56	6	24	6	24

Abbreviations: IADLs – Instrumental activities of daily living, BADLs – Basic activities of daily living

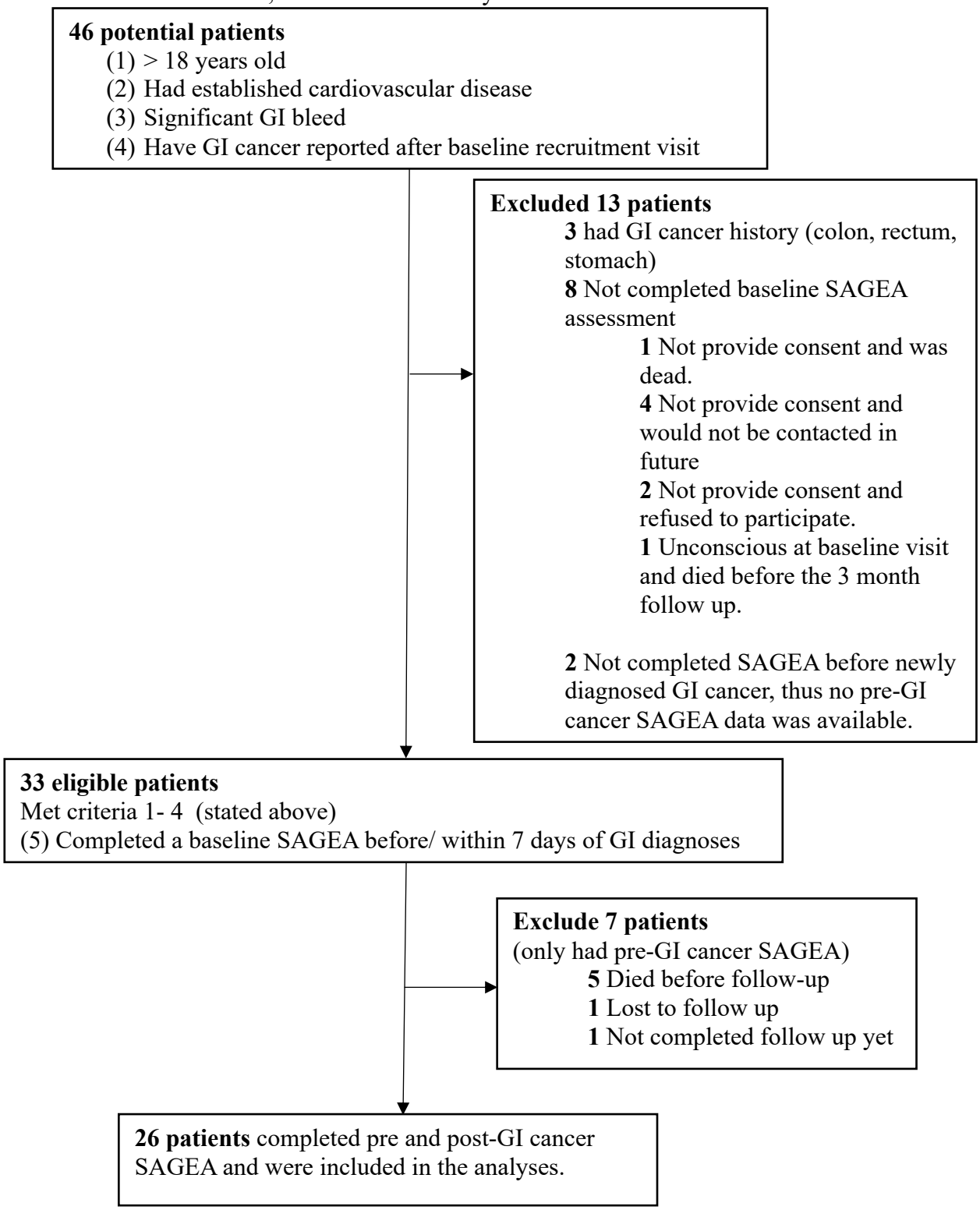


Figure 1. Flow chart on the process of retrospective selection of patients for inclusion.

Note. Abbreviation: GI – Gastrointestinal, SAGEA – Standard Assessment of Global Everyday Activities.

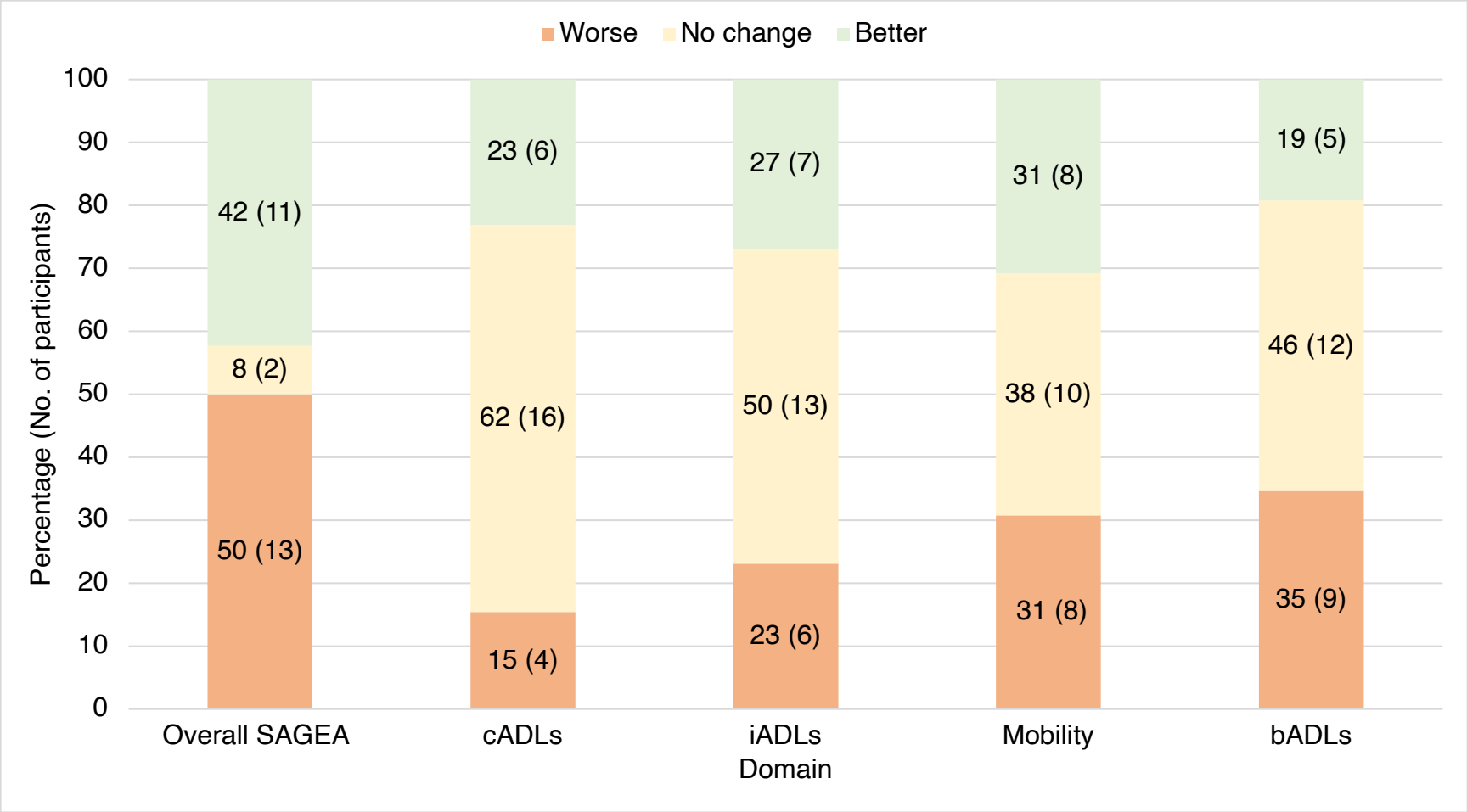


Figure 2. Change in function after GI cancer diagnosis.

Note. Abbreviation: SAGEA - Standard Assessment of Global Activities in the Elderly, cADLs – cognitive activities of daily living, iADLs – instrumental activities of daily living, bADLs – basic activities of daily living.

Supplemental files

Supplemental file 1

Items and scoring for the Standard Assessment of Global Activities in the Elderly (SAGEA)

Items and individual item scoring

Over the past month, did you have any difficulties with the following:

Item scoring: none (0) or some -> if some, then mild (1), moderate (2), or severe (3)

1. Keeping your attention or 'train of thought' during a conversation?
2. Remembering things that happened a few days before? (e.g. conversation, people visiting)
3. Ability to switch between things that are happening at the same time? (e.g. making tea and talking to someone)

Over the past month, did you perform any of the following activities:

Item scoring: no (0) or yes -> if yes, difficulty? none (0) mild (1), moderate (2), or severe (3).

If ALL items 4-10 were not performed -> give the maximum score of 3 point.

4. Playing a game or reading a book that requires concentration (e.g. of games: crosswords, checkers, chess)
5. Finding your way around a new building? (e.g. hospital/clinic)
6. Organizing a trip or social activities? (e.g. vacation or family occasion) (score the activity that the person finds to be the more difficult of the two)
7. Doing your own finances or shopping? (score the activity that the person finds to be the more difficult of the two)
8. Organizing and taking your medications?
9. Preparing a meal and/or doing laundry? (score the activity that the person finds to be more difficult of the two)
10. a) Driving? Do not drive (go to 10b)

10. b) Using public transportation? Do not use (go to 11)

Over the past month, did you perform any of the following activities:

Item scoring: no (0) or yes -> if yes, difficulty? none (0) mild (1), moderate (2), or severe (3). If requires help, add 1 point to maximum score of 3 points for that item. If person did not do item 12, score 3 points.

11. Using stairs? (one flight) If yes, did you require help?

10. Walking? (about 10m or 32ft or 14 steps) If yes, did you require help?

Over the past month, did you perform any of the following activities:

Item scoring: no (3) or yes -> if yes, difficulty? none (0) mild (1), moderate (2), or severe (3). If requires help, add 1 point to a maximum score of 3 points for that item. If person did not do activity, score 3 points.

11. Dressing? If yes, did you require help?
12. Transferring from bed to chair? If yes, did you require help?
13. Bathing or toileting? (score the activity that the person finds to be more difficult of the two)

Overall Scoring:

Total score is computed by adding the 4 subscores. Range 0 (no difficulty) to 24 (severe difficulty).

Cognitive (c) ADL subscore: Add scores for items 1-3. Range 0 (no difficulty) to 9 points (severe difficulty)

Instrumental (i) ADL: Compute maximum score for items 4-10. If all of items 4-10 are answered "no", i.e. person did not do any of these activities, score as 3 points. Range 0 (no difficulty) to 3 points (severe difficulty).

Mobility subscore: Compute maximum score for items 11-12. Range 0 (no difficulty) to 3 points (severe difficulty).

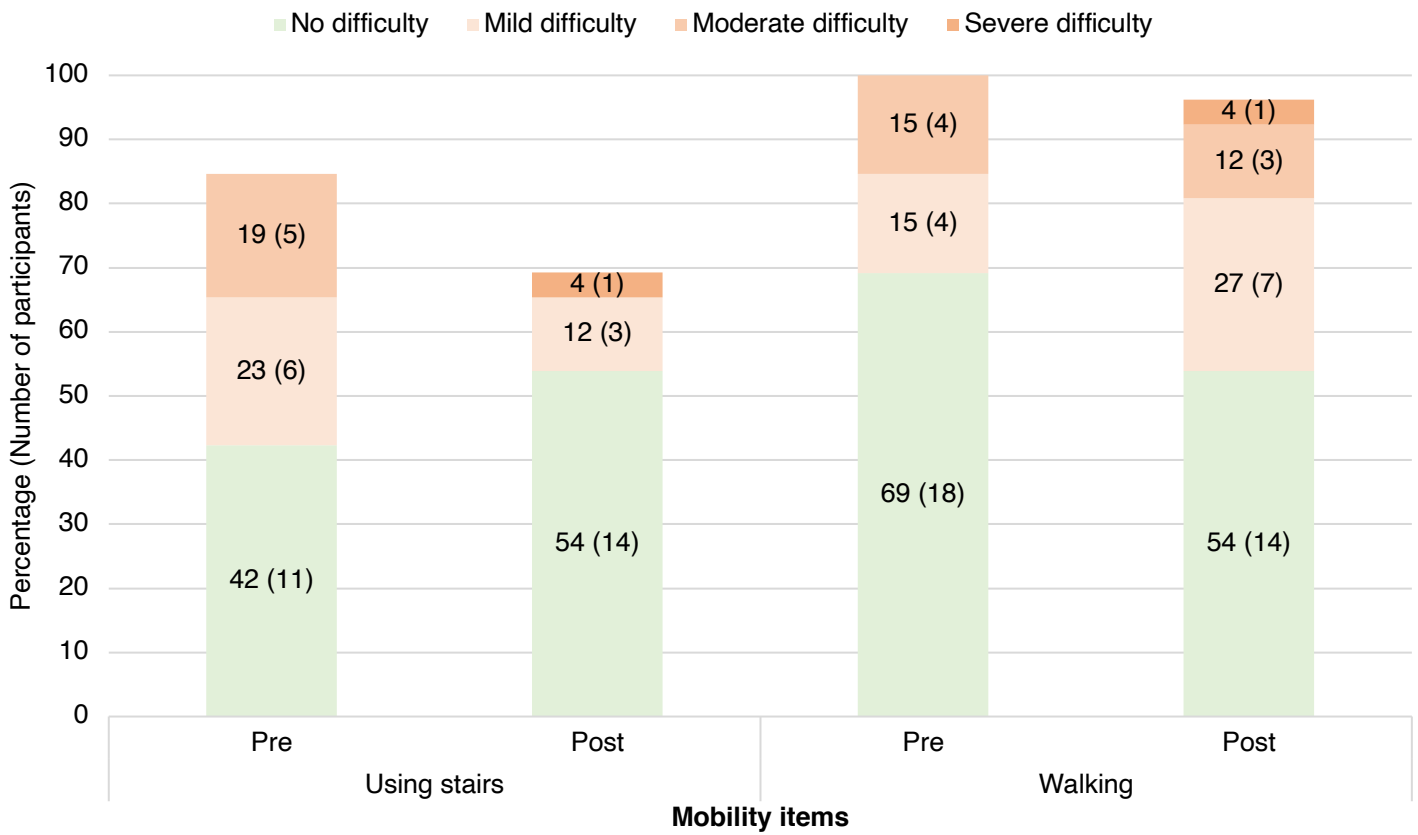
Basic (b) ADL subscore: Add scores for items 13-15. Range 0 (no difficulty) to 9 points (severe difficulty).

Supplemental table 1.

The Standard Assessment of Global Activities in the Elderly (SAGEA) Scores and the number of performed tasks pre- and post- GI cancer among those who performed the same or more tasks (n=8) and those who performed fewer tasks (n=18)

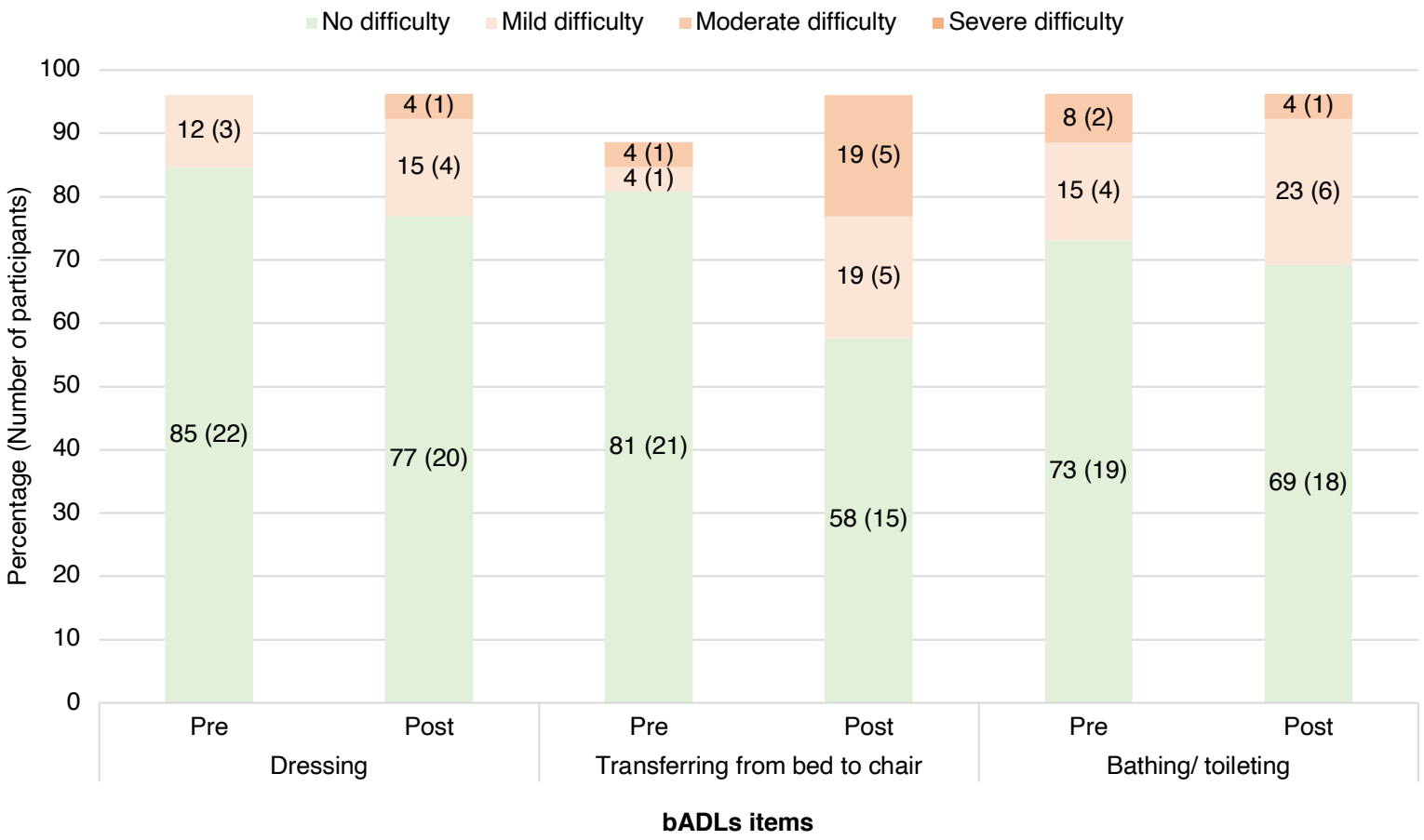
Characteristics	Performed the same number of / more tasks; n=8	Performed less tasks; n=18
SAGEA score		
Pre GI cancer, Median (IQR)	5.5 (6.5)	2.5 (4)
Post GI cancer, Median (IQR)	3 (7.5)	3.5 (6)
Difference, Median (IQR)	-1 (4.5)	+2 (5)
Number of tasks performed		
Pre GI cancer, Median (IQR),	9.5 (2.5)	11 (1)
Post GI cancer, Median (IQR)	10 (2)	8 (4)
Difference, Median (IQR)	+0.5 (1.5)	-3 (4)

Note. Abbreviation: SD – Standard deviation. IQR – Interquartile Range, SAGEA – Standard Assessment of Global Activities in the Elderly.



Supplemental figure 1. 1 Mobility-related tasks: Difficulty in performing at pre and post GI cancer states.

Note. Abbreviation: Pre – Pre GI cancer, Post – Post GI cancer.



Supplemental figure 1. 2 bADLs (basic activities of daily living)-related tasks: Difficulty in performing at pre and post GI cancer states.

Note. Abbreviation: Pre – Pre GI cancer, Post – Post GI cancer.

Chapter 3

Systematic review: Patient-reported outcome measures of fatigue in inflammatory bowel disease

Systematic review: Patient-reported outcome measures of fatigue in inflammatory bowel disease

Vanessa S.K. Fan [1], Ava Mehdipour [1], Heather K. O'Grady [1], Julie Richardson [1,2], Lauren E. Griffith [2], Kristin L. Campbell [3], Jackie Bosch[1,4]

[1] School of Rehabilitation Science, McMaster University

[2] Department of Health Research Methods, Evidence, and Impact, McMaster University

[3] Department of Physical Therapy, The University of British Columbia

[4] Population Health Research Institute

Corresponding author: Jackie Bosch

This manuscript is submitted to the *Inflammatory Bowel Diseases*.

Financial disclosures

None.

Conflict of interest statement

No conflict of interest was declared.

Summary

Fatigue is an important construct in inflammatory bowel disease (IBD) but not commonly measured. This review identified the Functional Assessment of Chronic Illness Therapy-Fatigue and IBD-fatigue scale as the most psychometrically robust and feasible measures for research and clinical use.

Abstract

Background: Fatigue is common for inflammatory bowel disease (IBD) patients, however there is no consensus on the optimal measurement tool to assess fatigue.

Aims: To identify standardized patient-reported outcome measures (PROMs) of fatigue in adults with IBD, evaluate their psychometric properties, and recommend PROM(s) that is/are the most psychometrically robust and feasible for research and clinical use.

Methods: Eight databases were first searched from 2015-2020 to identify fatigue PROMs used in IBD clinical research. Studies on each PROMs' psychometric properties were then searched in MEDLINE, EMBASE, CINAHL, and PsycINFO from inception to September 2020. The study quality, psychometric robustness of the PROMs, and quality of evidence were evaluated following the COnsensus-based Standards for the selection of health Measurement INstruments guideline.

Results: First search included 111 articles and identified 16 PROMs. Second search identified 9 articles on psychometric properties for 8 PROMs only. All 8 PROMs demonstrated validity (with moderate-high quality evidence) and 5 PROMs demonstrated reliability (intraclass correlation coefficient: 0.65-0.98; with very low-low quality evidence). Content validity (with high quality evidence) was only demonstrated for the IBD-Fatigue (IBD-F) scale (English). Responsiveness (with high quality evidence) and the minimal clinically important changes were only demonstrated in the Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F). Evidence on content validity, structural validity, cross-cultural validity, internal consistency, and responsiveness of PROMs was lacking for most PROMs.

Conclusion: The FACIT-F and IBD-F(English) are recommended for research and clinical use. Further research is required to establish responsiveness and minimal clinically important

change.

Keywords

Inflammatory bowel disease, Crohn's disease, ulcerative colitis, fatigue, patient-reported outcome measures

Introduction

Canada has the highest prevalence of inflammatory bowel disease (IBD) in the world,[1] affecting 270,000 Canadians.[2] IBD is chronic inflammation in the gastrointestinal tract of unknown aetiology, which can be broadly categorized into Crohn's disease and ulcerative colitis. Fatigue is the second most commonly reported concern by IBD patients, second to diarrhea, in both active and inactive states.[3] The impact of fatigue on an individual's physical, cognitive, and emotional abilities in turn affects social and daily functioning.[4] Subsequently, fatigue impaired health-related quality of life in IBD patients. [5]

In 2017, the management of IBD-fatigue was recognised as the top research and clinical priority by the nurses European Crohn's and Colitis Organisation and IBD patients through international Delphi survey and priority setting activities. [6,7] However, research on interventions is limited.[8] Identification of an appropriate outcome measure for clinical research of fatigue may assist researchers in better testing the efficacy of potential intervention approaches; however, assessing fatigue is challenging, at least in part due to a lack of consensus on the definition of fatigue.[9] For this review we have used the definition by Van Langenberg et al. whose research has been focused on the IBD-related fatigue; (1) physical fatigue that results in the inability to initiate and complete certain activities; (2) cognitive fatigue that results in impaired concentration and loss of memory, or (3) emotional/affective fatigue that leads to a decrease in motivation and mood,[10] which is not resolved by prolonged rest or sleep.[11] Fatigue can be measured as an unidimensional or multidimensional construct.[12] Unidimensional measures focus on 1 dimension (e.g. the impact of fatigue) and are usually shorter than multidimensional measures. A conceptual framework of IBD-related fatigue is proposed to illustrate that fatigue can be measured by its experience and impact (see Figure 1).

Fatigue can be measured by both performance-based and patient-reported outcome measures (PROMs).[9] Performance-based fatigue outcome measures evaluate physical and

cognitive functional changes that are attributable to fatigue.[10] One example is an isokinetic dynamometer that measures the decrease in quadriceps muscle contraction over a prolonged period of time, as an indicator of muscle fatigue.[9] PROMs capture the physical, cognitive, and emotional impacts of fatigue from the patients' perspective. These PROMs can provide a holistic picture of the patients' perception of the day-to-day functional issues associated with fatigue that performance-based measures do not provide. Therefore, this review focused on PROMs of IBD-related fatigue. Numerous unidimensional and multidimensional PROMs are used to assess fatigue and the effectiveness of interventions for fatigue in IBD research, but they are not commonly used in clinical practice,[14] perhaps because of a lack of clarity as to which outcome measure is reliable, valid, and responsive as well as feasible to administer.[13,15]

In this review, we aimed to determine the most robust PROM(s) to evaluate IBD-fatigue in research and clinical settings. Psychometrically robust PROMs should be reliable, valid, and responsive to change.[16,17] In addition to the psychometric robustness, PROMs should be clinically useful and feasible (i.e. easy to administer, analyse and interpret scoring; short completion time [<10 minutes]; low patient burden).[18] To address this aim, we have (1) identified standardised fatigue PROMs used in adult IBD clinical research; (2) identified studies on the psychometric properties of the PROMs, (3) appraised the methodological quality of studies and the psychometric properties of the PROMs; and (4) made recommendations regarding the most robust measure(s) for research and clinical use.

Methods

The protocol for this review was registered on the International Prospective Register of Systematic Reviews (PROSPERO) database (Registration no.: CRD42020204033).

Identifying fatigue PROMs

The search strategy to identify fatigue PROMs used in studies of adults with IBD was developed in consultation with a librarian (Details on the search strategy can be found in text, Supplementary Data Content 1). Peer-reviewed articles, abstracts, and conference proceedings were searched in the following databases: MEDLINE (OVID), EMBASE (OVID), CINAHL, PsycINFO (OVID), Cochrane Library, Allied and Complementary Medicine Database (AMED), Health and Psychosocial Instruments, and Mental Measurement Yearbook from 2015-September 2020, to identify PROMs that are currently used in research. Articles were included if they included; (1) adults (age ≥16) with IBD, and (2) a standardized fatigue PROM (described or used). The PROMs were included if their construct measured was fatigue. We translated the abstracts in non-English articles and included the article if it included the name of the fatigue PROM. Titles and abstracts were screened by one reviewer (VF). Full-text articles were screened for PROMs independently by three reviewers (VF, AM, HO), who identified the same PROMs (100% agreement).

Identifying studies on the psychometric properties of fatigue PROMs

Once the PROMs were identified, we searched for studies on the psychometric properties of each identified PROM in the adult IBD population. Search terms included the name of the PROMs (and if appropriate, the abbreviation), IBD, and the psychometric properties (including reliability, validity, responsiveness, interpretability, and feasibility) (Details on the search strategy can be found in text, Supplementary Data Content 2). MEDLINE (OVID), EMBASE (OVID), CINHAL (via ProQuest), and PsycINFO were searched from inception to September 17, 2020, recognizing that psychometric studies would precede the use of PROMs in research. Studies were included if they; (1) were published in English, (2) tested the psychometric properties of the identified PROMs, and (3) included an adult (age ≥16) IBD population. Abstracts without full articles were ineligible. For PROMs not developed in the IBD

population, the original article on the development of the PROM was identified and hand-searched for information on the construct(s) being measured, recall period, and scoring of the PROMs. Title, abstract, and the full-text of articles were independently screened by two reviewers (VF, AM), and a third reviewer (JB) resolved conflicts when there was disagreement. The interrater percent agreement for full-text review was 97.3%. Studies on the normative values and general information on the administration of measures were hand-searched. Two reviewers (VF, AM) independently extracted the following data according to the COnsensus-based Standards for the selection of health status Measurement INstrument (COSMIN) guideline: characteristics of the included samples, characteristics of the PROMs (including the construct(s), target population, recall period, subscales, number of items, description of score), psychometric properties, score interpretability, and administration of the PROMs. [19]

Evaluation of methodological quality and the psychometric properties of PROMs

The COSMIN guideline was used to structure the evaluation of the methodological quality of studies and the psychometric properties of the PROMs.[19] We first assessed the methodological quality of the studies, then the robustness of the psychometric property for the PROM, and finally rated the overall quality of evidence for each psychometric property. Details on how each process was conducted are provided below.

i. Evaluation of the methodological quality of studies

Evaluation of the study quality was guided by the COSMIN Risk of Bias Checklist.[20] The Risk of Bias Checklist was developed by international measurement experts and we used it to evaluate study quality based on the standards for study design and preferred statistical methods for each psychometric property.[20,21] The Risk of Bias Checklist has been shown to have an adequate interrater agreement, with two-thirds of the evaluators agreeing on about 80% of the

items.[22] We examined studies on the following 8 psychometric properties: content validity, construct validity (convergent and known-groups validity), structural validity, cross-cultural validity, internal consistency, test-retest reliability, measurement error, and responsiveness of PROMs. There is no accepted gold standard for measuring IBD-fatigue, therefore criterion validity of PROMs was not examined. Two reviewers (VF, AM) evaluated each study quality independently, assigning an overall rating (very good, adequate, doubtful or inadequate) based on the lowest rating for any of the standards for each psychometric property.[20]

ii. Evaluation of the psychometric robustness of PROMs

We determined whether the results of the psychometric property study demonstrated adequate psychometric robustness using the COSMIN Criteria for Good Measurement Property. [19]

Each result was rated as either sufficient (+), insufficient(-), or indeterminate(?) for each psychometric property (See table, Supplementary Data Content 3, which lists the updated COSMIN Criteria for Good Measurement Property).[19] Each translated measure or subscale of a multidimensional PROM was evaluated separately.[19] Subscales were evaluated separately if part of a multidimensional measure that did not use a total summative score, but instead it was intended that each subscale is scored separately.

The PROM was considered to have adequate psychometric robustness based on the following criteria: for content validity, there was evidence that the items are relevant, comprehensive, and comprehensible to IBD patients; for construct validity (convergent and known-groups validity), if 75% of the study results met our hypotheses on the relationship between measures (See table, Supplementary Data Content 4, which lists the specific correlations for each PROM and subscale with different sub-construct of fatigue) [19,23]; for convergent validity testing for PROMs and subscales, our review team created the minimally

acceptable correlations for each sub-constructs of fatigue, recognizing that some constructs are likely more correlated (See table, Supplementary Data Content 4, which lists the specific correlations for each PROM and subscale with different sub-construct of fatigue); for known-groups validity, subgroup with active IBD state should have a statistically higher level of fatigue than those in inactive IBD state or healthy individuals; for structural validity, there needed to have evidence on exploratory or confirmatory factor analysis.[19,23] Cross-cultural validity was examined for the translated measures. For cross-cultural validity, a similar item response between different language groups was required [19,23]; for internal consistency, evidence on structural validity to support the unidimensionality of (sub)scale and the Cronbach's alpha ≥ 0.7 for each unidimensional (sub)scale was required [19,23]; for reliability, evidence of an intraclass correlation coefficient or weighted kappa ≥ 0.7 [19,23] was required or measurement error had to be less than the minimal important change [19,23]; A PROM was considered to be responsive if 75% of the study results met our hypotheses on the relationship between change scores (See table, Supplementary Data Content 4, which lists the specific correlations for each PROM and subscale with different sub-construct of fatigue).[19,23] Two reviewers (VF, AM) independently evaluated the psychometric properties. If there was any disagreement in the ratings between reviewers (VF,AM), a third reviewer (JB) was consulted to make a final decision.

After evaluating each result on the psychometric property study, the results from the individual studies were then descriptively summarised and evaluated for each psychometric property of each PROMs. A final rating, based on the COSMIN Criteria for Good Measurement Property, was provided for the overall psychometric robustness (sufficient [+], insufficient[-], indeterminate[?], or inconsistent [±]).[19] If there was not enough information to evaluate the psychometric property, an indeterminate rating would be given.[19] Unexplained inconsistencies across results led to a rating of inconsistent.[19] The psychometric robustness of the

multidimensional PROMs was rated based on the majority of the results (>75%) from the subscales.

iii. Evaluation of the overall quality of evidence on the psychometric robustness

The overall quality of evidence on the psychometric robustness for each PROM was assessed using the modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, in which a rating of high, moderate, low, or very low, was assigned to each psychometric property.[19] The quality of evidence was assumed to be high and downgraded by different levels based on the risk of bias from the study quality, the pooled sample size, and the consistency of results (See table, Supplementary Data Content 5, which shows the criteria for evaluating the quality of evidence using the modified GRADE criteria).[19] The quality of evidence was not evaluated for those psychometric properties with *only* indeterminate or inconsistent ratings.[19] The quality of evidence on the psychometric robustness of multidimensional PROMs was based on the majority of the results (>75%) from the subscales.

The results on psychometric robustness and the quality of evidence were interpreted summatively. A high quality of evidence on sufficient/insufficient psychometric robustness means that 'We are very confident that the estimated psychometric property is sufficient/insufficient and close to the true psychometric property'. [19,24] A moderate quality of evidence on sufficient/insufficient psychometric robustness means that 'We are moderately confident that the estimated psychometric property is sufficient/ insufficient and likely to be close to the true psychometric property'. [19,24] A low quality of evidence on sufficient/insufficient psychometric robustness means that 'We have limited confidence that estimated psychometric property is sufficient/ insufficient and may be substantially different from the true psychometric property'. [19,24] A very low quality of evidence on sufficient/ insufficient quality psychometric robustness means that 'We have very little confidence that

the estimated psychometric property is sufficient/ insufficient and is likely to be substantially different from the true psychometric property'. [19,24]

Recommendation of measures

The most 'robust' PROM for research use should demonstrate: (i) at least moderate quality evidence on sufficient content validity because content validity (i.e. Items are relevant, comprehensive, and comprehensible to assess fatigue in the IBD population) is the most important psychometric property suggested by the COSMIN guideline [19]; (ii) at least moderate quality evidence on any type of sufficient validity, including construct, structural, and cross-cultural validity; (iii) at least moderate quality evidence on any type of sufficient reliability, including internal consistency, test-retest reliability, and measurement error; (iv) at least moderate quality evidence on sufficient responsiveness; and (v) data on the minimal clinical important change so that researchers can understand whether the change in fatigue assessment is clinically meaningful beyond the statistical significance. In addition to the psychometric robustness, the most 'robust' PROM for clinical use should be clinically feasible, which is: easy to access; available in multiple administrative modes or translated versions of the measure with the evidence to demonstrate equal psychometric robustness across; understandable for people with low literacy or non-English speakers; does not require training to administer, score and interpret scoring; and can be completed within a short time (<10 minutes) with a low patient burden. [18]

Results

IBD-related Fatigue PROMs

The results of the search for PROMs are summarised in Figure 2. In short, 2718 articles were screened, of which 2369 were excluded (primarily because the articles were not related to the IBD population), and 349 had a full-text review. Full-text review excluded an additional 238 articles (primarily because the name of the fatigue measure was not provided), leaving 111

articles that identified 16 IBD-related fatigue PROMs. The identified PROMs were the Fatigue Severity Scale (FSS), the Daily Fatigue Impact Scale (D-FIS), the Functional Assessment of Chronic Illness Therapy- Fatigue (FACIT-F), the Inflammatory Bowel Disease Fatigue Scale (IBD-F), the Multidimensional Fatigue Inventory-20 (MFI-20), the Multidimensional Assessment of Fatigue (MAF), the Modified Fatigue Impact Scale (M-FIS), and the Fatigue Questionnaire (FQ), the Checklist Individual Strength, the Piper Fatigue Scale, the Revised Piper Fatigue Scale, the Patient-Reported Outcomes Measurement Information System – Fatigue, the Fatigue Symptom Inventory, the Visual Analogue Scale of Fatigue, the Brief Fatigue Inventory, and the Fatigue Impact Scale.

Psychometric data on the IBD-related fatigue PROMs

The search results for articles describing the psychometric testing of each PROM are summarised in Figure 3. In short, 558 articles were screened, of which 370 were excluded based on relevance, resulting in full-text review of 188 articles. Full-text review excluded an additional 175 articles. Nine articles, reporting on data from 8 studies, described the psychometric properties of 8 of the 16 IBD-related fatigue PROMs. Studies on the psychometric data were available for the FSS-Spanish version (1 study) [5]; the D-FIS-Spanish version (1 study) [5]; the FACIT-F(1 study) [25]; the IBD-F (5 studies: 2 for English [26,27], each for Brazilian-Portuguese [28], Greek [29], and Danish [30]); the MFI-20 (1 study) [26]; the MAF (1 study) [26]; the M-FIS-Spanish version (1 study) [5]; and the FQ-Norwegian version (1 Study and 1 Supplementary article [31,32]). The 8 PROMs without psychometric data were not included in this review (Details can be found in Figure 2).

Characteristics of the IBD-related fatigue PROMs

The characteristics of the 8 PROMs are described in Table 1. Three of the PROMs (FSS, D-FIS, FACIT-F) are unidimensional and the remaining 5 (IBD-F, MFI-20, MAF, M-FIS, FQ) are multidimensional, with 2-5 dimensions. Only 1 of the 8 PROMs was developed for IBD population. A total score is used for most multidimensional PROMs, but only subscores are used for 2 PROMs (IBD-F and MFI-20).

Sample characteristics

A total of 2188 patients with IBD, both Crohn's disease and ulcerative colitis subtypes, were included from 8 studies, with sample sizes ranging from 61 to 567. In general, the study samples had slightly more females than males (average proportion of females: 55.2%); ranged in age from 34 to 57 years; and were comprised of individuals in both active and inactive IBD states. Full study details are included in the table, Supplementary Data Content 6.

Evaluation of the psychometric robustness of PROMs

1. *Validity*

1.1 Content validity

Content validity was not assessed for any of the unidimensional measures. Content validity was assessed in only 1 multidimensional measure, the English version of IBD-F.[27] Sixteen IBD patients were interviewed and indicated that items were comprehensive, comprehensible, more relevant, and more specific to their IBD-fatigue than other generic measures.[27] Content validity of the IBD-F(English) was evaluated as sufficient with high quality evidence. The evaluation details on content validity (including relevance, comprehensiveness, and comprehensibility) of the IBD-F (English) can be found in table, Supplementary Data Content 7.

1.2 Construct validity

Studies on construct validity (convergent and known-groups validity) were available for all unidimensional and multidimensional PROMs, with sample sizes varying from 61-465.[5,25,27–31] In brief, the robustness of construct validity was evaluated as sufficient for most PROMs and subscales (Unidimensional measures: FSS [Spanish], D-FIS [Spanish], FACIT-F; Multidimensional measures: IBD-F [English, Brazil, Greek, Danish], MFI-20, MAF, M-FIS [Spanish], FQ [Norwegian]-Physical fatigue) with moderate to high quality of evidence, except for the mental fatigue subscale of the FQ (Norwegian). The robustness of construct validity was evaluated as inconsistent for the FQ (Norwegian) mental fatigue subscale because it was moderately correlated with the mental health subscale of the quality of life measure; however, it was unable to discriminate between healthy controls and IBD patients. The details on the study quality, overall psychometric robustness, and its quality of evidence on construct validity of measures can be found in table, Supplementary Data Content 8.

1.3 Structural validity

Studies on structural validity were only available for the IBD-F (English, Greek), and the FQ (Norwegian), with sample sizes varying from 61-465.[27,29,31] Only the IBD-F (English) found a two-factorial structure and demonstrated high quality evidence of sufficient structural validity.[27] The robustness of structural validity was evaluated as indeterminate for the IBD-F (Greek) and the FQ (Norwegian) because structural validity was not assessed for the whole measure of the IBD-F (Greek) and there was no information on how well the data adequately fit the hypothesized model of the FQ (Norwegian). The details on the study quality, overall psychometric robustness, and its quality of evidence on structural validity of measures can be found in table, Supplementary Data Content 9.

1.4 Cross-cultural validity

Cross-cultural validity was not assessed for any of the seven translated versions of PROMs, namely the FSS (Spanish), the D-FIS (Spanish), the IBD-F (Brazilian-Portuguese, Greek, Danish), the M-FIS (Spanish), and the FQ (Norwegian).

2. *Reliability*

2.1 Internal consistency

Studies on internal consistency were available for the FACIT-F, the IBD-F (English, Brazil, and Greek versions), and the FQ (Norwegian), with sample sizes varying from 61 to 465.[25,27,28,31] The Cronbach's alpha of the PROMs and subscales were summarized as follows: FACIT-F (0.94-0.95), the IBD-F (English: 0.91-0.98; Brazil: 0.95-0.98; Greek: 0.9-0.97) and the FQ (Norwegian: 0.73-0.89) (See table, Supplementary Data Content 9, which demonstrates the Cronbach's alpha values for the measures and subscales). [25,27,28,31] For most PROMs and subscales with data on their Cronbach's alphas, the robustness of internal consistency was indeterminate and the study quality was 'doubtful' because of insufficient details on the structural validity. [19] Only the IBD-F (English) demonstrated high quality evidence of sufficient internal consistency. The details on the study quality, overall psychometric robustness, and its quality of evidence on internal consistency of measures can be found in table, Supplementary Data Content 9.

2.2 Test-retest reliability

Studies on test-retest reliability were available for the FACIT-F, the IBD-F (English, Brazil, Greek, Danish), the MFI-20, the MAF, and the FQ (Norwegian), with sample sizes varying from 22 to 123. [25,27–31] The test-retest time interval ranged from 48 hours to over 6 months across studies. [25,27–31] Most studies were in adequate or very good quality. [25,27,29–31] Yet one study assessed the test-retest reliability of IBD-F (Brazil) within 48-72 hours, [28] in which the time interval was considered too short and may introduce recall bias and therefore given a

'doubtful' rating to the study quality for IBD-F (Brazil). Most PROMs and subscales demonstrated sufficient test-retest reliability with the intraclass correlations (ICCs) >0.7, except the general fatigue subscale of the MFI-20 (ICC:0.65). The ICCs were summarized as follows: FACIT-F (0.81), IBD-F (English: 0.74-0.83; Brazil: 0.92-0.97; Greek: 0.88-0.9; Danish: 0.88-0.94), MFI-20 (0.65-0.84), MAF (0.74), and FQ (Norwegian: 0.88-0.98) (See table, Supplementary Data Content 10, which shows the ICCs of the measures and subscales).[25,27–31] The quality of evidence on test-retest reliability varied from very low to low, as most had only data from a single study and with sample sizes of less than 100. The details on the study quality, overall psychometric robustness, and its quality of evidence on test-retest reliability of measures can be found in table, Supplementary Data Content 10.

2.3 Measurement error

Measurement error was only assessed for the IBD-F (Brazil). [28] Lage et al.'s study (n=118) found that the standard error of measurement and minimal detectable change of the total IBD-F (Brazil) were 4.8 and 6.05 respectively. [28] The psychometric robustness was evaluated as indeterminate because we could not determine whether the measurement error was smaller than the minimal important change (not assessed).

3. Responsiveness

Of the 3 unidimensional measures, responsiveness was only assessed for the FACIT-F. [25] The change in the FACIT-F corresponded to the change of Physician's Global Assessment of disease activity in terms of direction and magnitude in 209 IBD patients (Much better = -11.8; slightly better = -2.6 ; same = 0.7; slightly worse = +2.4 ; much worse = +5.2). [25] Thus, the FACIT-F had high quality evidence on sufficient responsiveness. The minimal clinical important change of FACIT-F was 2.4-2.6. [25]

Of the 5 multidimensional measures, responsiveness was only assessed for the IBD-F (Brazil). The change in the IBD-F(Brazil) total score was strongly correlated with the change

in disease activity in Crohn's disease (r=0.81;n=81), but weakly correlated in ulcerative colitis (r=0.24; n=37) subgroup. [28] The robustness of responsiveness was evaluated as inconsistent because the strengths of the correlations were not consistent and not fully aligned with our expected correlation (>0.3) across both Crohn's disease and ulcerative colitis subgroups. The minimal detectable change of the total IBD-F (Brazil) score was 6.05. [28] Yet, the minimal clinically important change of the IBD-F (Brazil) was not assessed. [28]

Summary results on the psychometric robustness, interpretability of scoring, and feasibility of the PROMs

Table 2 and Table 3 show the summary results on the psychometric robustness, data on the interpretability of scoring, and the feasibility/ utility of the unidimensional and multidimensional measures, respectively. The unidimensional FACIT-F and the multidimensional IBD-F (English) have the most robust psychometric properties for research and clinical use.

Discussion

This is the first systematic review to identify standardised IBD-related fatigue PROMs, evaluate their psychometric properties, and recommend the most robust measure(s) for research and clinical use. Our evaluation process indicated that the FACIT-F and IBD-F (English) are the most robust PROMs for clinical and research use.

We identified PROMs that assess different facets of fatigue that align with our conceptual framework of IBD-related fatigue, looking sometimes at experience and/or impact of fatigue. Numerous unidimensional and multidimensional PROMs were developed due to varying definitions of fatigue. Most of our reviewed PROMs are multidimensional and focus on the impact of fatigue on function. As all different aspects of fatigue experience can

additively or synergistically impact function, PROMs assessing the overall impact of fatigue might provide a more representative and holistic assessment than PROMs assessing only fatigue experience. Also, assessing the impact of fatigue on function might be more objective than assessing fatigue experience alone because fatigue experience can vary daily or even hourly in IBD patients. [27] On the other hand, it is important to appreciate the unique contribution of the PROMs assessing fatigue experience. As fatigue takes time to impact function, PROMs assessing fatigue experience might be useful to evaluate fatigue at the early onset of IBD before fatigue affects function.

In this review content validity is the most important psychometric property for a PROM for IBD as suggested by the COSMIN guideline because it is crucial for a PROM to reflect all the key and relevant aspects of IBD-related fatigue. Yet, the content validity of the generic PROMs (i.e. PROMs developed in disease population other than IBD) have not been tested in the IBD population. It is unclear whether the items of the generic PROMs are still comprehensive in assessing IBD-related fatigue. However, the PROMs included in our review contain very similar questions which supports face validity. Remaining challenges include the lack of evidence on responsiveness for many IBD-related PROMs and the lack of cross-cultural validation of translated PROMs. This highlights the need for future research to address these gaps to enable researchers to accurately determine the efficacy of interventions and assess the intervention outcomes across global studies using cross-culturally validated PROMs.

The recommendation for the most 'robust' PROM for research and clinical use is the FACIT-F because it demonstrates construct validity, test-retest reliability, and responsiveness. It is the only PROM that has data on the minimal clinically important change, is also easy to administer, score, and interpret, making it a preferred measure for IBD clinical research and practice. [13] However, the FACIT-F is a unidimensional measure assessing only the impact of fatigue, and is therefore less likely to detect changes in fatigue that do not affect function (e.g.

fatigue experience). This may be important if outcome measurement needs to detect the impact of fatigue on function. Also, the FACIT-F can be easily used for practice because of its brevity.

The IBD-F (English) is an alternative choice. It is the only PROM developed to measure IBD-specific fatigue concerns and is multidimensional, assessing fatigue experience, the impact of fatigue, and perceived causes of fatigue. [27] It has been translated into three languages, also tested in IBD patients, and can be administered in online or paper-and-pencil format. It is free for access, easy to administer, score, and interpret. The IBD-F(English) demonstrated content, construct, structural validity, internal consistency, and test-retest reliability. There is no evidence on responsiveness or the minimally important statistical or clinical change. Although both FACIT-F and IBD-F (English) primarily assess the impact of fatigue, the IBD-F (English) includes specific questions on self-care, physical, cognitive, emotional function, and some unique aspects such as sexual and interpersonal relationships, self-confidence, and quality of life. More detail on specific areas of impact may be of importance if researchers expect change particularly in these areas. This potential benefit must be considered against the lack of data on responsiveness. The IBD-F (English) would be preferred over the FACIT-F if researchers are interested in measuring multidimensions of fatigue (i.e. experience and impact of fatigue), with the caveat that responsiveness of the measure has not been established. The IBD-F (English) may also be more clinically useful than the FACIT-F as the first section can serve as a screening tool to identify individuals with fatigue, with only those requiring so completing the more detailed assessment in Section Two.

The MFI-20 and the MAF are promising candidates because both are multidimensional but shorter (16-20 items) than the IBD-F (40 items) and demonstrate at least moderate quality evidence on validity and reliability. The MFI-20 assesses various aspects of fatigue experience (physical, mental, general fatigue, reduced activity and motivation), which the IBD-F and the MAF do not distinctively measure. Researchers can use the MFI-20 subscales to understand

which aspect of fatigue is most severe which will help develop a more effective intervention targeting the key aspect. The MAF shows promise because it measures the timing, distress, severity, and impact of fatigue on daily activities with a single score. The MAF is advantageous in their ease of administration and interpretation of scoring. Future research is needed to examine the responsiveness of the MFI-20 and the MAF.

Our results must be interpreted in light of limitations of our review. First, additional psychometric data may exist in studies not designed to examine psychometric properties of the PROMs (e.g. convergent validity could be reflected from the association between fatigue and other related constructs). We choose studies designed to test the psychometric properties as they would be most relevant and more robust. We did not search out data within studies not designed to test the psychometric properties of the PROMs because those studies may not be conducted in both Crohn's disease, ulcerative colitis, inactive, and active IBD subgroups. Therefore, the psychometric data from those studies may have limited values and may not be generalisable to a representative overall IBD population.

Second, COSMIN provides a robust yet conservative evaluation of the quality of evidence. The use of the 'lowest count score' in assessing the study quality may overestimate the risk of study bias. Therefore our estimates are likely an underestimate of the psychometric robustness. However, this would apply to all the PROMs assessed equally and therefore would not affect the overall conclusions. Lastly, fatigue PROMs that have not been validated in IBD patients may be useful which warrants future study to explore.

Fatigue management has been identified as a priority for IBD patients, therefore assessment of fatigue and ability to detect change over time is key. This review systematically evaluated the psychometric robustness of standardised fatigue PROMs used in IBD research. We recommend the unidimensional FACIT-F and the multidimensional IBD-F (English) as the

most robust measures for research and clinical use. Additional research on the responsiveness of these measures is needed.

Acknowledgement

We acknowledge the contribution of Dovina Constance (clinical specialist nurse), Chelsea Sherrington (clinical psychologist), and Brenda O'Connor (research nurse) at the Inflammatory Bowel Disease (IBD) centre at Mount Sinai Hospital (Toronto). They provided valuable clinical insights regarding the use of fatigue measures in the IBD population.

Reference

1. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2017;390(10114):2769-2778.
2. Kaplan G, Benchimol E, Bernstein C, et al. Impact of Inflammatory Bowel Disease in Canada. 2018:1-232. https://www.crohnsandcolitis.ca/Crohns_and_Colitis/documents/reports/2018-Impact-Report-LR. Accessed on January 10, 2021.
3. Singh S, Blanchard A, Walker JR, et al. Common Symptoms and Stressors Among Individuals With Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol*. 2011;9(9):769-775.
4. Dhalaigh DN, Greaney AM, Farrell D. Inflammatory bowel disease fatigue: an analysis of definitions, risk factors and impact and their implications for assessment and management. *Gastrointest Nurs*. 2020;18(5):42-50.
5. Castillo-Cejas MD, Robles V, Borruel N, et al. Questionnaries for measuring fatigue and its impact on health perception in inflammatory bowel disease. *Rev Esp Enferm Dig*. 2013;105(3):144-153.
6. Dibley L, Bager P, Czuber-Dochan W, et al. Identification of Research Priorities for Inflammatory Bowel Disease Nursing in Europe: a Nurses-European Crohn's and Colitis Organisation Delphi Survey. *J Crohns Colitis*. 2017;11(3):353-359.
7. Hart AL, Lomer M, Verjee A, et al. What Are the Top 10 Research Questions in the Treatment of Inflammatory Bowel Disease? A Priority Setting Partnership with the James Lind Alliance. *J Crohns Colitis*. 2017;11(2):204-211.
8. Farrell D, Artom M, Czuber-Dochan W, et al. Interventions for fatigue in inflammatory bowel disease. *Cochrane database Syst Rev*. 2020;4:CD012005.

9. Van Langenberg DR, Della Gatta P, Warmington SA, et al. Objectively measured muscle fatigue in Crohn's disease: Correlation with self-reported fatigue and associated factors for clinical application. *J Crohn's Colitis*. 2014;8(2):137-146.

10. Van Langenberg DR, Gibson PR. Systematic review: Fatigue in inflammatory bowel disease. *Aliment Pharmacol Ther*. 2010;32(2):131-143.

11. Cella D, Peterman A, Passik S, et al. Progress toward guidelines for the management of fatigue. *Oncology*. 1998;12(11A):369-377.

12. Whitehead L. The Measurement of Fatigue in Chronic Illness: A Systematic Review of Unidimensional and Multidimensional Fatigue Measures. *J Pain Symptom Manage*. 2009;37(1):107-128.

13. Borren NZ, Woude CJ van der, Ananthakrishnan AN. Fatigue in IBD: epidemiology, pathophysiology and management. *Nat Rev Gastroenterol Hepatol*. 2019;16:247-259.

14. Czuber-Dochan W, Norton C, Bredin F, et al. Healthcare professionals' perceptions of fatigue experienced by people with IBD. *J Crohn's Colitis*. 2014;8(8):835-844.

15. Czuber-Dochan W, Ream E, Norton C. Review article: Description and management of fatigue in inflammatory bowel disease. *Aliment Pharmacol Ther*. 2013;37(5):505-516.

16. Gadotti IC, Vieira ER, Magee DJ. Importance and clarification of measurement properties in rehabilitation. *Brazilian J Phys Ther*. 2006;10:137-146.

17. de Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011.

18. Greenhalgh J, Long AF, Brettle AJ, et al. Reviewing and selecting outcome measures for use in routine practice. *J Eval Clin Pract*. 1998;4(4):339-350.

19. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.

20. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;27(5):1171-1179.

21. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Qual Life Res*. 2010;19(4):539-549.

22. Mokkink LB, Terwee CB, Gibbons E, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) Checklist. *BMC Med Res Methodol*. 2010;10(82).

23. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.

24. Schünemann H, Brożek J, Guyatt G OA, et al. GRADE handbook for grading quality of evidence and strength of recommendations. www.guideline development.org/handbook. Published 2013. Accessed January 15, 2021.

25. Tinsley A, Macklin EA, Korzenik JR, et al. Validation of the functional assessment of chronic illness therapy-fatigue (FACIT-F) in patients with inflammatory bowel disease. *Aliment Pharmacol Ther*. 2011;34(11-12):1328-1336.

26. Norton C, Czuber-Dochan W, Bassett P, et al. Assessing fatigue in inflammatory bowel disease: comparison of three fatigue scales. *Aliment Pharmacol Ther*. 2015;42(2):203-211.

27. Czuber-Dochan W, Norton C, Bassett P, et al. Development and psychometric testing of inflammatory bowel disease fatigue (IBD-F) patient self-assessment scale. *J Crohns Colitis*. 2014;8(11):1398-1406.

28. Lage AC, Oliveira CC, Batalha APDB, et al. The inflammatory bowel disease-fatigue patient self-assessment scale: Translation, cross-cultural adaptation and psychometric properties of the brazilian version (IBD-F Brazil). *Arq Gastroenterol*. 2020;57(1):50-63.

29. Varbobitis I, Kokkotis G, Gizis M, et al. The IBD-F Patient Self-Assessment Scale Accurately Depicts the Level of Fatigue and Predicts a Negative Effect on the Quality of Life of Patients With IBD in Clinical Remission. *Inflamm Bowel Dis*. 2020;27(6):826-835.

30. Vestergaard C, Dahlerup JF, Bager P. Validation of the Danish version of Inflammatory Bowel Disease Self-assessment Scale. *Dan Med J*. 2017;64(8).

31. Jelsness-Jørgensen LP, Bernklev T, Henriksen M, et al. Chronic fatigue is more prevalent in patients with inflammatory bowel disease than in healthy controls. *Inflamm Bowel Dis*. 2011;17(7):1564-1572.

32. Jelsness-Jorgensen LP. The Fatigue Questionnaire has a good test-retest profile in IBD. *Aliment Pharmacol Ther*. 2012;35(5):621-622.

33. Multidimensional Assessment of Fatigue (MAF). American College of Rheumatology Website. https://www.rheumatology.org/I-Am-A/Rheumatologist/Research/Clinician-Researchers/Multidimensional-Assessment-of-Fatigue-MAF. Accessed May 19, 2021.

34. Modified Fatigue Impact Scale. Shirley Ryan AbilityLab. https://www.sralab.org/rehabilitation-measures/modified-fatigue-impact-scale. Accessed May 19, 2021.

35. Jackson C. The Chalder Fatigue Scale (CFQ 11). *Occup Med (Chic Ill)*. 2015;65(1):86.

36. Shahid A, Wilkinson K, Marcu S, et al. Chalder Fatigue Scale. *STOP, THAT and One Hundred Other Sleep Scales*. New York, NY: Springer New York; 2012:97-98.

37. Montan I, Löwe B, Cella D, et al. General Population Norms for the Functional

Assessment of Chronic Illness Therapy (FACIT)-Fatigue Scale. *Value Heal J Int Soc Pharmacoeconomics Outcomes Res*. 2018;21(11):1313-1321.

38. Bager P, Vestergaard C, Juul T, et al. Population-based normative data for the inflammatory bowel disease fatigue scale – IBD-F. *Scand J Gastroenterol*. 2018;53(10-11):1274-1279.

39. Hinz A, Barboza CF, Barradas S, et al. Fatigue in the general population of Colombia - normative values for the multidimensional fatigue inventory MFI-20. *Onkologie*. 2013;36(7-8):403-407.

40. Loge JH, Ekeberg O, Kaasa S. Fatigue in the general Norwegian population: normative data and associations. *J Psychosom Res*. 1998;45(1):53-65.

Tables

Table 1 Characteristics of inflammatory bowel disease-related fatigue patient-reported outcome measures (PROMs)

Name and abbreviation of PROM (Year of development)	**Construct of (sub)scale(s)**	**Target population**	**Recall period**	**Item (n)**	**Scoring and its interpretability** Scale, response options, score range (method of scoring), interpretation of score	**Administration** Administration format and time, training required, Access and available language of PROM
Unidimensional						
Fatigue Severity Scale, FSS (1989)	Fatigue severity	Systematic lupus erythematosus, Multiple sclerosis	7 days	9	7-point scale; 'Strongly disagree' to 'Strongly agree' Total score: 9-63 (Summed responses) Average score: 1-7 (Summed responses/number of responses) Higher score means greater fatigue severity.	Administration format & time: Self-reported pencil-and-paper; <5 minutes Training required: No Request for access: https://eprovide.mapi-trust.org/instruments/fatigue-severity-scale Available languages: 38 Translations, including English Validated language in IBD patients: Spanish
Daily Fatigue Impact Scale, D-FIS (2002)	Impact of fatigue	Participants with flu like symptoms	Within today	8	5-point scale; 'No problem' to 'Extreme problem' Total score: 0-32 (Summed ordinal responses) Higher score means greater fatigue impact during the day.	Administration format & time: Self-reported pencil-and-paper; 5-10 minutes Training required: Not reported Request for access: https://eprovide.mapi-trust.org/instruments/daily-fatigue-impact-scale Available languages: English, Spanish Validated language in IBD patients: Spanish
Functional Assessment of Chronic Illness Therapy-Fatigue, FACIT-F (1997)	Impact of Fatigue	Cancer patients with low hemoglobin level	7 days	13	5-point scale; 'Not at all' to 'Very much' Total score: 0-52 (Summed ordinal responses) Lower score means greater fatigue.	Administration format & time: Self-reported pencil-and-paper, interview format; <10 minutes Training required: 6th grade reading level is required Request for access: www.facit.org/measures/FACIT-F Available languages: 62 Translations, including English Validated language in IBD patients: English

Multidimensional						
IBD-Fatigue scale, IBD-F (2014)	3 Subscales: 1. Severity and frequency of fatigue (7 items) 2. Impact of fatigue (30 items) 3. Factors contributing to fatigue (3 items)	IBD	14 days	40	Scoring is only available for Section 1 and 2, with 5-point scale. 'No fatigue/None of the time' to 'Severe fatigue/ All of them time' Section 2 will only be completed if clients select answers between responses 1 to 4 in any 1 item in Section 1. Score for Section 1,2 : 0-20, 0-120 (Summed ordinal responses within each section) Calculated score = Summed ordinal responses/ (120 – no. of 'not applicable'options x4) x 120.[27] Higher score means greater fatigue severity (Section 1) and impact (Section 2). Section 3: Open-ended format	Administration format & time: Self-reported pencil-and-paper (4 pages), online version; 10-12 minutes Training required: Not reported Free for access: Online version of IBD-F is available from www.fatigueinibd.co.uk; with automatic scoring program. Available and validated languages in IBD patients: English, Greek, Danish, Brazilian-Portuguese
Multidimensional Fatigue Inventory, MFI-20 (1995)	5 Subscales: 1. General fatigue 2. Physical fatigue 3. Reduced activity 4. Mental fatigue 5. Reduced motivation (4 items for each subscale)	Cancer patients receiving radiotherapy, patients with chronic fatigue syndrome.	During the previous days	20	5-point scale; 'Yes, that is true' to 'No, that is not true' Each subscore: 4-20 (10 items require reverse scoring [Item 2,5,9,10,13,14,16,17,18,19]; Summed responses within each subscale) General fatigue: Item 1,5,12,16; Physical fatigue: Item 2,8,14,20; Reduced Activity: Item 7,11,13,19; Mental fatigue: Item 4,9,15,18; Reduced motivation: Item 1,5,12,16 Higher score means greater fatigue.	Administration format & time: Self-reported pencil-and-paper; Not reported Training required: Not reported Request for access: https://eprovide.mapi-trust.org/instruments/multidimensional-fatigue-inventory Available languages: 75 Translations Validated languages in IBD patients: English

Multidimensional Assessment of Fatigue, MAF (1995)	4 Dimensions†: Severity, distress, degree of interference in activities of daily living, and timing of fatigue † Not intended to be used as subscales	Patients with rheumatoid arthritis	7 days	16	10-point scale for item #1-14 on fatigue impact 'Not at all' to 'A great deal' 4-point Scale for item #15-16 on timing and frequency of fatigue Item 15: 'Hardly any days' to 'Every day'; Item 16: 'Decreased' to 'Increased' Total score: 1-50 (Sum up items #1,2,3, average of item # 4-14, and item #15 x 2.5; item #16 is not included in calculation of the final score) Higher score indicates a higher level of fatigue, fatigue distress, and its impact on the activities of daily living.	Administration format & time: Self-reported pencil-and-paper; <10 minutes[33] Training required: No Request for access: https://eprovide.mapi-trust.org/instruments/multidimensional-assessment-of-fatigue Available languages: 52 Translations Validated language in IBD patients: English
Modified Fatigue Impact Scale, M-FIS (2005)	3 Subscales: Impact of fatigue on 1. Physical function (9 items) 2. Cognitive function (10 items) 3. Psychosocial function (2 items)	Multiple sclerosis	Previous month	21	5-point scale; 'No problem' to 'Extreme problem' Physical subscore: 0-36 (Summed ordinal responses from item 4,6,7,10,13,14,17,20,21) Cognitive subscore: 0-40 (Summed ordinal responses from item 1,2,3,5,11,12,15,16,18,19) Psychosocial subscore: 0-8 (Summed ordinal responses from item 8,9) Total score: 0-84 (Summed ordinal responses from all items) Higher score indicates greater fatigue.	Administration format & time: Self-reported pencil-and-paper; <10 minutes[34] Training required: Reading manual[34] Request for access: https://eprovide.mapi-trust.org/instruments/modified-fatigue-impact-scale Available languages: 46 Translations, including English Validated language in IBD patients: Spanish

Fatigue Questionnaire, FQ (1993)	2 Subscales: 1. Physical Fatigue (7 items) 2. Mental fatigue (4 items)	Chronic fatigue syndrome	Not provided	11	2 ways of scoring: <u>1) 4-point scale</u> 'Better than usual' to 'Much worse than usual Physical fatigue: 0-21(Summed ordinal responses from items 1-7) Mental fatigue: 0-12. (Summed ordinal responses from items 8-11) Total score: 0-33 (Summed all ordinal responses) Higher score means greater fatigue. <u>2) Bimodal scoring on the global score</u>[35] 0 – 'Better than usual', ' No more than usual'; 1 – 'Worse than usual', 'Much worse than usual' Global binary score: 0-11 Global binary score < 3 indicates no fatigue. Global binary score >4 indicates severe fatigue.	Administration format & time: Self-reported pencil-and-paper, interview; 3-5 minutes[36] Training required: Not reported Request for access: https://eprovide.mapi-trust.org/instruments/chalder-fatigue-scale Available languages: 9 Translation, including English Validated language in IBD patients: Norwegian

Table 2. Unidimensional patient-reported outcome measures: Summary results on psychometric robustness, data on interpretability of scoring, and feasibility/utility of measures

PROM (row)/ **Psychometric property** (column)	**FSS (Spanish)**	**D-FIS (Spanish)**	**FACIT-F**
Validity			
Content	?	?	?
Construct	✓✓✓	✓✓✓✓	✓✓✓✓
Structural	?	?	?
Cross cultural	?	?	Not applicable
Reliability			
Internal consistency	?	?	?
Test-retest reliability	?	?	✓✓✓
Measurement Error	?	?	?
Responsiveness	?	?	✓✓✓✓
Interpretability of scoring Available data on the distribution of scores; ceiling/floor effect; score for UC and CD, inactive and active IBD subgroups, norm; minimal important change (MIC)	Scores for CD, UC, active and inactive IBD subgroups [5]	Scores for healthy individuals, active and inactive IBD subgroups[5]	Scores for active and inactive IBD states[25]; Norm data in German[37] ; MIC: 2.4-2.6 in IBD patients [25]
Feasibility/utility Pros and cons	Pros: Easy to administer, score, and interpret single scoring. Cons: Limited information from the unidimensional measure.	Pros: Easy to administer, score, and interpret single scoring. Cons: Limited information from the unidimensional measure. The recall period is within 24 hours, which might not be too applicable to assess fatigue as a chronic condition in IBD patients. It could be too burdensome for patients to complete daily.	Pros: Easy to administer, score, and interpret single scoring. The information on MIC allows clinicians and researchers to understand whether the change is meaningful to patients. Cons: Limited information from the unidimensional measure.

Key: High (✓✓✓✓)/moderate (✓✓✓)/low (✓✓)/very low(✓) quality evidence on sufficient psychometric robustness; (?): Psychometric robustness was indeterminate.
Abbreviation: FSS – Fatigue Severity Scale, D-FIS – Daily Fatigue Impact Scale, FACIT-F - Functional Assessment of Chronic Illness Therapy – Fatigue, CD – Crohn's disease, UC – Ulcerative Colitis, IBD – Inflammatory bowel disease

Table 3. Multidimensional patient-reported outcome measures: Summary results on psychometric robustness, data on interpretability of scoring, and feasibility/utility of measures

PROM (row)/ **Psychometric property** (column)	**IBD-F** English	**IBD-F** Brazilian Portuguese	**IBD-F** Greek	**IBD-F** Danish	**MFI-20**	**MAF**	**M-FIS (Spanish)**	**FQ (Norwegian)**
Validity								
Content	✓✓✓	?	?	?	?	?	?	?
Construct	✓✓✓✓	✓✓✓✓	✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓✓	✓✓✓	✓✓✓✓
Structural	✓✓✓✓	?	?	?	?	?	?	?
Cross cultural	Not applicable	?	?	?	Not applicable	Not applicable	?	?
Reliability								
Internal consistency	✓✓✓✓	?	?	?	?	?	?	?
Test-retest reliability	✓✓	✓✓	✓✓	✓✓	✓✓	✓✓	?	✓
Measurement Error	?	?	?	?	?	?	?	?
Responsiveness	?	± (CD: +, UC: -)	?	?	?	?	?	?
Interpretability of scoring Available data on the distribution of scores; ceiling/floor effect; score for UC and CD, inactive and active IBD subgroups, norm; minimal important change (MIC)	Distribution of score in UK IBD patients[27]; Scores for CD, UC, active, and inactive IBD subgroups[27]	Floor effect[28]; Scores for CD, UC, active, and inactive IBD subgroups[28]	Distribution of score in Greek IBD patients[29]	No ceiling/ floor effect[30]; Scores for active and inactive IBD groups[30]; Normative data in Danish population[38]	Normative data in Colombia population[39]	None	None	Scores for UC, CD subgroups, and healthy controls[31]; Normative data in Norwegian population[40]
Feasibility/utility Pros and cons	Pros: Free for access; Most number of the translated versions being tested in IBD patients; Multiple administration modalities; The perceived causes and management strategies for IBD-fatigue are asked in IBD-F, which can foster understanding on fatigue management. Section 1 can serve as a screening tool for further assessment by Section 2. Cons: 40 items could be burdensome for patients with severe fatigue to complete.				Pros: Easy to administer. It measures each aspect of fatigue experience at a granular level. Cons: Reverse scoring for some items; Subscores are not intended to sum up to a total score for interpretation. Subscores are required to analyse and interpret separately, which might be more complex than single scoring.	Pros: Easy to administer and interpret single scoring. Cons: A special formula is required to score.	Pros: Easy to administer, score, and interpret scoring. Cons: None.	Pros: Easy to administer, score, and interpret scoring. Cons: There is no clear recall period, which may introduce variability in responses across timepoints/ individuals.

† The psychometric robustness of the multidimensional PROM is determined based on the majority of the results (>75%) from the subscales because these PROMs are often used as a whole.

Key: High (✓✓✓✓)/moderate (✓✓✓)/low (✓✓)/very low(✓) quality evidence on sufficient psychometric robustness; The psychometric robustness was indeterminate (?), inconsistent(±), sufficient(+), or insufficient (-). Abbreviation: IBD-F – Inflammatory Bowel Disease Fatigue Scale, MFI-20 – Multidimensional Fatigue Inventory-20, MAF – Multidimensional Assessment of Fatigue, M-FIS – Modified Fatigue Impact Scale, FQ – Fatigue Questionnaire, CD – Crohn's Disease, UC – Ulcerative Colitis

Figures

Figure 1. A proposed conceptual framework of the inflammatory bowel disease-related fatigue

The multifactorial causes of inflammatory bowel disease (IBD)-related fatigue can lead to physical, cognitive, and emotional fatigue. [13] The experience and impact of fatigue are interdependent but slightly distinct. Different aspects of IBD-fatigue experience can additively or synergistically impact on physical, cognitive, and psychosocial functioning. [10] Ultimately, fatigue can impact on the health-related quality of life (HRQoL).

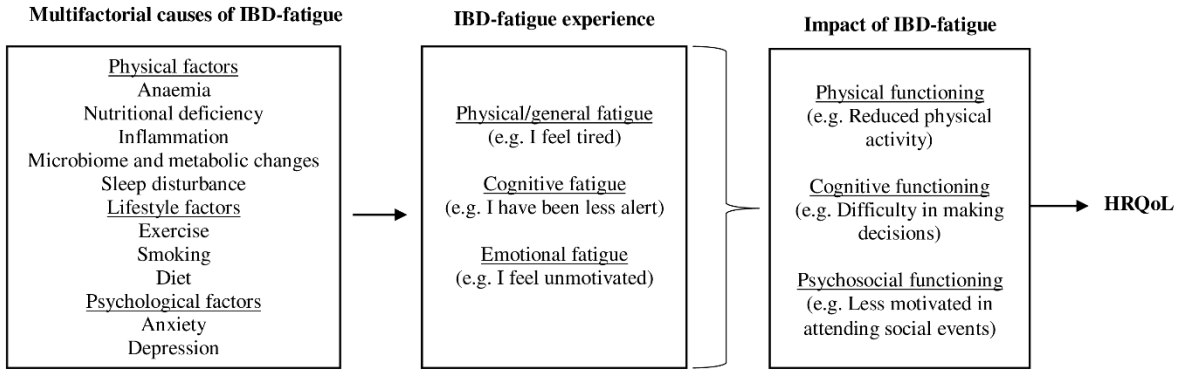


Figure 2. PRISMA diagram for identifying fatigue patient-reported outcome measures in adults with inflammatory bowel disease (IBD).

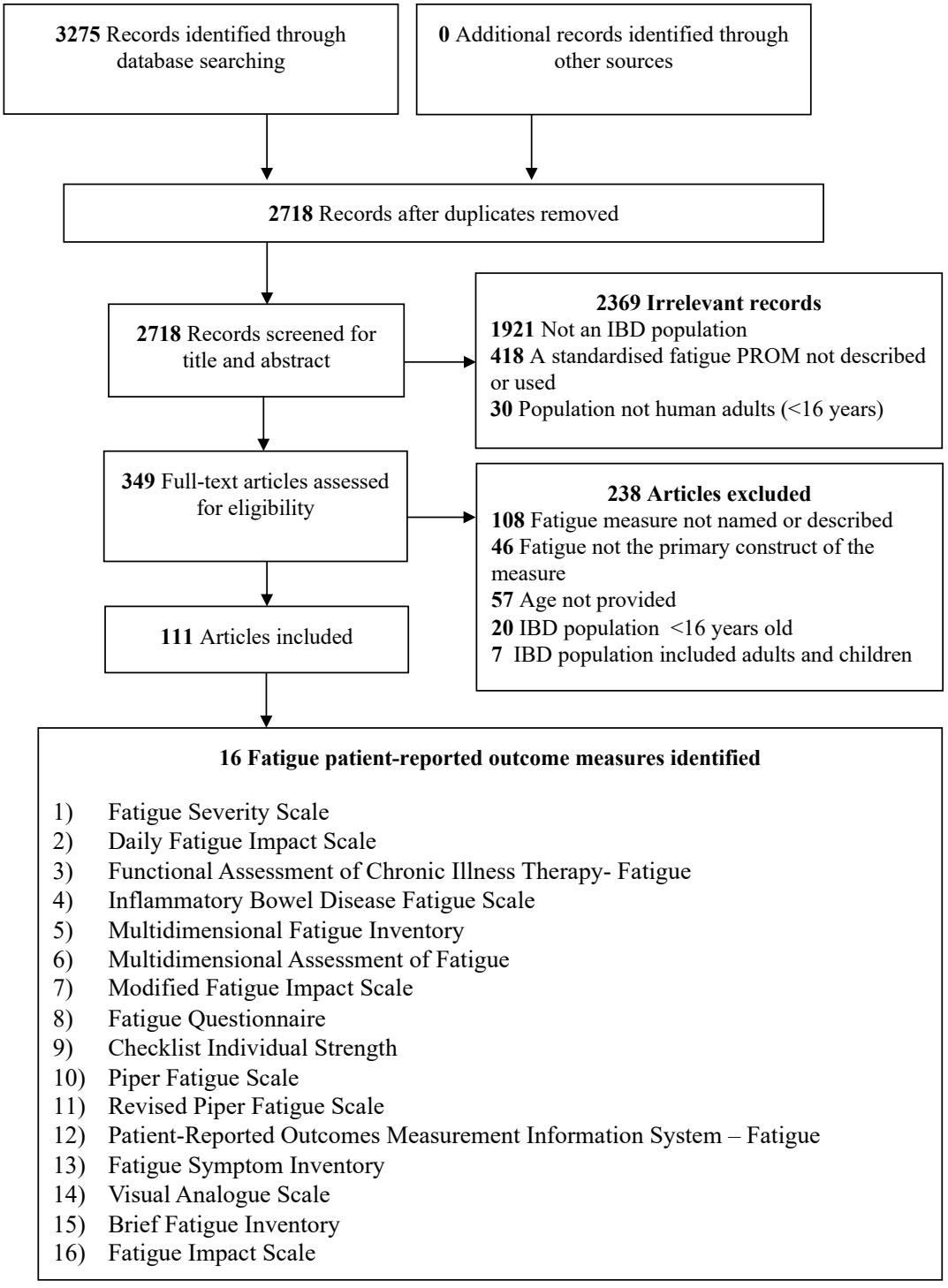
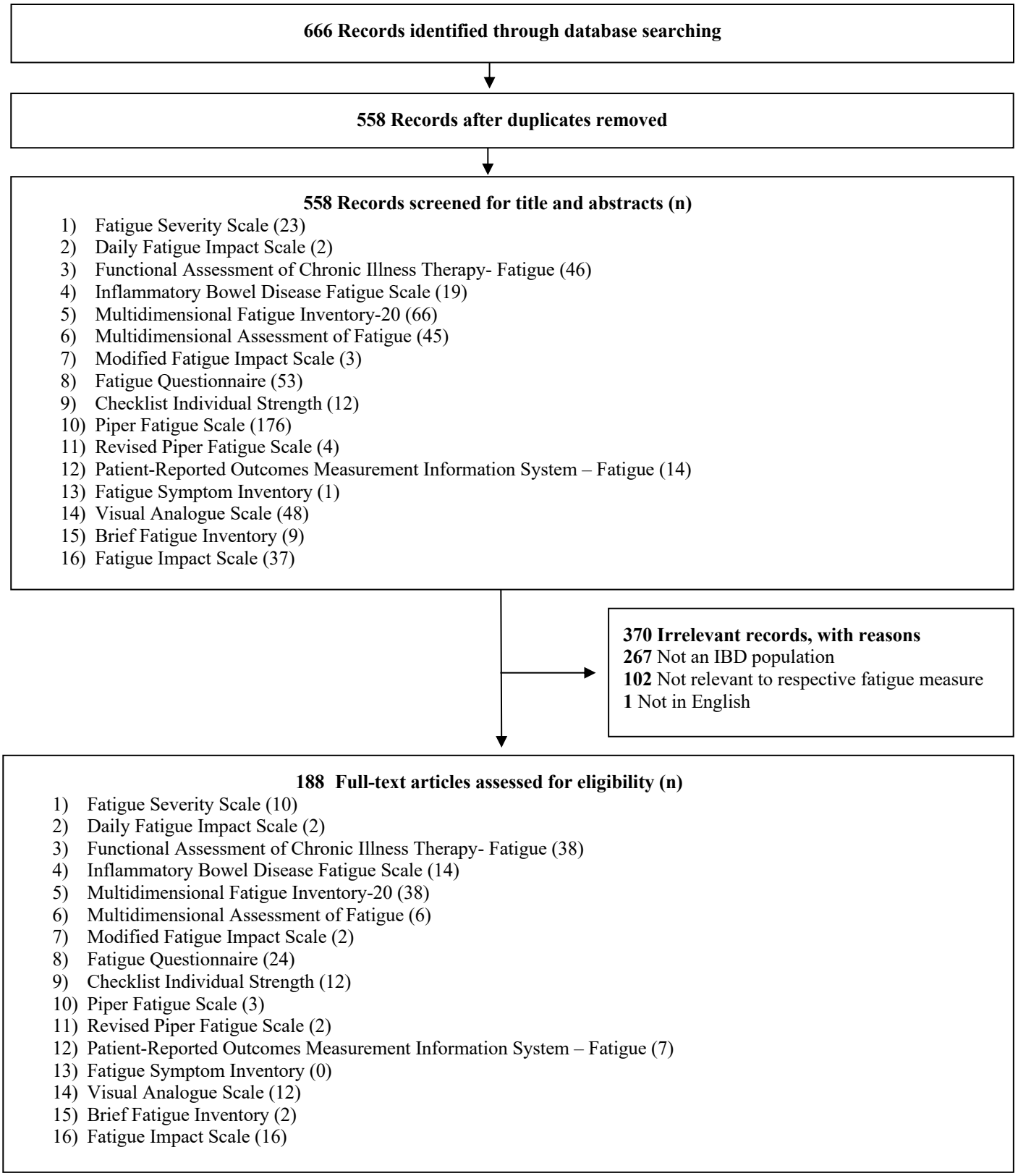
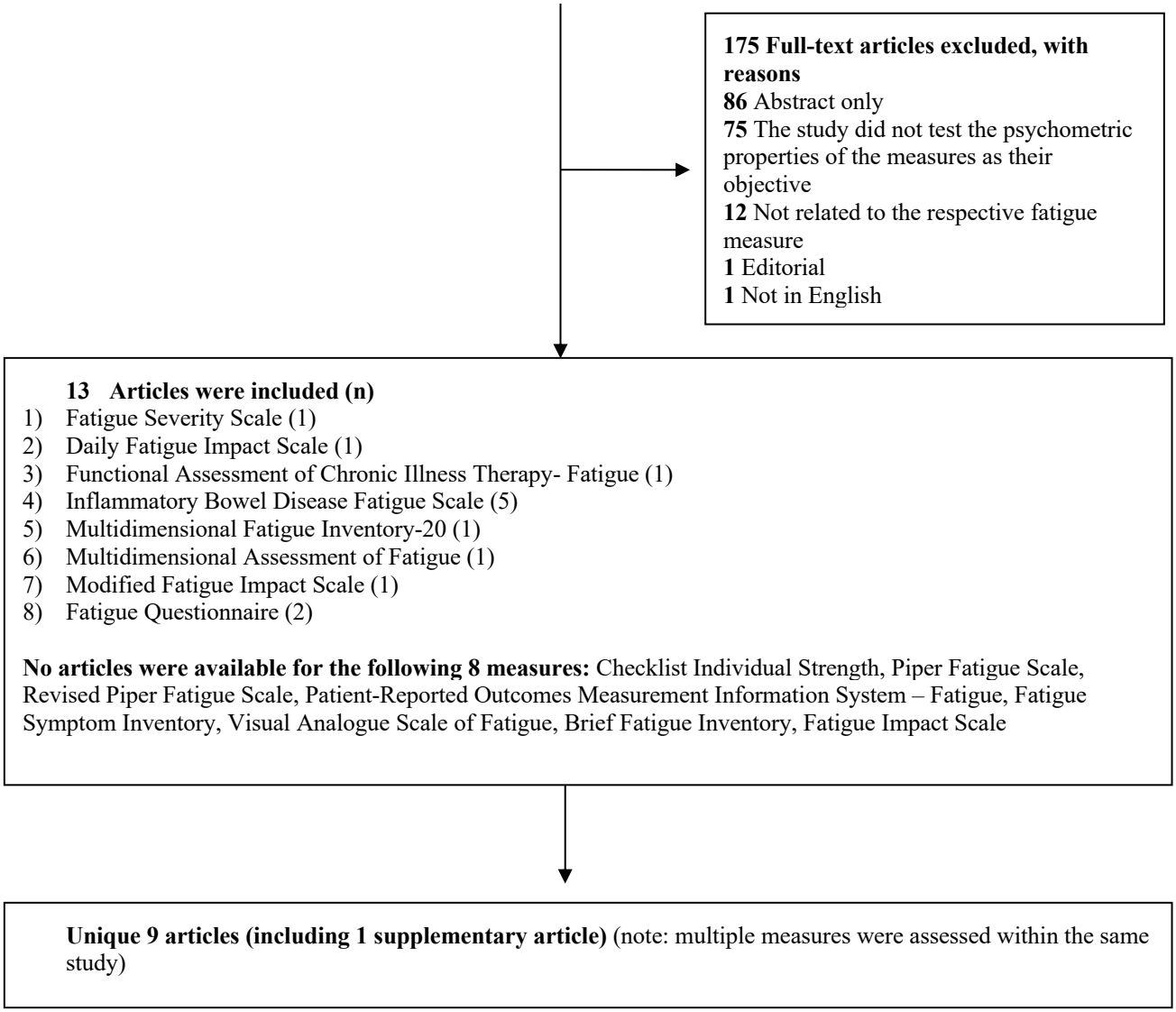


Figure 3. PRISMA diagram for identifying articles on the psychometric properties of the identified patient-reported outcome measures.




175 Full-text articles excluded, with reasons
86 Abstract only
75 The study did not test the psychometric properties of the measures as their objective
12 Not related to the respective fatigue measure
1 Editorial
1 Not in English
13 Articles were included (n)
1) Fatigue Severity Scale (1)
2) Daily Fatigue Impact Scale (1)
3) Functional Assessment of Chronic Illness Therapy- Fatigue (1)
4) Inflammatory Bowel Disease Fatigue Scale (5)
5) Multidimensional Fatigue Inventory-20 (1)
6) Multidimensional Assessment of Fatigue (1)
7) Modified Fatigue Impact Scale (1)
8) Fatigue Questionnaire (2)
No articles were available for the following 8 measures: Checklist Individual Strength, Piper Fatigue Scale, Revised Piper Fatigue Scale, Patient-Reported Outcomes Measurement Information System – Fatigue, Fatigue Symptom Inventory, Visual Analogue Scale of Fatigue, Brief Fatigue Inventory, Fatigue Impact Scale
Unique 9 articles (including 1 supplementary article) (note: multiple measures were assessed within the same study)

Supplementary Data Content 1. Search Strategy for the first search in identifying fatigue patient-reported outcome measures in adults with inflammatory bowel disease

The combination of the following 3 concepts were run in the search: inflammatory bowel disease (IBD), fatigue, and measurement.

Here are the subject headings and free text of the 3 concepts for MEDLINE (OVID).

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
exp inflammatory bowel disease/ exp Colitis/ exp Colitis, Ulcerative/ exp Crohn Disease/ Ileitis/ exp Proctocolitis/ Exp enterocolitis / Exp / Exp Enteritis/ inflammatory bowel disease.mp. IBD.mp. Colitis*.mp. Crohn*.mp. jejunoileitis.mp. ileocolitis.mp. ileitis.mp. Proctocolitis.mp. Proctosigmoiditis.mp. Enterocolitis.mp. Pancolitis.mp. Proctitis.mp. Enteritis.mp.	exp Fatigue/ Lethargy exp Asthenia/ exp Apathy/ exp Muscle Fatigue/ exp Mental Fatigue/ fatigue*.mp. letharg*.mp. astheni*.mp. apath*.mp. fatigability.mp. wear*.mp. tire*.mp. exhaust*.mp. lacklustre.mp. languor.mp. listless*.mp. lassitude.mp. ((low or lack or loss or lost) adj3 (energy or vigo*r)).mp.	exp Self-Assessment/ exp Patient Outcome Assessment/ exp Patient Reported Outcome Measures/ exp Self Report/ exp "Surveys and Questionnaires"/ or Checklist/ exp Visual Analog Scale/ self assess*.mp. patient assess*.mp. patient report*.mp. self report*.mp. survey*.mp. questionnaire*.mp. or assess*.mp. clinical tool*.mp. Measur*.mp. checklist*.mp. Visual Analog Scale*.mp. instrument*.mp. tool*.mp. scale*.mp. inventor*.mp. list*.mp. test*.mp. VAS*.mp. PROM*.mp. subscale* index.mp. form*.mp.

The subject headings and free text of the 3 concepts for EMBASE (OVID) were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
exp inflammatory bowel disease/ exp colitis/ exp ulcerative colitis/ exp Crohn disease/ exp ileitis/ exp proctocolitis/ exp enterocolitis/ exp proctitis/ exp enteritis/ exp colon Crohn disease/ inflammatory bowel disease*.mp. IBD.mp. Colitis*.mp Crohn*.mp. jejunoileitis.mp. ileocolitis.mp. ileitis.mp. proctocolitis.mp. Proctosigmoiditis.mp. enterocolitis.mp. Pancolitis.mp. proctitis.mp. Enteritis.mp.	fatigue/ Lethargy/ exp Asthenia/ exp Apathy/ exp exhaustion/ exp listlessness/ exp lassitude/ exp muscle fatigue/ fatigue*.mp. letharg*.mp. astheni*.mp. apath*.mp. fatigability.mp. wear*.mp. tire*.mp. exhaust*.mp. lacklustre.mp. languor.mp. listless*.mp. lassitude.mp. ((low or lack or loss or lost) adj3 (energy or vigo*r)).mp.	exp self evaluation/ exp patient-reported outcome/ exp self report/ exp short survey/ exp questionnaire/ exp outcome assessment/ exp "Outcome Assessment (Health Care)"/ exp checklist/ exp visual analog scale/ exp patient assessment/ exp clinical assessment tool/ exp Measurement/ self assess*.mp. patient assess*.mp. patient report*.mp. self report*.mp. survey*.mp. questionnaire*.mp. assess*.mp. clinical tool*.mp. Measur*.mp. checklist*.mp. Visual Analog Scale*.mp. instrument*.mp. tool*.mp. scale*.mp. inventor*.mp. list*.mp. test*.mp. VAS*.mp. PROM*.mp. subscale* index.mp. form*.mp.

The subject headings and free text of the 3 concepts for PsychINFO database were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
exp Colitis/ exp Ulcerative Colitis/ inflammatory bowel disease*.mp. IBD.mp. Colitis*.mp Crohn*.mp. jejunoileitis.mp. ileocolitis.mp. ileitis.mp. proctocolitis.mp. Proctosigmoiditis.mp. enterocolitis.mp. Pancolitis.mp. proctitis.mp. Enteritis.mp.	exp fatigue/ exp asthenia/ exp Apathy/ fatigue*.mp. letharg*.mp. astheni*.mp. apath*.mp. fatigability.mp. wear*.mp. tire*.mp. exhaust*.mp. lacklustre.mp. languor.mp. listless*.mp. lassitude.mp. ((low or lack or loss or lost) adj3 (energy or vigo*r)).mp.	exp self evaluation/ exp Patient Reported Outcome Measures/ exp Self Report/ exp survey/ exp questionnaires/ exp "Checklist (Testing)"/ exp Measurement/ patient report*.mp. self report*.mp. survey*.mp. questionnaire*.mp. assess*.mp. clinical tool*.mp. Measur*.mp. checklist*.mp. Visual Analog Scale*.mp. instrument*.mp. tool*.mp. scale*.mp. inventor*.mp. list*.mp. test*.mp. VAS*.mp. PROM*.mp. subscale*.mp. index.mp. form*.mp.

The subject headings and free text of the 3 concepts for the Allied and Complementary Medicine Database (AMED) were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
exp inflammatory bowel disease/ exp Colitis ulcerative/ exp Crohn disease/ inflammatory bowel disease*.mp. IBD.mp. Colitis*.mp Crohn*.mp. jejunoileitis.mp. ileocolitis.mp. ileitis.mp. proctocolitis.mp. Proctosigmoiditis.mp. enterocolitis.mp. Pancolitis.mp. proctitis.mp. Enteritis.mp.	exp Fatigue/ exp Muscle Fatigue/ exp Fatigue mental/ fatigue*.mp. letharg*.mp. astheni*.mp. apath*.mp. fatigability.mp. wear*.mp. tire*.mp. exhaust*.mp. lacklustre.mp. languor.mp. listless*.mp. lassitude.mp. ((low or lack or loss or lost) adj3 (energy or vigo*r)).mp.	exp Self assessment/ exp Questionnaires/ exp Patient assessment/ exp clinical assessment scales/ exp Measurement/ patient report*.mp. self report*.mp. survey*.mp. questionnaire*.mp. assess*.mp. clinical tool*.mp. Measur*.mp. checklist*.mp. Visual Analog Scale*.mp. instrument*.mp. tool*.mp. scale*.mp. inventor*.mp. list*.mp. test*.mp. VAS*.mp. PROM*.mp. subscale*.mp. index.mp. form*.mp.

The subject headings and free text of the 3 concepts for the CINAHL database were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
MH 'Inflammatory Bowel Diseases+" MH "Colitis+" MH "Colitis, Ulcerative" MH "Crohn Disease" MH "Ileitis+" MH "Enterocolitis+" MH "Enteritis+" "inflammatory bowel disease*" "IBD" "colitis*" "crohn*" "jejunoileitis" "ileocolitis" "ileitis" "proctocolitis" "Proctosigmoiditis" "enterocolitis" "Pancolitis" "proctitis" "Enteritis"	MH "Fatigue+" MH "Asthenia" MH "Apathy" MH "Muscle Fatigue” MH "Muscle Weakness+" MH "Mental Fatigue" "fatigue*" "letharg*" "astheni*" "apath*" "fatigability" "wear*" "tire*" "exhaust*" "lacklustre" "languor" "listless*" "lassitude"	MH "Self Assessment" MH "Patient-Reported Outcomes" MH "Surveys" MH "Questionnaires+" MH "Outcome Assessment" MH "Checklists" MH "Visual Analog Scaling" MH "Patient Assessment" MH "Clinical Assessment Tools" MH "Scales" MH "Research Instruments" "patient report*" "self report*" "Survey*" "questionnaire*" "assess*" "clinical tool*" "measur*" "checklist*" "visual analog scale*" "instrument*" "tool*" "scale*" "inventor*" "list*" "test*" "VAS*" "PROM*" "subscale*" "index" "form*"

The subject headings and free text of the 3 concepts for the Cochrane Library were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
MeSH descriptor: [Inflammatory Bowel Diseases] explode all trees MeSH descriptor: [Colitis] explode all trees MeSH descriptor: [Crohn Disease] explode all trees MeSH descriptor: [Ileitis] explode all trees MeSH descriptor: [Proctocolitis] explode all trees MeSH descriptor: [Enterocolitis] explode all trees MeSH descriptor: [Proctitis] explode all trees MeSH descriptor: [Enteritis] explode all trees (inflammatory bowel disease*):ti,ab,kw (Word variations have been searched) (IBD):ti,ab,kw (Word variations have been searched) (Colitis*):ti,ab,kw (Word variations have been searched) (Crohn*):ti,ab,kw (Word variations have been searched) (jejunoileitis):ti,ab,kw (Word variations have been searched) (ileocolitis):ti,ab,kw (Word variations have been searched) (ileitis):ti,ab,kw (Word variations have been searched) (Proctocolitis):ti,ab,kw (Word variations have been searched) (Proctosigmoiditis):ti,ab,kw (Word variations have been searched) (Enterocolitis):ti,ab,kw (Word variations have been searched)	MeSH descriptor: [Fatigue] explode all trees MeSH descriptor: [Lethargy] explode all trees MeSH descriptor: [Asthenia] explode all trees MeSH descriptor: [Apathy] explode all trees MeSH descriptor: [Muscle Fatigue] explode all trees MeSH descriptor: [Mental Fatigue] explode all trees (fatigue*):ti,ab,kw (Word variations have been searched) (letharg*):ti,ab,kw (Word variations have been searched) (astheni*):ti,ab,kw (Word variations have been searched) (apath*):ti,ab,kw (Word variations have been searched) (fatigability):ti,ab,kw (Word variations have been searched) (wear*):ti,ab,kw (Word variations have been searched) (tire*):ti,ab,kw (Word variations have been searched) (exhaust*):ti,ab,kw (Word variations have been searched) (lacklustre):ti,ab,kw (Word variations have been searched)	MeSH descriptor: [Self-Assessment] explode all trees MeSH descriptor: [Patient Outcome Assessment] explode all trees MeSH descriptor: [Patient Reported Outcome Measures] explode all trees MeSH descriptor: [Self Report] explode all trees MeSH descriptor: [Surveys and Questionnaires] explode all trees MeSH descriptor: [Checklist] explode all trees MeSH descriptor: [Visual Analog Scale] explode all trees (self assess*):ti,ab,kw (Word variations have been searched) (patient assess*):ti,ab,kw (Word variations have been searched) (patient report*):ti,ab,kw (Word variations have been searched) (self report*):ti,ab,kw (Word variations have been searched) (survey*):ti,ab,kw (Word variations have been searched) (questionnaire*):ti,ab,kw (Word variations have been searched) (assess*):ti,ab,kw (Word variations have been searched) (clinical tool*):ti,ab,kw (Word variations have been searched) (measur*):ti,ab,kw (Word variations have been searched) (checklist):ti,ab,kw (Word variations have been searched) (visual analog scale):ti,ab,kw (Word variations have been searched) (instrument):ti,ab,kw (tool*):ti,ab,kw (Word variations have been searched) (scale*):ti,ab,kw (Word variations have been searched)

(Pancolitis):ti,ab,kw (Word variations have been searched) (Proctitis):ti,ab,kw (Word variations have been searched) (Enteritis):ti,ab,kw (Word variations have been searched)	(languor):ti,ab,kw (Word variations have been searched) (listless*):ti,ab,kw (Word variations have been searched) (lassitude):ti,ab,kw (Word variations have been searched) (((low or lack or loss or lost) N3 (energy or vigo*r))):ti,ab,kw (Word variations have been searched)	(inventor*):ti,ab,kw (Word variations have been searched) (list*):ti,ab,kw (Word variations have been searched) (test*):ti,ab,kw (Word variations have been searched) (VAS*):ti,ab,kw (Word variations have been searched) (PROM*):ti,ab,kw (Word variations have been searched) (subscale*):ti,ab,kw (Word variations have been searched) (index):ti,ab,kw (Word variations have been searched) (form*):ti,ab,kw (Word variations have been searched)

The subject headings and free text of the 3 concepts for the Health and Psychosocial Instruments database were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
inflammatory bowel disease.mp. IBD.mp. Colitis*.mp. Crohn*.mp. jejunoileitis.mp. ileocolitis.mp. ileitis.mp. Proctocolitis.mp. Proctosigmoiditis.mp. Enterocolitis.mp. Pancolitis.mp. Proctitis.mp. Enteritis.mp.	fatigue*.mp. letharg*.mp. astheni*.mp. apath*.mp. fatigability.mp. wear*.mp. tire*.mp. exhaust*.mp. lacklustre.mp. languor.mp. listless*.mp. lassitude.mp. ((low or lack or loss or lost) adj3 (energy or vigo*r)).mp.	patient report*.mp. self report*.mp. survey*.mp. questionnaire*.mp. assess*.mp. clinical tool*.mp. Measur*.mp. checklist*.mp. Visual Analog Scale*.mp. instrument*.mp. tool*.mp. scale*.mp. inventor*.mp. list*.mp. test*.mp. VAS*.mp. PROM*.mp. subscale* index.mp. form*.mp.

The subject headings and free text of the 3 concepts for the Mental Measurement Yearbook database were as follows.

Concept 1 – Inflammatory bowel disease	Concept 2 – Fatigue	Concept 3 – Measure
inflammatory bowel disease*.mp. IBD.mp. Colitis*.mp Crohn*.mp. jejunoileitis.mp. ileocolitis.mp. ileitis.mp. proctocolitis.mp. Proctosigmoiditis.mp. enterocolitis.mp. Pancolitis.mp. proctitis.mp. Enteritis.mp.	fatigue*.mp. letharg*.mp. astheni*.mp. apath*.mp. fatigability.mp. wear*.mp. tire*.mp. exhaust*.mp. lacklustre.mp. languor.mp. listless*.mp. lassitude.mp. ((low or lack or loss or lost) adj3 (energy or vigo*r)).mp.	patient report*.mp. self report*.mp. survey*.mp. questionnaire*.mp. assess*.mp. clinical tool*.mp. Measur*.mp. checklist*.mp. Visual Analog Scale*.mp. instrument*.mp. tool*.mp. scale*.mp. inventor*.mp. list*.mp. test*.mp. VAS*.mp. PROM*.mp. subscale* index.mp. form*.mp.

Supplementary Data Content 2. Search Strategy for the second search in identifying studies on psychometric properties of inflammatory bowel disease (IBD)-related fatigue patient-reported outcome measures

The combination of the following 3 concepts were run in the search: Name and the abbreviation of the IBD-related fatigue patient-reported outcome measures, IBD, and the psychometric properties filters suggested by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN).[1] The psychometric properties include validity, reliability, responsiveness, interpretability, and feasibility.

The search strategies for MEDLINE (OVID), EMBASE (OVID), CINHAL (via ProQuest), and PsycINFO are as follows.

Here are the subject headings and free text of the 3 concepts for MEDLINE (OVID).

Concept 1 Name and the abbreviation of the identified fatigue measure
e.g. Multidimensional Fatigue Inventory.mp., MFI.mp.

Concept 2 - IBD
exp inflammatory bowel disease/
exp Colitis/
exp Colitis, Ulcerative/
exp Crohn Disease/
Ileitis/
exp Proctocolitis/
Exp enterocolitis /
Exp Proctitis/
Exp Enteritis/
inflammatory bowel disease.mp.
IBD.mp.
Colitis*.mp.
Crohn*.mp.
jejunoileitis.mp.
ileocolitis.mp.
ileitis.mp.
Proctocolitis.mp.
Proctosigmoiditis.mp.
Enterocolitis.mp.
Pancolitis.mp.
Proctitis.mp.
Enteritis.mp.

Concept 3 - Psychometric properties
1.(instrumentation or methods).fs.
2. (Validation Studies or Comparative Study).pt.
3. exp Psychometrics/
4. psychometr*.ti, ab.
5. (clinimetr* or clinometr*).tw.
6. exp "Outcome Assessment (Health Care)"/
7. outcome assessment.ti, ab.
8. outcome measure*.tw.
9. exp Observer Variation/
10. observer variation.ti, ab.
11. exp Health Status Indicators/
12. exp "Reproducibility of Results"/
13. reproducib*.ti, ab.
14. exp Discriminant Analysis/
15. (reliab* or unreliab* or valid* or coefficient or homogeneity or homogeneous or "internal consistency").ti, ab.

16. (cronbach* and (alpha or alphas)).ti, ab.
17. (item and (correlation* or selection* or reduction*)).ti, ab.
18. (agreement or precision or imprecision or "precise values" or test-retest).ti, ab.
19. (test and retest).ti, ab.
20. (reliab* and (test or retest)).ti, ab.
21. (stability or interrater or inter-rater or intrarater or intra-rater or intertester or inter-tester or intratester or intra-tester or interobserver or inter-observer or intraobserver or intraobserver or intertechnician or inter-technician or intratechnician or intra-technician or interexaminer or inter-examiner or intraexaminer or intra-examiner or interassay or interassay or intraassay or intra-assay or interindividual or inter-individual or intraindividual or intra-individual or interparticipant or inter-participant or intraparticipant or intra-participant or kappa or kappa's or kappas or repeatab*).ti, ab.
22. ((replicab* or repeated) and (measure or measures or findings or result or results or test or tests)).ti, ab.
23. (generaliza* or generalisa* or concordance).ti, ab.
24. (intraclass and correlation*).ti, ab.
25. (discriminative or "known group" or factor analysis or factor analyses or dimension* or subscale*).ti, ab.
26. (multitrait and scaling and (analysis or analyses)).ti, ab.
27. (item discriminant or interscale correlation* or error or errors or "individual variability").ti, ab.
28. (variability and (analysis or values)).ti, ab.
29. (uncertainty and (measurement or measuring)).ti, ab.
30. ("standard error of measurement" or sensitiv* or responsive*).ti, ab.
31. ((minimal or minimally or clinical or clinically) and (important or significant or detectable) and (change or difference)).ti, ab.
32. (small* and (real or detectable) and (change or difference)).ti, ab.
33. (meaningful change or "ceiling effect" or "floor effect" or "Item response model" or IRT or Rasch or "Differential item functioning" or DIF or "computer adaptive testing" or "item bank" or "cross-cultural equivalence").ti, ab.
34. specificity.ti, ab.
35. sensitivity.ti, ab.
34. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35

Here are the subject headings and free text of the 3 concepts for EMBASE (OVID).

Concept 1 Name and the abbreviation of the identified fatigue measure
e.g. Multidimensional Fatigue Inventory.mp., MFI.mp.

Concept 2 - IBD
exp inflammatory bowel disease/
exp colitis/
exp ulcerative colitis/
exp Crohn disease/
exp ileitis/
exp proctocolitis/
exp enterocolitis/
exp proctitis/
exp enteritis/
exp colon Crohn disease/

inflammatory bowel disease*.mp.
IBD.mp.
Colitis*.mp
Crohn*.mp.
jejunoileitis.mp.
ileocolitis.mp.
ileitis.mp.
proctocolitis.mp.
Proctosigmoiditis.mp.
enterocolitis.mp.
Pancolitis.mp.
proctitis.mp.
Enteritis.mp.

Concept 3 - Psychometric properties
‘intermethod comparison’/exp OR ‘data collection method’/exp OR ‘validation study’/exp OR ‘feasibility study’/exp OR ‘pilot study’/exp OR ‘psychometry’/exp OR ‘reproducibility’/exp OR reproducib*:ab,ti OR ‘audit’:ab,ti OR psychometr*:ab,ti OR clinimetr*:ab,ti OR clinometr*:ab,ti OR ‘observer variation’/exp OR ‘observer variation’:ab,ti OR ‘discriminant analysis’/exp OR ‘validity’/exp OR reliab*:ab,ti OR valid*:ab,ti OR ‘coefficient’:ab,ti OR ‘internal consistency’:ab,ti OR (cronbach*:ab,ti AND (‘alpha’:ab,ti OR ‘alphas’:ab,ti)) OR ‘item correlation’:ab,ti OR ‘item correlations’:ab,ti OR ‘item selection’:ab,ti OR ‘item selections’:ab,ti OR ‘item reduction’:ab,ti OR ‘item reductions’:ab,ti OR ‘agreement’:ab,ti OR ‘precision’:ab,ti OR ‘imprecision’:ab,ti OR ‘precise values’:ab,ti OR ‘test-retest’:ab,ti OR (‘test’:ab,ti AND ‘retest’:ab,ti) OR (reliab*:ab,ti AND (‘test’:ab,ti OR ‘retest’:ab,ti)) OR ‘stability’:ab,ti OR ‘interrater’:ab,ti OR ‘inter-rater’:ab,ti OR ‘intrarater’:ab,ti OR ‘intra-rater’:ab,ti OR ‘intertester’:ab,ti OR ‘inter-tester’:ab,ti OR ‘intratester’:ab,ti OR ‘intratester’:ab,ti OR ‘interobeserver’:ab,ti OR ‘inter-observer’:ab,ti OR ‘intraobserver’:ab,ti OR ‘intraobserver’:ab,ti OR ‘intertechnician’:ab,ti OR ‘inter-technician’:ab,ti OR

‘intratechnician’:ab,ti OR ‘intratechnician’:ab,ti OR ‘interexaminer’:ab,ti OR ‘inter-examiner’:ab,ti OR ‘intraexaminer’:ab,ti OR ‘intraexaminer’:ab,ti OR ‘interassay’:ab,ti OR ‘inter-assay’:ab,ti OR ‘intraassay’:ab,ti OR ‘intra-assay’:ab,ti OR ‘interindividual’:ab,ti OR ‘inter-individual’:ab,ti OR ‘intraindividual’:ab,ti OR ‘intra-individual’:ab,ti OR ‘interparticipant’:ab,ti OR ‘inter-participant’:ab,ti OR ‘intraparticipant’:ab,ti OR ‘intraparticipant’:ab,ti OR ‘kappa’:ab,ti OR ‘kappas’:ab,ti OR ‘coefficient of variation’:ab,ti OR repeatab*:ab,ti OR (replicab*:ab,ti OR ‘repeated’:ab,ti AND (‘measure’:ab,ti OR ‘measures’:ab,ti OR ‘findings’:ab,ti OR ‘result’:ab,ti OR ‘results’:ab,ti OR ‘test’:ab,ti OR ‘tests’:ab,ti)) OR generaliza*:ab,ti OR generalisa*:ab,ti OR ‘concordance’:ab,ti OR (‘intraclass’:ab,ti AND correlation*:ab,ti) OR ‘discriminative’:ab,ti OR ‘known group’:ab,ti OR ‘factor analysis’:ab,ti OR ‘factor analyses’:ab,ti OR ‘factor structure’:ab,ti OR ‘factor structures’:ab,ti OR ‘dimensionality’:ab,ti OR subscale*:ab,ti OR ‘multitrait scaling analysis’:ab,ti OR ‘multitrait scaling analyses’:ab,ti OR ‘item discriminant’:ab,ti OR ‘interscale correlation’:ab,ti OR ‘interscale correlations’:ab,ti OR (‘error’:ab,ti OR ‘errors’:ab,ti AND (measure*:ab,ti OR correlat*:ab,ti OR evaluat*:ab,ti OR ‘accuracy’:ab,ti OR ‘accurate’:ab,ti OR ‘precision’:ab,ti OR ‘mean’:ab,ti)) OR ‘individual variability’:ab,ti OR ‘interval variability’:ab,ti OR ‘rate variability’:ab,ti OR ‘variability analysis’:ab,ti OR (‘uncertainty’:ab,ti AND (‘measurement’:ab,ti OR ‘measuring’:ab,ti)) OR ‘standard error of measurement’:ab,ti OR sensitiv*:ab,ti OR responsive*:ab,ti OR (‘limit’:ab,ti AND ‘detection’:ab,ti) OR ‘minimal detectable concentration’:ab,ti OR interpretab*:ab,ti OR (small*:ab,ti AND (‘real’:ab,ti OR ‘detectable’:ab,ti) AND (‘change’:ab,ti OR ‘difference’:ab,ti)) OR ‘meaningful change’:ab,ti OR ‘minimal important change’:ab,ti OR ‘minimal important difference’:ab,ti OR ‘minimally important change’:ab,ti OR ‘minimally important difference’:ab,ti OR ‘minimal detectable change’:ab,ti OR ‘minimal detectable difference’:ab,ti OR ‘minimally detectable change’:ab,ti OR ‘minimally detectable difference’:ab,ti OR ‘minimal real change’:ab,ti OR ‘minimal real difference’:ab,ti OR ‘minimally real change’:ab,ti OR ‘minimally real difference’:ab,ti OR ‘ceiling effect’:ab,ti OR ‘floor effect’:ab,ti OR ‘item response model’:ab,ti OR ‘irt’:ab,ti OR ‘rasch’:ab,ti OR ‘differential item functioning’:ab,ti OR ‘dif’:ab,ti OR ‘computer adaptive testing’:ab,ti OR ‘item bank’:ab,ti OR ‘cross-cultural equivalence’:ab,ti

Here are the subject headings and free text of the 3 concepts for CINAHL.

Concept 1 Name and the abbreviation of the identified fatigue measure
e.g. Multidimensional Fatigue Inventory.mp., MFI.mp.

Concept 2 - IBD
MH 'Inflammatory Bowel Diseases+"
MH "Colitis+"
MH "Colitis, Ulcerative"
MH "Crohn Disease"
MH "Ileitis+"
MH "Enterocolitis+"
MH "Enteritis+"

"inflammatory bowel disease*"
"IBD"
"colitis*"
"crohn*"
"jejunoileitis"
"ileocolitis"
"ileitis"
"proctocolitis"
"Proctosigmoiditis"
"enterocolitis"
"Pancolitis"
"proctitis"
"Enteritis"

Concept 3 - Psychometric properties
(MH "Psychometrics") or (TI psychometr* or AB psychometr*) or (TI clinimetr* or AB clinimetr*) or (TI clinometr* OR AB clinometr*) or (MH "Outcome Assessment") or (TI outcome assessment or AB outcome assessment) or (TI outcome measure* or AB outcome measure*) or (MH "Health Status Indicators") or (MH "Reproducibility of Results") or (MH "Discriminant Analysis") or ((TI reproducib* or AB reproducib*) or (TI reliab* or AB reliab*) or (TI unreliab* or AB unreliab*)) or ((TI valid* or AB valid*) or (TI coefficient or AB coefficient) or (TI homogeneity or AB homogeneity)) or (TI homogeneous or AB homogeneous) or (TI "coefficient of variation" or AB "coefficient of variation") or (TI "internal consistency" or AB "internal consistency") or (MH "Internal Consistency+") or (MH "Reliability+") or (MH "Measurement Error+") or (MH "Content Validity+") or "hypothesis testing" or "structural validity" or "cross-cultural validity" or (MH "Criterion-Related Validity+") or "responsiveness" or "interpretability" or (TI reliab* or AB reliab*) and ((TI test or AB test) OR (TI retest or AB retest)) or (TI stability or AB stability) or (TI interrater or AB interrater) or (TI inter-rater or AB inter-rater) or (TI intrarater or AB intrarater) or (TI intra-rater or AB intrarater) or (TI intertester or AB intertester) or (TI inter-tester or AB inter-tester) or (TI intratester or AB intratester) or (TI intra-tester or AB intra-tester) or (TI interobserver or AB interobserver) or (TI inter-observer or AB inter-observer) or (TI intraobserver or AB

intraobserver) or (TI intra-observer or AB intra-observer) or (TI intertechnician or AB intertechnician) or (TI inter-technician or AB inter-technician) or (TI intratechnician or AB intratechnician) or (TI intra-technician or AB intra-technician) or (TI interexaminer or AB interexaminer) or (TI inter-examiner or AB inter-examiner) or (TI intraexaminer or AB intraexaminer) OR (TI intra-examiner or AB intra-examiner) or (TI intra-examiner or AB intraexaminer) or (TI interassay or AB interassay) or (TI inter-assay or AB inter-assay) or (TI intraassay or AB intraassay) or (TI intra-assay or AB intra-assay) or (TI interindividual or AB interindividual) or (TI inter-individual or AB inter-individual) OR (TI intraindividual or AB intraindividual) or (TI intra-individual or AB intra-individual) or (TI interparticipant or AB interparticipant) or (TI inter-participant or AB inter-participant) or (TI intraparticipant or AB intraparticipant) or (TI intra-participant or AB intra-participant) or (TI kappa or AB kappa) or (TI kappa's or AB kappa's) or (TI kappas or AB kappas) or (TI repeatab* or AB repeatab*) or (TI responsive* or AB responsive*) or (TI interpretab* or AB interpretab*)

Here are the subject headings and free text of the 3 concepts for PsychINFO.

Concept 1 Name and the abbreviation of the identified fatigue measure
e.g. Multidimensional Fatigue Inventory.mp., MFI.mp.

Concept 2 - IBD
exp Colitis/
exp Ulcerative Colitis/

inflammatory bowel disease*.mp.
IBD.mp.
Colitis*.mp
Crohn*.mp.
jejunoileitis.mp.
ileocolitis.mp.
ileitis.mp.
proctocolitis.mp.
Proctosigmoiditis.mp.
enterocolitis.mp.
Pancolitis.mp.
proctitis.mp.
Enteritis.mp.

Concept 3 - Psychometric properties
cl("Psychometrics & Statistics & Methodology" OR "Research Methods & Experimental Design") OR (psychometr* OR clinimetr* OR clinometr* OR "outcome assessment" OR "outcome measure*" OR "observer variation" OR reproducib* OR reliab* OR unreliab* OR valid* OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR reliab* OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa's OR kappas OR repeatab* OR generaliza* OR generalisa* OR concordance OR discriminative OR "known group" OR "factor analys*" OR dimension* OR subscale* OR "item discriminant" OR "interscale correlation*" OR error* OR "individual variability" OR "standard error of measurement" OR sensitiv* OR responsive* OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence") OR ("cronbach* alpha*" OR "replicab* measure*" OR "replicab* finding*" OR "replicab* result*" OR "replicab* test*" OR "repeated measure*" OR "repeated finding*" OR "repeated result*" OR "repeated test*" OR "item correlation*" OR "item selection*" OR "item reduction*" OR "Test retest" OR "intraclass correlation*" OR "multitrait scaling analys*" OR "uncertainty measur*" OR "variability analys*" OR "variability

value*” OR “minimal* important change” OR “minimal* important difference” OR “minimal* significant change” OR “minimal* significant difference” OR “minimal* detectable change” OR “minimal* detectable difference” OR “clinical* important change” OR “clinical* important difference” OR “clinical* significant change” OR “clinical* significant difference” OR “clinical* detectable change” OR “clinical* detectable difference” OR “small* real change” OR “small* real difference” OR “small* detectable change” OR “small* detectable difference”) OR (SU.EXACT.EXPLODE(“Measurement”) OR SU.EXACT.EXPLODE(“Error Analysis”) OR SU.EXACT.EXPLODE(“Test Construction”) OR SU.EXACT.EXPLODE(“Interrater Reliability”) OR SU.EXACT.EXPLODE(“Content Analysis”) OR SU.EXACT.EXPLODE(“Error of Measurement”) OR SU.EXACT.EXPLODE(“Factor Structure”) OR SU.EXACT.EXPLODE(“Testing Methods”) OR SU.EXACT.EXPLODE(“Statistical Reliability”) OR SU.EXACT.EXPLODE(“Consistency (Measurement)”) OR SU.EXACT.EXPLODE(“Computer Assisted Testing”) OR SU.EXACT.EXPLODE(“Factor Analysis”) OR SU.EXACT.EXPLODE(“Prediction”) OR SU.EXACT.EXPLODE(“Statistical Validity”) OR SU.EXACT.EXPLODE(“Prediction Errors”))

Supplementary Data Content 3. Updated COSMIN Criteria for Good measurement[2,3]

Measurement property	Definitions	Rating†	Criteria
Validity			
Content Validity	The degree to which the content of a PROM is an adequate reflection of the construct to be measured	+	‡The ratings on relevance, comprehensiveness and the comprehensibility are +
		-	‡The ratings on relevance, comprehensiveness and the comprehensibility are -
		±	‡Unexplained inconsistency of ratings on relevance, comprehensiveness and the comprehensibility across studies
Hypotheses testing for construct validity	The degree to which the scores of a PROM are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the instrument validly measures the construct to be measured	+	The result is in accordance with the hypotheses (as stated in Supporting information 4) §
		?	No hypothesis defined (by the review team) §
		-	The result is not in accordance with our hypothesis §
Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured	+	**CTT:** CFA: CFI or TLI or comparable measure >0.95 OR RMSEA <0.06 OR SRMR <0.08 **IRT/Rasch:** No violation of unidimensionality : CFI or TLI or comparable measure > 0.95 OR RMSEA < 0.06 OR SRMR < 0.08 AND no violation of local independence: residual correlations among the items after controlling for the dominant factor < 0.20 OR Q3's < 0.37 AND no violation of monotonicity: adequate looking graphs OR item scalability > 0.30

			AND adequate model fit IRT: $\chi 2 > 0.001$ Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardised values > −2 and < 2 **EFA:** The choice of EFA, the criteria for retaining factors, the number of factors to be rotated, the choice of the rotation method, interpretation of the final factor structure are well-justified. The sample size is sufficient to support the use of EFA. The missing data and the distribution of the data are properly handled and examined to allow a valid EFA. Eigenvalues or percentages of variance are reported.[4]
		?	CTT/EFA: Not all information for '+' reported IRT/Rasch: model fit not reported
		-	Criteria for '+' not met
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 important differences found between group factors (such as age, gender, language) in multiple group factor analysis OR no important DIF for group factors (McFadden's $R^2 < 0.02$)
		?	No multiple group factor analysis OR DIF analysis performed
		-	Important differences between group factors OR DIF was found
Reliability			
Internal consistency	The degree of the interrelatedness among the items	+	At least low evidence ¶ for sufficient structural validity AND Cronbach's alpha(s) ≥ 0.70 for each unidimensional scale or subscale
		?	Criteria for "At least low evidence ¶ for sufficient structural validity" not met
		-	At least low evidence ¶ for sufficient structural validity AND Cronbach's alpha(s) < 0.70 for each unidimensional scale or subscale
Test-retest reliability	The extent to which scores for patients who have not changed are the same for repeated measurement under over time (test-retest).	+	ICC or weighted Kappa ≥ 0.70
		?	ICC or weighted Kappa not reported
		-	ICC or weighted Kappa < 0.70
Measurement error	The systematic and random error of a patient's	+	SDC or LoA < MIC
		?	MIC not defined

	score that is not attributed to true changes in the construct to be measured.	-	SDC or LoA > MIC
Responsiveness			
Responsiveness	The ability of an instrument to detect change over time in the construct to be measured	+	The result is in accordance with the hypothesis (as stated in Supporting information 4) [§]
		?	No hypothesis defined (by the review team) [§]
		-	The result is not in accordance with the hypothesis[§]

This table was extracted from Table 1 of Terwee and Prinsen et al.'s work. [2,3] Criteria for exploratory factor analysis was added for structural validity.[4]

[†] The result on the psychometric property was rated as either (+) sufficient, (-) insufficient, (?) indeterminate, or (±) inconsistent.

[‡]Criteria for content validity were extracted from the User manual of COSMIN methodology for assessing the content validity of PROMs.[5]

[§]The results of all studies are summarized and then evaluated based on 75% of the results.

[¶] The low evidence for sufficient structural validity was defined according to the modified GRADE approach.

Abbreviation: PROM – Patient-reported outcome measure, CTT - classical test theory, CFA - confirmatory factor analysis, CFI - comparative fit index, TLI - Tucker-Lewis index, RMSEA - Root Mean Square Error of Approximation, SRMR – Standardised root mean residual, IRT – Item Response Theory, EFA – Exploratory factor analysis, DIF - differential item functioning, ICC - intraclass correlation coefficient, SDC - smallest detectable change, LoA - limits of agreement, MIC - minimal important change.

Supplementary Data Content 4. Construct validity and responsiveness testing: Hypotheses sets for (sub)scales with different sub-construct of fatigue

For (sub)scale with general/ physical **fatigue experience** (e.g. I feel weak/ lack of energy) as key construct: • IBD-Fatigue scale (Section 1) • Multidimensional Fatigue Inventory – Physical fatigue, General fatigue subscales • Fatigue questionnaire – Physical fatigue subscale
1. Correlations with (changes in) assessments measuring similar construct of 'fatigue' (i.e. impact of fatigue) should be greater ≥ 0.5.
2. Correlations between (changes in) assessments measuring IBD-specific quality of life (assessed by IBD-specific quality of life measure) and disease activity should be ≥ 0.5.
3. Correlations between (changes in) assessments measuring related, but dissimilar constructs should be ≥ 0.3. A related but dissimilar construct refers to physical health, mental health, reduced motivation, emotional fatigue, mental fatigue, and health-related quality of life measured by generic measures.
For (sub)scale with **mental/ cognitive fatigue** as key construct: • Multidimensional Fatigue Inventory – Mental fatigue subscale • Fatigue questionnaire – Mental Fatigue subscale
1. Correlations with (changes in) assessments measuring similar construct of 'fatigue' (i.e. the impact of fatigue) should be ≥ 0.5.
2. Correlations between (changes in) assessments measuring related, but dissimilar constructs should be ≥ 0.3. A related but dissimilar construct refers to the general/ physical fatigue experience and general mental health.
For (sub)scale with **emotional fatigue (i.e. reduced motivation)** as key construct • Multidimensional Fatigue Inventory – Reduced motivation subscale
1. Correlations with (changes in) assessments measuring similar construct of 'fatigue' (i.e. impact of fatigue) should be ≥ 0.5.
2. Correlations between (changes in) assessments measuring related, but dissimilar constructs should be ≥ 0.3. A related but dissimilar construct refers to the general/ physical fatigue experience.
For (sub)scale with **impact of fatigue** (e.g. Fatigue interferes with my work, family, or life) as key construct: • Fatigue Severity Scale • Daily Fatigue Impact Scale

• Functional Assessment of Chronic Illness Therapy-Fatigue • IBD-Fatigue scale (Section 2) • Multidimensional Fatigue Inventory – Reduced activity subscale • Multidimensional Assessment of Fatigue • Modified Fatigue Impact Scale • Fatigue questionnaire (Total score)
1. Correlations with (changes in) assessments measuring similar construct of 'fatigue' should be ≥ 0.5. Similar construct of fatigue refers to reduced activity associated with fatigue, physical fatigue, mental fatigue, emotional fatigue, and the general/ physical fatigue experience.
2. Correlations between (changes in) assessments measuring disease activity and IBD-specific quality of life (assessed by IBD-specific quality of life measure) should be ≥ 0.5.
3. Correlations between (changes in) assessments measuring health-related quality of life (assessed by generic measures) should be ≥ 0.3.
For all (sub)scales of fatigue:
1. Correlations with (changes in) assessment measuring unrelated constructs should be <0.3. 2. Subgroup with active disease state should experience a higher level of fatigue than those in remission/ healthy controls. 3. The change score of the disease activity should correspond to the change score of fatigue measure in terms of the direction and magnitude.

Supplementary Data Content 5. Instructions on downgrading the quality of evidence on the overall psychometric robustness of patient-reported outcome measures using the modified GRADE approach [2,6]

The quality of evidence was assumed to be high and downgraded by different levels (-1/-2/-3), based on the following 3 GRADE factors. [2,6]

GRADE factors	Reasons and levels for downgrading (-1/-2/-3 level)
1. Risk of Bias based on the methodological quality of the study	No risk of bias: One study with very good quality or multiple studies with at least adequate quality Serious risk of bias (-1): Only one study of adequate quality or multiple studies of doubtful quality available Very serious risk of bias (-2): Only one study of doubtful quality or multiples studies of inadequate quality available Extremely serious risk of bias (-3): Only one study of inadequate quality available
2. Inconsistency	The quality of evidence will be downgraded by the unexplained inconsistency across results. Serious (-1), Very serious (-2) inconsistency: The level of inconsistency was determined by the review team.
3. Imprecision †	Serious imprecision (-1): If the pooled sample size is between 50 - 100. Very serious imprecision (-2): If the pooled sample size is < 50.

† The principle of imprecision did not apply to content validity, structural validity, and cross cultural validity because the sample size was already taken into account in the Risk of Bias Checklist. [2]

Supplementary Data Content 6. Characteristics of the study samples

	Population			Disease characteristics				Instrument administration			
PROM Study reference	**Sample size, n**	**Age Mean (SD)[range] year**	**Gender % female**	**Disease**	**Disease duration mean (SD)[range] year**	**Disease indices used**	**Disease severity**	**Setting**	**Country**	**Language**	**Response rate**
Unidimensional measures											
FSS (Spanish), D-FIS (Spanish)[7]	99 (CD: 55, UC: 44)	CD: 34 UC: 43	CD: 56.4%, UC: 56.8%	CD, UC	CD: 7. UC: 10	Harvey-Bradshaw Index for CD: Inactive – score < 3 Colitis Activity Index for UC Inactive – score <6	CD active: 50.9% CD inactive: 49% UC active: 54.5% UC inactive: 45.5%	The Crohn-Colitis Care Unit of the Hospital Universitari Vall d'Hebron	Spain	English	Not given
FACIT-F [8]	209 (CD: 132, UC: 77)	CD: 38.2 (12.2) UC: 39.5 (13.3)	CD: 53.8% UC: 51.9%	CD, UC	Not given	Harvey Bradshaw Index for CD: Active - score ≥ 5, Inactive – score ≤ 4 Simple Clinical Colitis Activity Index: Active - score ≥ 5, Inactive – score ≤ 4	Active CD: 15.2% Active UC: 19.5%	Massachusetts General Hospital Crohn's and Colitis Center	United States	English	Not given
Multidimensional measure											
IBD-F (English) [9]	Phase 1: item generation – 20 (CD: 11, UC: 9)	48.8 (14.9)	50%	CD, UC	9.4 [1-37]	Harvey Bradshaw Index for CD: Active - score ≥ 5, Inactive – score ≤ 4 Simple Clinical Colitis	Active CD: 88% Inactive CD: 18% Active UC: 78% Inactive UC: 22%	Postal questionnaire to the registered IBD members in Crohn's and Colitis UK	United Kingdom	English	77%

	Phase 2: face and content validity –16	54.3 (18.0)	50%	CD, UC	14.4 [4-38]	Activity Index: Active - score ≥ 5, Inactive – score ≤ 4	Active CD: 62% Inactive CD: 38% Active UC: 100% Inactive UC: 0%				
	Phase 3: initial piloting – 30	49.8 (15.1)	50%	CD, UC	17.3 [3-49]		Active CD: 44% Inactive CD: 56% Active UC: 57% Inactive UC: 43%				
	Phase 4: test-retest reliability testing- 36	46.8 (13.8)	50%	CD, UC	10.5 [3-38]		Active CD: 33% Inactive CD: 67% Active UC: 23% Inactive UC: 77%				
	Phase 5: initial construct validity testing - 465	56.9 (13.8)	68.3%	CD, UC	26 [1-60]		Active CD: 69% Inactive CD: 31% Active UC: 76% Inactive UC: 43%				
IBD-F (Brazil) [10]	Phase 1 Translation & cross cultural adaptation: 42	CD: 38.1 (12), UC: 43.1 (14.3)	CD: 67.6% UC: 75%	CD, UC	CD: 84 months (63 months), UC: 100 months (92 months)	Harvey Bradshaw Index for CD: Active - score ≥ 5, Inactive – score ≤ 4	Active CD: n=14 Inactive CD: n=18 Mild UC: n=4	The Gastroenterology Outpatient clinic	Brazil	Brazilian-Portuguese	Not given

	(CD: 34, UC: 8)					Truelove and Witts for UC Inactive: ≤2 or 3 stools/day, without blood and/or pus in the stools, and no systemic symptoms Mild: 4 stools/day, with or without blood, no systemic involvement, and increased inflammatory markers Moderate: >4 stools/day with minimal systemic symptoms and increased inflammatory markers Severe: >6 stools/day with blood and evidence of systemic involvement, such as fever, tachycardia, anemia, and erythrocyte sedimentation above thirty	Moderate UC: n=3 Severe UC: n=0				
	Phase 2 Measurement properties testing: 118 (CD:81, UC:37)	CD: 43.1 (11), UC: 48.9 (11.7)	CD: 66.7% UC: 54.1%	CD, UC	CD: 108 months (82 months) UC: 132 months (99 months)		Active CD: n=34 Inactive CD: n=47 Mild UC: n=26 Moderate UC: n=10 Severe UC: n=1				96% completed the survey.

IBD-F (Greek) [11]	61 (CD: 42, UC:19)	35.6 (13.3)	39.30%	CD, UC	4.5 months, [1-4.5 months]	Harvey Bradshaw Index for CD: Active - score > 4 Inactive – score ≤ 4 Simple Clinical Colitis Activity Index: Active - score > 4, Inactive – score ≤ 4	CD: Median of Harvey Bradshaw Index [range]=2 [0-8] UC: Median of Simple clinical colitis activity index [range]=1.5 [0-10]	The IBD department of the hospital in Athens (including inpatients and outpatients)	Greece	Greek version	Not given
IBD- F (Danish) [12]	Reliability testing: 66 Convergent validity phase: 159	CD: 41 (31-53) [21-81] UC:40 (29-52) [20-79]	CD:58.1%, UC: 73.4 %	CD,UC	No information provided.	Self reported-disease activity - disease flare or disease in remission	1st timepoint (first test of test-retest reliability & convergent validity testing): Active CD:9.5% Active UC: 18% 2nd timepoint (retest reliability): Active CD: 13.3 % Active UC: 16.1%	Outpatient clinic of Aarhus University Hospital	Danish	Danish version, Electronic	54%
MFI, MAF [9]	465 (CD: 301,UC: 164	56.9 (13.8) [22-90]	69%	CD, UC	26 [1-60]	Harvey Bradshaw Index for CD: Active - score ≥ 5, Inactive – score ≤ 4 Simple Clinical Colitis Activity Index:	Inactive CD: 31%, Active CD: 69% Inactive UC: 43%, UC active: 57%	Postal questionnaire to the registered IBD members in Crohn's and Colitis UK	United Kingdom	English	77%

						Active - score ≥ 5, Inactive – score ≤ 4					
FQ (Norwegian) [13]	140 (CD: 48, UC: 92)	CD:40(15) [19-69]. UC: 46.9 (5.8)[20-82]	CD: 75%. UC: 47%	CD, UC	CD: 9.2 (9.6); UC: 8.5 (9.5)	Simple Crohn's Disease Activity Index for CD Simple Clinical Colitis Activity Index for UC (The authors did not provide the cutoff score to determine disease activity)	CD: Mean score (SD)=3.9 (2.7) (Patients were either in remission/ mild to moderate disease activity state; but not in the moderate-to-severe, severe-fulminant disease state) UC: Mean score (SD) =2.8 (2.4)	3 outpatient clinics in the Southeastern part of Norway	Norway	Norwegian	Not given
Abbreviation: PROM – Patient-reported outcome measure, SD – Standard deviation, IBD – Inflammatory bowel disease, CD – Crohn's disease subgroup, UC – Ulcerative colitis, FSS - Fatigue Severity Scale, D-FIS - Daily Fatigue Impact Scale, FACIT-F - Functional Assessment of Chronic Illness Therapy – Fatigue, IBD-F - Inflammatory bowel disease fatigue scale, MFI - Multidimensional Fatigue Inventory, MAF - Multidimensional Assessment of Fatigue; M-FIS - Modified Fatigue Impact Scale, FQ - Fatigue Questionnaire											

Supplementary Data Content 7. Content validity evaluation of the IBD-Fatigue (English) scale

	PROM development study[1]	**Content validity study**	**Rating of reviewers**	**Overall rating**	**Quality of evidence**
Criteria (column)/ **Rating** (row)	(+): sufficient, (-): insufficient, (±): inconsistent, (?): indeterminate				High, moderate, low, very low
Relevance					
1) Are the included items relevant for measuring fatigue?	+	?	+		
2) Are the included items relevant for the IBD patients?	+	?	+		
3) Are the included items relevant for the evaluative use of measure?	+	?	+		
4) Are the response options appropriate?	+	?	+		
5) Is the recall period appropriate?	+	?	+		
Relevance rating				+	High[†]
Comprehensiveness					
6) Are all key concepts included?	+	?	+		
Comprehensiveness rating				+	High[†]
Comprehensibility					
7) Are the PROM instructions understood by the IBD patients as intended?	+	?			
8) Are the PROM items and response options understood by the IBD patients as intended?	+	?			
9) Are the PROM items appropriately worded?			+		
10) Do the response options match the question?			+		
Comprehensibility rating				+	High[†]
Content validity rating				+	High[†]

[†] We did not downgrade the quality of evidence on content validity, despite the results on content validity was only found from the PROM development study (not from the content validity study). It is because the evidence on content validity from the PROM development study seems to be robust enough. The authors assessed relevance, comprehensibility, and comprehensiveness of the measure in 16 IBD patients through well justified and appropriate methods (i.e. cognitive interviews involving both 'think aloud' and verbal probing techniques, iterative testing and analysis using descriptive matrix, and the use of appropriate topic guide). The IBD-F draft was also reviewed by a group of experts (including gastroenterologists, IBD nurse, nurse academics, project researchers) through face-to-face meetings and emails. It is less likely to have the risk of bias on the items' content given that the relevant and important parties were involved in the design of the measure.

Supplementary Data Content 8. Construct validity (Convergent and known-groups validity) of the patient-reported outcome measures (PROMs): Results on the study quality, overall psychometric robustness, and its quality of evidence

PROMs	Pooled sample size	Study quality (Very Good, Adequate, Doubtful, Inadequate)	Convergent validity Comparator measure [Construct(s)] (Correlation)		Known-groups validity	No. of results met our hypotheses (%), by (sub)scale	Overall robustness on construct validity (+: sufficient, ±:inconsistent)	Quality of evidence (High, moderate, low, very low) (Reason of downgrading quality)
Unidimensional								
FSS (Spanish)	99	Very Good[7]	M-FIS, D-FIS [Fatigue impact] (0.77-0.82) [7] IBDQ-36 [IBD-related QoL] (-0.72) [7] CCAI/SCCAI [Disease activity]: CD (0.67), UC (0.5) [7]		Not assessed	5 out of 5 (100%)	+	Moderate (Serious imprecision, n<100)
D-FIS (Spanish)	236	Very Good[7] (Convergent) Very Good[7] (Known-group)	FSS, M-FIS [Fatigue impact] (0.82-0.84) [7] IBDQ-36 [IBD-related QoL] (-0.81) [7] CCAI/SCCAI [Disease activity]: CD (0.26-0.67), UC (0.34-0.5) [7]		Difference between IBD patient group and the healthy control group ($p<0.005$); active and inactive IBD groups ($p<0.001$).[7]	7 out of 9 (77.8%)	+	High
FACIT-F	209	Very Good[8] (Convergent) Doubtful[8] (Known-group)	HBI/SCCAI [Disease activity]: CD (-0.49), UC (-0.59)[8]		Significant difference between active and inactive CD (-4.6 points) and UC (-8.5 points) subgroups.[8]	3 out of 4 (75%)	+	High
Multidimensional								
IBD-F (English)	465	Very Good[9]	Section 1	MFI-20 [Fatigue experience† and impact] (0.44-0.73)[9]	Not assessed	10 out of 11 (90.9%)	+	High
			Section 2	MFI-20 [Fatigue experience† and impact] (0.56-0.78)[9]	Not assessed	11 out of 11 (100%)	+	High
IBD-F (Brazil)	118	Very Good[10]	Total score	FACIT-F [Fatigue impact] (-0.67)[10] HADS [Anxiety & Depression] (0.14)[10]	Not assessed	1 out of 2 (50%)	+ (Higher weight on correlation between fatigue measures)	High
IBD-F (Greek)	61	Very Good[11]	Section 1	FSS [Fatigue impact] (0.69) [11]	Not assessed	1 out of 1 (100%)	+	Moderate (Serious imprecision, n<100)

			Section 2	SIBDQ [IBD-related QoL](0.69) [11]	Not assessed	1 out of 1 (100%)	+	Moderate (Serious imprecision, n<100)
IBD-F (Danish)	159	Very Good[12]	Section 1	MFI-20 [Fatigue experience† and impact] (0.66-0.88)[12] SHS [Generic HRQoL] (0.31-0.61) [12]	Not assessed	10 out of 10 (100%)	+	High
			Section 2	MFI-20 [Fatigue experience† and impact] (0.69-0.85)[12], SHS [Generic HRQoL] (0.35-0.67)[12]	Not assessed	9 out of 10 (90%)	+	High
MFI-20	465	Very Good[9]	General fatigue	IBD-F Section 1 [Fatigue experience]†: CD (0.63), UC (0.66) [9] IBD-F Section 2, MAF [Fatigue impact]: CD (0.59-0.63), UC (0.62-0.66) [9]	Not assessed	6 out of 6 (100%)	+	High
			Physical fatigue	IBD-F Section 1 [Fatigue experience]†:CD (0.51), UC (0.54) [9] IBD-F Section 2, MAF [Fatigue impact]: CD (0.56-0.62), UC (0.59-0.61) [9]	Not assessed	6 out of 6 (100%)	+	High
			Reduced activity	IBD-F Section 1 [Fatigue experience]†: CD (0.48), UC (0.47) [9] IBD-F Section 2, MAF [Fatigue impact]: CD (0.54-0.61), UC (0.55-0.58) [9]	Not assessed	4 out of 6 (66.7%)	+	Moderate (Inconsistency across results)
			Reduced motivation	IBD-F Section 1 [Fatigue experience]†: CD (0.51), UC (0.49) [9] IBD-F Section 2, MAF [Fatigue impact]: CD (0.53-0.61), UC (0.48-0.57) [9]	Not assessed	5 out of 6 (83.3%)	+	High
			Mental fatigue	IBD-F Section 1 [Fatigue experience]†: CD(0.5), UC(0.44) [9] IBD-F Section 2, MAF [Fatigue impact]: CD (0.54-0.61), UC (0.37-0.56) [9]	Not assessed	5 out of 6 (83.3%)	+	High
MAF	465	Very Good[9]	Total score	IBD-F Section 1,2 MFI-20-General, physical fatigue, reduced activity subscales [Fatigue	Not assessed	12 out of 14 (85.7%)	+	High

				experience† and impact]: CD (0.54-0.8), UC (0.55-0.73) [9] MFI-20 – [Reduced motivation]: CD (0.53), UC (0.48) [9] MFI-20 – [Mental fatigue]: CD (0.54), UC (0.37) [9]				
M-FIS (Spanish)	99	Very Good [7]	Total score	FSS, D-FIS [Fatigue impact] (0.77-0.84) [7] IBDQ-36 [IBD-related QoL] (-0.76) [7] CDAI/ SCCAI [Disease activity]: CD (0.46), UC (0.39) [7]	Not assessed	5 out of 5 (100%)	+	Moderate (Serious imprecision, n<100)
FQ (Norwegian)	140	Very Good[13]	Physical fatigue	SF-36 - Physical and mental health (0.38-0.63) [13]	Able to discriminate between healthy controls and IBD patients ($p<0.001$) [13]	2 out of 2 (100%)	+	High
			Mental fatigue	SF-36 - Mental health (0.3-0.4) [13]	Unable to discriminate between healthy controls and IBD patients[13]	1 out of 2 (50%)	±	Not given
			Total score	Not assessed	Able to discriminate between healthy controls and IBD patients ($p<0.001$) [13]	1 out of 1 (100%)	+	High

† Fatigue experience refers to general/ physical fatigue experience (e.g. I feel weak), which is related but distinct from the impact of fatigue (e.g. Fatigue interferes with my work, family, or life).

Abbreviation: FSS - Fatigue Severity Scale, D-FIS - Daily Fatigue Impact Scale, CD – Crohn's disease, UC – Ulcerative colitis, CDAI - Crohn's Disease Activity Index (for CD), SCCAI - Simple Clinical Colitis Activity Index (for UC), IBDQ-36 – Inflammatory Bowel Disease Questionnaire, FACIT-F - Functional Assessment of Chronic Illness Therapy – Fatigue, HBI - Harvey-Bradshaw Index (for CD), IBD-F - Inflammatory bowel disease fatigue scale, MFI - Multidimensional Fatigue Inventory-20, HADS - Hospital Anxiety and Depression Scale, SIBDQ – Short Inflammatory Bowel Disease Questionnaire, SHS – Short Health Scale, HRQoL- Health related quality of life, MAF - Multidimensional Assessment of Fatigue; M-FIS - Modified Fatigue Impact Scale, FQ - Fatigue Questionnaire, SF36 - 36-Item Short Form Survey

Supplementary Data Content 9. Structural validity and internal consistency of the patient-reported outcome measures (PROMs): Results on the study quality, overall psychometric robustness, and its quality of evidence

Psychometric property	Structural validity					Internal consistency					
PROMs	**Pooled sample size**	**Study quality†**	**Results**	**Overall robustness‡**	**Quality of evidence§**	**Pooled sample size**	**Study quality†**	**Results** Cronbach's alpha		**Overall robustness‡**	**Quality of evidence§**
Unidimensional											
FSS (Spanish)	Not assessed					Not assessed					
D-FIS (Spanish)	Not assessed					Not assessed					
FACIT-F	Not assessed					209	Doubtful[8]	CD: 0.95 [8] , UC: 0.94 [8]		?	Not given
Multidimensional											
IBD-F (English)	465	Very Good[9]	Exploratory factor analysis: Two-factorial structure[9] Section 1:83-93% of the variants were explained across items Section 2: 63% of the variants were explained	+	High	465	Very Good[9]	Section 1 Section 2	0.91[9] 0.98[9]	+	High
IBD-F (Brazil)	Not assessed					118	Doubtful[10]	Section 1 Section 2	0.95[10] 0.98[10]	?	Not given
IBD-F (Greek)	61	Inadequate[11]	Section 1:88.3% of the variants were explained. [11] Structural validity was not assessed for the whole measure.	?	Not given	61	Doubtful[11]	Section 1 Section 2	0.90[11] 0.97[11]	?	Not given
IBD-F (Danish)	Not assessed					Not assessed					
MFI-20	Not assessed					Not assessed					
MAF	Not assessed					Not assessed					
M-FIS (Spanish)	Not assessed					Not assessed					

FQ (Norwegian)	140	Adequate[13]	Confirmatory factor analysis: Two-factorial structure [13] ; Unclear how well the data adequately fit the hypothesized model	?	Not given	140	Doubtful[13]	Physical Fatigue Mental Fatigue Total Fatigue	0.89[13] 0.73[13] 0.89[13]	?	Not given

† Study quality was rated either 'Very Good', 'Adequate', 'Doubtful' or 'Inadequate', following the COSMIN Risk of Bias Checklist.[12]
‡ Overall psychometric robustness was rated either sufficient (+), insufficient (-) or indeterminate (?), based on the updated COSMIN criteria for good measurement property.[2]
§ The quality of evidence was graded either 'high', 'moderate', 'low', or 'very low'.[2] Rating on the quality of evidence would not be given if the psychometric robustness was indeterminate. Abbreviation: FSS-Fatigue Severity Scale, D-FIS-Daily Fatigue Impact Scale, FACIT-F- Functional Assessment of Chronic Illness Therapy-Fatigue, CD-Crohn's disease, UC-Ulcerative colitis, IBD-F-Inflammatory bowel disease fatigue scale, MFI-20 - Multidimensional Fatigue Inventory-20, MAF-Multidimensional Assessment of Fatigue; M-FIS- Modified Fatigue Impact Scale, FQ-Fatigue Questionnaire

Supplementary Data Content 10. Test-retest reliability of the patient-reported outcome measures (PROMs): Results on the study quality, overall psychometric robustness, and its quality of evidence

PROMs	Pooled sample size	Time interval	Study quality (Very Good, Adequate, Doubtful, Inadequate)	Results (Intraclass correlation)		Overall robustness (+: sufficient, - : insufficient)	Quality of evidence (High, moderate, low, very low) (Reason of downgrading quality)
Unidimensional							
FSS (Spanish)	Not assessed						
D-FIS (Spanish)	Not assessed						
FACIT-F	66	Within 6 months	Very Good[8]		0.81[8]	+	Moderate (Serious imprecision, n<100)
Multidimensional							
IBD-F (English)	36	Within 6 weeks	Very Good [9]	Section 1 Section 2	0.74[9] 0.83[9]	+ +	Low (Very serious imprecision, n<50)
IBD-F (Brazil)	123	Within 48-72 hours	Doubtful[10]	Section 1 Section 2 Total score	0.92[10] 0.97[10] 0.97 [10]	+ + +	Low (Very serious risk of bias from 1 doubtful study; risk of recall bias)
IBD-F (Greek)	61	Within 30-45 days	Adequate[11]	Section 1 Section 2	0.88[11] 0.90[11]	+ +	Low (Serious risk of bias from 1 adequate study, ICC model was not described; serious imprecision, n<100)
IBD-F (Danish)	66	Over 2 weeks	Adequate[12]	Section 1 Section 2	0.88[12] 0.94[12]	+ +	Low (Serious risk of bias from 1 adequate study, ICC model was not described; serious imprecision, n<100)
MFI-20	35	Within 6 weeks	Very Good[9]	General fatigue Physical fatigue Reduced activity Reduced motivation Mental fatigue	0.65[9] 0.77[9] 0.78[9] 0.81[9] 0.84[9]	- + + + +	Low (Very serious imprecision, n<50)
MAF	35	Within 6 weeks	Very Good[9]	Total score	0.74 [9]	+	Low (Very serious imprecision, n<50)
M-FIS (Spanish)	Not assessed						
FQ (Norwegian)	22	Over 6 months	Adequate[13]	Physical fatigue Mental Fatigue Total fatigue	0.98[13] 0.88[13] 0.98[13]	+ + +	Very low (Serious risk of bias from 1 adequate study; No evidence on whether the test conditions were similar; very serious imprecision, n<50)

Abbreviation: FSS - Fatigue Severity Scale, D-FIS - Daily Fatigue Impact Scale, FACIT-F - Functional Assessment of Chronic Illness Therapy – Fatigue, IBD-F - Inflammatory bowel disease fatigue scale, MFI-20 - Multidimensional Fatigue Inventory-20, MAF - Multidimensional Assessment of Fatigue; M-FIS - Modified Fatigue Impact Scale, FQ - Fatigue Questionnaire

Reference for the supplementary data content:

1. Terwee CB, Jansma EP, Riphagen II, et al. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual life Res*. 2009;18(8):1115-1123.
2. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.
3. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
4. de Vet HCW, Adèr HJ, Terwee CB, et al. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Qual life Res*. 2005;14(5):1203-1221,1223-1224.
5. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual life Res*. 2018;27(5):1159-1170.
6. Schünemann H, Brożek J, Guyatt G OA, et al. GRADE handbook for grading quality of evidence and strength of recommendations. www.guideline development.org/handbook. Published 2013. Accessed January 15, 2021.
7. Castillo-Cejas MD, Robles V, Borruel N, et al. Questionnaries for measuring fatigue and its impact on health perception in inflammatory bowel disease. *Rev Esp Enferm Dig*. 2013;105(3):144-153.
8. Tinsley A, Macklin EA, Korzenik JR, et al. Validation of the functional assessment of chronic illness therapy-fatigue (FACIT-F) in patients with inflammatory bowel disease. *Aliment Pharmacol Ther*. 2011;34(11-12):1328-1336.

9. Czuber-Dochan W, Norton C, Bassett P, et al. Development and psychometric testing of inflammatory bowel disease fatigue (IBD-F) patient self-assessment scale. *J Crohns Colitis*. 2014;8(11):1398-1406.

10. Lage AC, Oliveira CC, Batalha APDB, et al. The inflammatory bowel disease-fatigue patient self-assessment scale: Translation, cross-cultural adaptation and psychometric properties of the brazilian version (IBD-F Brazil). *Arq Gastroenterol*. 2020;57(1):50-63.

11. Varbobitis I, Kokkotis G, Gizis M, et al. The IBD-F Patient Self-Assessment Scale Accurately Depicts the Level of Fatigue and Predicts a Negative Effect on the Quality of Life of Patients With IBD in Clinical Remission. *Inflamm Bowel Dis*. 2020;27(6):826-835.

12. Vestergaard C, Dahlerup JF, Bager P. Validation of the Danish version of Inflammatory Bowel Disease Self-assessment Scale. *Dan Med J*. 2017;64(8).

13. Jelsness-Jørgensen LP, Bernklev T, Henriksen M, et al. Chronic fatigue is more prevalent in patients with inflammatory bowel disease than in healthy controls. *Inflamm Bowel Dis*. 2011;17(7):1564-1572.

Chapter 4

Discussion

Chapter 4 - Discussion

'Inability to perform and participate in daily activities' were found to be the most prevalent unmet needs among post-treatment CRC survivors (Den Bakker et al., 2018; Sodergren et al., 2019), and these deficits could last from 1- 10 years after diagnosis (Den Bakker et al., 2018; Schneider et al., 2007; Sodergren et al., 2019). Functional deficits are associated with poor quality of life (Schag et al., 1994) and survival (Braithwaite et al., 2010), and high healthcare burden (Chavan et al., 2020). However, the type and extent of functional deficits that post-treatment CRC survivors experience remains unclear. A better understanding of the extent and type of functional deficits will at least enable health care practitioners to inform CRC survivors if there are at risks of functional deficits and will hopefully enable health care practitioners to work with CRC survivors to prevent or minimize functional deficits. To provide health practitioners and CRC survivors with this information, we need data on whether functional deficits exist and the types of functional deficits that they have and methods for better exploring the potential causes of the deficits. My thesis work endeavoured to explore the extent and type of functional deficits and propose possible ways to explore the mechanisms underlying functional deficits.

My second thesis chapter demonstrated some functional deficits among individuals after GI cancer diagnosis, however this was in line with the normative data for functional deficits in those with the age of 75-85. Data on the functional status of GI cancer survivors can inform CRC survivors because CRC is the most prevalent type of GI cancer.

While the overall functional ability of those newly diagnosed with GI cancer did not demonstrate an immediate decline, there were a large number who stopped doing at least one functional task or had more difficulty performing the task. After the diagnosis of GI cancer, about 70% of participants were no longer performing at least one functional task that they were performing before diagnosis. The tasks that were no longer performed were primarily IADLs, such as finding your way around a new building (11 out of 18 participants), organising a trip/ social activities (8 out of 13 participants) and preparing a meal and/or doing laundry (9 out of 18 participants). A substantial proportion of participants (11 out of 25 participants) had more difficulty walking after a GI cancer diagnosis.

These findings were based on a small sample and are therefore hypothesis-generating. They imply that individuals are likely to experience some functional deficits after a diagnosis of GI cancer. These observations have 3 implications for CRC survivors, clinicians, and researchers. First, our findings could better inform CRC survivors about their potential long term (8 months post-diagnosis) functional deficits at earlier stages. It is key for CRC survivors to be aware of early functional decline and recognise the potential strategies to mitigate the functional decline, such as aerobic exercise (Hammer et al., 2015) and self-management programs (Hegel et al., 2011).

Second, our findings encourage clinicians to discuss the potential functional deficits that CRC survivors may experience and promote activity participation. Routine assessment of functional outcomes may be challenging. Early discussion of potential functional deficits might help CRC survivors proactively address the deficits as they occur.

Third, we explored the types of functional deficits that CRC survivors (CRC being the most prevalent type of GI cancer) potentially experience. Our exploratory findings help direct future research to focus on both functional performance and activity participation among CRC survivors. Our findings identified that GI cancer survivors experienced loss in participation of IADLs and increased difficulty in walking, however our data do not explain the cause behind the participation loss or increased difficulty with walking. Understanding the contributing factors behind the functional deficits will help researchers design more effective and specific interventions targeting the root causes of the functional limitation that CRC survivors have. Our data indicate the functional status about 8-12 months after GI cancer diagnosis. We do not know whether our participants underwent treatments and their treatment types received, however a course of chemotherapy can last for 6 months and it starts 1-2 months after surgery (depending on the stage, drugs, and treatment types) (*Treatments for Colon Cancer*, n.d.). Therefore our participants were likely at the end of their course of treatment or after treatments, and it is possible that function improves over time. Therefore it is important to examine whether our observed functional deficits among GI cancer survivors continue to understand the long-term impact of functional deficits.

Our findings demonstrate some functional deficits after GI cancer diagnosis. Therefore, understanding the underlying mechanism causing the functional deficits is the next important step to help us determine the most appropriate intervention. To do this, we may need to explore additional assessments that specifically target those mechanisms affecting function, which leads me to study the assessment of fatigue in Chapter 3.

Fatigue is the most common symptom experienced by CRC survivors (O'Gorman et al., 2018). Physical, emotional, and cognitive fatigue can directly impact on functional performance. It is likely that fatigue is therefore contributing, at least in part to the functional deficits identified in Chapter 2. To better understand how fatigue is involved, we need measures of fatigue that are sensitive and comprehensive to reflect the CRC-related fatigue and its associative factors. Although the pathogenesis of CRC-specific fatigue remains unknown, CRC-related fatigue can consist of physical, emotional, and cognitive aspects. It is likely to result from a combination of cancer treatment side effects, GI symptoms (e.g. pain, diarrhoea), the dysregulation of the gut-brain-axis due to microbiome imbalance, depression, and anxiety about cancer reoccurrence. As the causes of CRC-specific fatigue might differ from generic cancer-related fatigue, the experience and impact of fatigue may differ between CRC post-treatment survivors and the general cancer survivors population.

Common fatigue measures for those with cancer include the Functional Assessment of Cancer Therapy Instrument-Fatigue (FACIT-F), the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Fatigue subscale, and the Fatigue Questionnaire (Minton & Stone, 2009). However, these measures have only focused on physical fatigue and have not been validated solely in the CRC survivors population. Therefore, we need a fatigue measure that assesses all the important aspects of fatigue related to CRC. The causes of fatigue might be comparable between IBD and CRC, which include the chronic GI symptoms, depression and anxiety about the cancer reoccurrence or flare-up of IBD, and dysregulation of the gut-brain axis. Therefore I conducted a systematic review to identify fatigue measures

commonly used in IBD research, evaluated their content and psychometric properties, so I could recommend the most psychometrically robust and clinically feasible measure for the IBD and CRC populations .

Our systematic review identified 16 measures and recommended the FACIT-F and the IBD-Fatigue (English) (IBD-F) for both research and clinical use in patients with IBD. These measures were the ones that have the highest quality and most evidence on the psychometric properties and are clinically feasible. The FACIT-F is reliable, valid, responsive, and has data on the minimally clinically important change in the IBD population. Also, it is easy to administer, score and interpret, which can be easily used for practice. Yet the FACIT-F is a unidimensional measure assessing the impact of fatigue and is therefore less likely to capture change in fatigue experience. This may be important if outcome measurement needs to be brief in assessing the impact of fatigue on function.

The IBD-F (English) could be an alternative which provides a more in-depth understanding of CRC survivors' fatigue levels. The IBD-F (English) is developed in patients with IBD and intends to measure fatigue experience, the impact of fatigue, and perceived causes of fatigue in the IBD population. Although the measure developer intended to design this measure for IBD patients, the IBD-F (English) items are still applicable to CRC survivors. It is because the IBD-F (English) comprehensively assesses the impact of fatigue on self-care, physical, cognitive, emotional function, and some unique aspects such as sexual and interpersonal relationships, self-confidence, and quality of life. Although both FACIT-F and the IBD-F (English) primarily assess the impact of fatigue on daily activities, the IBD-F (English) measures the impact of

fatigue at a more granular level than the FACIT-F. For example, the IBD-F (English) measures how fatigue impacts work performance, cognition, walking, driving, washing and dressing, driving, and participation in the hobby and physical exercise that the FACIT-F does not assess. The IBD-F (English) might be more important if researchers expect changes particularly in these areas and it might help clinicians to understand which specific area is most affected by fatigue.

Also, the IBD-F (English) may be more clinically useful than the FACIT-F. It is because the first section of the IBD-F (English) can be used as a screening tool to identify individuals with fatigue for further assessment using the second section on the impact of fatigue. The screening section of the IBD-F (English) provides a more holistic and accurate level of fatigue severity than the commonly used 1-item question (i.e., "Do you feel unusually tired?") (Radbruch et al., 2008) because the IBD-F (English) captures more details on fatigue severity such as the highest, lowest, average and current fatigue level, and the proportion of time feeling fatigue.

Lastly, the third section of the IBD-F (English) would be particularly helpful in understanding the CRC-specific causes of fatigue because it asks about the perceived causes of fatigue and strategies that the individual had tried to manage fatigue. While the IBD-F(English) has shown to be reliable and valid in the IBD population, evidence on responsiveness is lacking. Therefore, the IBD-F (English) would be preferred over the FACIT-F if researchers are interested in measuring multidimensions of fatigue (i.e. causes, experience, and impact of fatigue), with the caveat that responsiveness of the measure has not been established.

Our findings also have implications for research. These recommendations were made on the

basis of very limited psychometric data on responsiveness and minimal clinically important change in the IBD population. While this should be a focus for future research to establish evidence, the FACIT-F and the IBD-F (English) are the most promising measures among the existing ones and they can be used for research and clinical practice. The FACIT-F and the IBD-F(English) have not been validated with CRC survivors and therefore consideration will have to be made in terms of the psychometric properties when interpreting results for those with CRC. It is key to validate the above potential measures in CRC survivors before using them.

Unique challenges in the process of conducting my research

There are very limited data on the functional outcomes and fatigue assessments among CRC survivors. Therefore, I have explored other populations that are similar in terms of pathologies, functional ability and fatigue experience to gain insight into how we better understand the functional needs of CRC survivors. In particular, the INTERBLEED study was not designed to understand the functional outcomes of CRC or GI cancer survivors. Therefore, our findings are not specific to CRC only. Yet our findings may indicate that CRC survivors experience some functional deficits. We had a small sample for our analysis, which shows that there may be functional deficits. However, these findings need replication with a much larger sample, for example, using data from the Canadian Longitudinal Study on Aging.

Another limitation is that we could not identify the most psychometrically robust and clinically feasible fatigue measure among CRC survivors because most generic cancer-related fatigue measures have not been validated solely on CRC population. The causes, experience and impact of fatigue may differ between CRC post-treatment survivors and the general cancer survivors.

Although the pathogenesis of CRC-specific fatigue remains unknown, the microbiome imbalance and the GI-related symptoms associated with CRC might influence fatigue in a similar fashion to those with IBD. Therefore, our findings on the psychometric data of fatigue measures in IBD can serve as a reference for the CRC population. However, it is important to validate our suggested measures in the CRC population before using them.

Future research

Future research is needed to confirm our preliminary findings on the functional abilities of CRC survivors and to understand their type and severity of functional deficits with respect to their treatment type and cancer site (colon or rectum). Future studies would ideally have an age- and gender-matched comparator group to understand whether the functional decline is due to CRC or ageing as a cofounding factor. It is also key to understand whether our observed functional deficits continue to worsen in the long term.

We have suggested the most promising fatigue measures for CRC survivors using IBD-related fatigue measures because of the close relationship to CRC-related fatigue, yet their psychometric properties require further validation in CRC survivors. Future research needs to establish evidence on responsiveness and minimal clinically important change, particularly important if the fatigue measure intends to measure change over time. Lastly, researchers need to understand the CRC-specific causes of fatigue to formulate an effective intervention to address fatigue.

Conclusion

I have attempted to understand the type and extent of functional deficits that CRC survivors experience and consider how to better assess the underlying mechanisms for functional deficits, likely to be at least in part caused by fatigue for CRC survivors. This novel information will help raise the awareness of the functional needs of CRC survivors and hopefully lead to an understanding of the underlying causes of the functional deficits, so that the functional abilities of CRC survivors can be at least maintained or potentially improved.

Reference

Braithwaite, D., Satariano, W. A., Sternfeld, B., Hiatt, R. A., Ganz, P. A., Kerlikowske, K., Moore, D. H., Slattery, M. L., Tammemagi, M., Castillo, A., Melisko, M., Esserman, L., Weltzien, E. K., & Caan, B. J. (2010). Long-term prognostic role of functional limitations among women with breast cancer. *Journal of the National Cancer Institute*, *102*(19), 1468–1477. https://doi.org/10.1093/jnci/djq344

Chavan, P. P., Kedia, S. K., & Yu, X. (2020). Impact of Physical and Functional Limitations on Health Care Utilization in Older Cancer Survivors: A Medicare Current Beneficiary Survey. *Journal of Aging and Health*, *32*(9), 987–997. https://doi.org/10.1177/0898264319872309

Den Bakker, C. M., Schaafsma, F. G., Huirne, J. A. F., Consten, E. C. J., Stockmann, H. B. A. C., Rodenburg, C. J., De Klerk, G. J., Bonjer, H. J., & Anema, J. R. (2018). Cancer survivors' needs during various treatment phases after multimodal treatment for colon cancer - Is there a role for eHealth? *BMC Cancer*, *18*(1), 1–12. https://doi.org/10.1186/s12885-018-5105-z

Hammer, M. J., Ercolano, E. A., Wright, F., Dickson, V. V., Chyun, D., & Melkus, G. D. (2015). Self-management for Adult Patients With Cancer: An Integrative Review. *Cancer Nursing*, *38*(2). https://journals.lww.com/cancernursingonline/Fulltext/2015/03000/Self_management_for_Adult_Patients_With_Cancer__An.12.aspx

Hegel, M. T., Lyons, K. D., Hull, J. G., Kaufman, P., Urquhart, L., Li, Z., & Ahles, T. A. (2011). Feasibility study of a randomized controlled trial of a telephone-delivered problem-solving-occupational therapy intervention to reduce participation restrictions in rural breast cancer survivors undergoing chemotherapy. *Psycho-Oncology*, *20*(10), 1092–1101.

https://doi.org/10.1002/pon.1830

Minton, O., & Stone, P. (2009). A systematic review of the scales used for the measurement of cancer-related fatigue (CRF). *Annals of Oncology*, *20*(1), 17–25. https://doi.org/10.1093/annonc/mdn537

O'Gorman, C., Stack, J., O'Ceilleachair, A., Denieffe, S., Gooney, M., McKnight, M., & Sharp, L. (2018). Colorectal cancer survivors: an investigation of symptom burden and influencing factors. *BMC Cancer*, *18*(1), 1022. https://doi.org/10.1186/s12885-018-4923-3

Radbruch, L., Strasser, F., Elsner, F., Gonçalves, J. F., Løge, J., Kaasa, S., Nauck, F., & Stone, P. (2008). Fatigue in palliative care patients — an EAPC approach. *Palliative Medicine*, *22*(1), 13–32. https://doi.org/10.1177/0269216307085183

Schag, C. A. C., Ganz, P. A., Wing, D. S., Sim, M.-S., & Lee, J. J. (1994). Quality of life in adult survivors of lung, colon and prostate cancer. *Quality of Life Research*, *3*(2), 127–141. https://doi.org/10.1007/BF00435256

Schneider, E. C., Malin, J. L., Kahn, K. L., Ko, C. Y., Adams, J., & Epstein, A. M. (2007). Surviving colorectal cancer. *Cancer*, *110*(9), 2075–2082. https://doi.org/https://doi.org/10.1002/cncr.23021

Sodergren, S. C., Wheelwright, S. J., Permyakova, N. V., Patel, M., Calman, L., Smith, P. W. F., Din, A., Richardson, A., Fenlon, D., Winter, J., Corner, J., & Foster, C. (2019). Supportive care needs of patients following treatment for colorectal cancer: risk factors for unmet needs and the association between unmet needs and health-related quality of life—results from the ColoREctal Wellbeing (CREW) study. *Journal of Cancer Survivorship*. https://doi.org/10.1007/s11764-019-00805-6

Treatments for colon cancer. (n.d.). Canadian Cancer Society. Retrieved September 5, 2021,

from https://cancer.ca/en/cancer-information/cancer-types/colorectal/treatment/colon-cancer