Doctoral theses at NTNU, 2020:352

Hilde Krogstad

Symptoms - prevalence and electronic assessment

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

Hilde Krogstad

Symptoms - prevalence and electronic assessment

Thesis for the Degree of Philosophiae Doctor

Trondheim, November 2020

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

ISBN 978-82-326-5050-7 (printed ver.)
ISBN 978-82-326-5051-4 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:352

Printed by NTNU Grafisk senter

Symptomer – forekomst og elektronisk kartlegging

Symptomer er definert som en subjektiv opplevelse av sykdomstegn og rapporteres av pasienten muntlig eller ved bruk av ulike skjema. En viktig del av pasientsentrert behandling er å spørre pasientene systematisk om deres symptomer og funksjonsnivå. Dagens pasientjournaler er i hovedsak elektroniske, og således ligger forholdene til rette for elektronisk innhenting av pasientrapporterte data. I denne avhandlingen har jeg vært en del av en gruppe som har utviklet et elektronisk kartleggingsverktøy (Eir). Eir er utviklet via tverrfaglige og trinnvise prosesser hvor gjentatte runder med brukertesting har ført til kontinuerlige forbedringer. Vi har undersøkt om pasientene foretrekker å rapportere symptomer på papir eller nettbrett og testet brukervennligheten av Eir i ulike settinger. Resultatene har vist at de fleste kreftpasientene foretrekker å bruke et elektronisk skjema, og at Eir er relevant og brukervennlig for kreftpasienter og leger i ulike settinger. Eir er designet for bruk i både klinisk praksis og forskning. Leger og forskere etterspør ofte referansedata for å kunne sammenligne resultater mellom pasienter og hva den generelle befolkningen rapporterer. For eksempel, hvis pasientenes symptomskåre er høyere eller lavere enn i den generelle befolkningen etter å ha justert for alder, kjønn og andre relevante bakgrunnsvariabler, kan det være indikasjon for oppfølging av potensiell sykdom eller bivirkninger av behandling. Vi har presentert de første norske referanseverdiene for M.D. Anderson Symptom Inventory, et mye brukt skjema blant kreftpasienter, fra den generelle norske voksne befolkningen. Samlet sett kan disse studiene kan bidra til å forbedre klinisk praksis ved å legge til rette for økt bruk av pasientrapporterte data.

Name of candidate:	Hilde Krogstad
Department:	Department of Clinical and Molecular Medicine Cancer Clinic, St. Olavs hospital, Trondheim University Hospital
Main supervisor:	Marianne Jensen Hjermstad
Co-supervisors:	Jon Håvard Loge Kari Sand Stein Kaasa
Public defence:	*12. November 2020*

Table of contents

Acknowledgements

The presented work was conducted at the Department of Clinical and Molecular Medicine, Norwegian University and Science of Technology (NTNU) and at the Cancer Clinic, St.Olavs hospital, Trondheim University Hospital. The work was funded by the NTNU and the Cancer Clinic.

I am grateful to all the participants in this project for their important contributions.

I want to thank my supervisors. First, my main supervisor, Marianne Jensen Hjermstad, who has always been available for guidance and rapid feedback, which has been crucial for the progress of the project. I also want to thank my co-supervisors; Kari Sand for always being positive and supportive, Jon Håvard Loge who has provided highly appreciated feedback on important issues in this project, and Stein Kaasa for offering me the opportunity to embark on this project. With your broad knowledge in research you all have contributed crucially to the scope and quality of this project.

To all my colleagues both at the Cancer Clinic and at the Research Group for Cancer and Palliative Care - thank you for friendship and support. You have encouraged and inspired me for years working together. Thanks to my colleagues at the Radiation Therapy Unit; Monika, Hanne, Mirjam, Martine, Boris and Tora for your support during the final work with this thesis. Furthermore, I want to thank the head of the Cancer Clinic, Arne Solberg, for facilitating the opportunity to combine clinical work and research activity. I am grateful to Ragnhild Green Helgås for always being supportive and answering all administrative questions throughout the PhD project, and to Gunnhild Jacobsen who has been my "neighbor" at the office during the entire PhD period.

Thanks to all co-authors of the three papers which constitutes the basis of this thesis. A special thanks to the Eir project group; Sunil Raj, Erik Løhre, Eivind Andersen and Tarje Halvorsen, for engagement and discussions. I would also like to thank Stine Marie Sundt-Hansen and Liv Ågot Hågensen for important contributions in paper III, as well as the Coordination Unit Orkdal region for their project management contribution.

I want to thank my family and friends for your understanding and encouragement. Thanks to my parents for unconditional support through all years. At last, I want to thank the four most

important persons in my life; my dear Håvard and our wonderful children Hedda, Nora and Ane.

Trondheim, June 2020

Hilde Krogstad

Abbreviations

BPI	Brief Pain Inventory
CI	Confidence Interval
COMBAT	Computer Based Assessment and Treatment
CTCAE	Common Toxicity Criteria for Adverse Events
EAPC	European Association for Palliative Care
ECOG	Eastern Cooperative Oncology Group
EHR	Electronic Health Record
EORTC QLQ-C30	European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Core Questionnaire 30
E-PROMs	Electronic Patient Reported Outcome Measures
ESAS	Edmonton Symptom Assessment Scale
FACT-G	Functional Assessment of Cancer-General Version
FDA	Food and Drug Administration
FQ	Fatigue Questionnaire
GAD 2	General Anxiety Disorder 2
GP	General Practitioner
HADS	Hospital Anxiety and Depression Scale
HCP	Health Care Professional
HRQoL	Health-Related Quality of Life
HUNT	Health Study of Nord-Trøndelag
ICC	Intraclass Correlation Coefficient
MID	Minimum Important Difference
MDASI	M.D. Anderson Symptom Inventory

NRS	Numerical Rating Scale
NTNU	Norwegian University of Science and Technology
PG-SGA	Patient-Generated Global Assessment
PHQ-9	Patient Health Questionnaire-9
PROMIS	Patient-Reported Outcomes Measurement Information System
PROMs	Patient-Reported Outcome Measures
PROs	Patient-Reported Outcomes
QoL	Quality of Life
RCT	Randomized Controlled Trial
REC	Regional Committee for Medical and Health Research Ethics
SCQ	Self-Administered Comorbidity Questionnaire
SD	Standard Deviation
SF-36	Short-Form Health Survey 36
SISAQoL	Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints
SUS	System Usability Scale
VAS	Visual Analog Scale
VDS	Verbal Descriptor Scale
WHO	World Health Organization

Norsk sammendrag

Symptomer er definert som en subjektiv opplevelse av sykdomstegn som bare kan oppleves og iakttas av den syke selv. En rekke forskjellige symptomer er vanlige i en generell befolkning, utgjør en hyppig årsak for kontakt med lege, men er ikke alltid et tegn på sykdom. Forekomst av symptomer er assosiert med faktorer som alder, kjønn, yrkessituasjon, kroniske sykdommer og psykiatriske tilstander. De hyppigst rapporterte symptomene blant kreftpasienter er fatigue, smerte, redusert matlyst, kvalme, tungpust og nedstemthet. Studier har vist at behandling av symptomer hos kreftpasienter ikke alltid er tilstrekkelig og at symptomer blir underrapportert i kliniske konsultasjoner. Utvikling av kartleggingsverktøy er et viktig tiltak for å forbedre diagnostikk og behandling av symptomer.

M.D. Anderson Symptom Inventory (MDASI) er et mye brukt skjema blant kreftpasienter. MDASI er et kortfattet skjema hvor pasientene kan rapportere hvilke symptomer de har og hvordan disse symptomene påvirker daglig funksjonsnivå. Alle symptomene i MDASI er imidlertid vanlige også i den generelle befolkningen, og hvordan symptomer påvirker funksjon er relevant i alle populasjoner. Referanseverdier fra den generelle befolkningen kan blant annet brukes for å sammenligne resultater mellom pasienter og hva den generelle befolkningen rapporterer. Dette kan være viktig for forskning og klinisk praksis. Vi utførte en tverrsnittsstudie hvor vi samlet inn referanseverdier for MDASI ved å undersøke forekomst og intensitet av symptomer i et utvalg fra den norske voksne befolkningen. De vanligste symptomene var fatigue, døsighet og smerte. Antall komorbide tilstander, høyere nivå av depresjonssymptomer og lavere utdanningsnivå samsvarte med høy opplevd symptombyrde. Ved bruk av referanseverdiene må det kontrolleres for disse faktorene.

Symptomkartlegging er tradisjonelt utført på papir, men dagens utvikling av helseinformasjonsteknologi gjør det mulig å presentere og integrere pasientdata fra ulike kilder, inkludert måling av symptomer. Vi har utviklet Eir, et elektronisk verktøy for kartlegging av symptomer hos kreftpasienter, via tverrfaglig og trinnvis utvikling. Gjentatte runder med testing med pasienter og leger etterfulgt av kontinuerlige endrings- og forbedringsprosesser har ført til at Eir versjon 3 (V3) har blitt et brukervennlig kartleggingsverktøy. Brukertesting av EirV3 i ulike situasjoner indikerer at verktøyet er anvendelig og godt akseptert i en heterogen populasjon av kreftpasienter. En komparativ studie som målte likheten mellom utfylling av symptomskåre på papirskjema sammenlignet

med elektronisk skjema viste utmerket ekvivalens mellom disse metodene. De fleste kreftpasientene foretrakk å bruke et elektronisk skjema (41%) eller hadde ingen preferanser (40%) angående metode. EirV3 er fortsatt i utvikling og brukes nå i en norsk klyngerandomisert studie der systematisk symptomkartlegging er en del av intervensjonen i studien. Videre arbeid bør fokusere på å integrere EirV3 i daglig klinisk praksis og i den elektroniske pasientjournalen, samt integrere referanseverdier for å forenkle fortolkningen av pasientenes egenrapporterte symptomskårer.

English summary

A symptom is defined as a feeling of disease or physical disturbance observed by the patient. Symptoms are common in the general population but do not necessarily indicate a disease or a disorder. The prevalence of symptoms is associated with factors such as age, gender, employment status, chronic conditions, psychiatric disorders etc. In patients with cancer, the most commonly reported symptoms are fatigue, pain, loss of appetite, nausea, dyspnea and depressive symptoms. The symptom prevalence in cancer patients is high despite medical advances and increased interest in and efforts regarding different methods of assessing symptoms. Moreover, the development of several symptom assessment tools does not seem to overcome the single most important barrier to optimal symptom management, namely inadequate symptom assessment.

A widely used symptom assessment tool is the M.D. Anderson Symptom Inventory (MDASI). The MDASI is a brief, reliable and valid tool for self-report of commonly experienced symptoms and how these symptoms interfere with daily functioning. The MDASI was originally designed for use in cancer patients. However, all symptoms included in the MDASI are also commonly experienced by the general population, and how self-reported severity of symptoms interferes with daily living is an important issue in all populations. Reference values are important to evaluate whether patients' symptom scores are above or below the mean values from the general population of the same age, gender or adjusted for other relevant background variables. As such, reference values facilitate the interpretation of scores for use in clinics and research settings. In a cross-sectional study, we collected reference values for the MDASI by examining the presence and intensity of common symptoms in a sample of the Norwegian adult population. The most frequent symptoms were fatigue, drowsiness and pain. Fatigue had the highest mean score. The presence of one or more comorbidities, increasing scores on depressive symptoms and lower level of education were associated with higher MDASI sum score.

Today's advances in health information technology permit immediate presentation and integration of patient data from various sources, including measurement of symptom scores. We have developed Eir, an electronic symptom assessment tool for use in cancer care, through multiprofessional, stepwise, and iterative processes. Iterative test rounds with end-users followed by continuous improvements led to a user-friendly symptom assessment tool, Eir Version 3 (V3). Usability testing of EirV3 in different settings indicates that the tool is

applicable and well accepted in a heterogeneous population of cancer patients. A comparative study examining equivalence between electronic and paper-based scores showed excellent agreement across methods. A majority of the patients preferred electronic assessment (41%) or had no preference (40%) regarding administration method. EirV3 is still in development and is currently implemented into the patient care pathways and clinical practice in a Norwegian cluster randomized trial on early integration of palliative care in oncology. Further work should address how to integrate EirV3 into daily clinical practice, in the electronic patient records and how to incorporate reference values to facilitate interpretation of patient self-reported scores.

List of papers

I Krogstad H, Loge JH, Grotmol K, Kaasa S, Kiserud C, Salvesen Ø, Hjermstad MJ.

Symptoms in the general Norwegian adult population - prevalence and associated factors

BMC Public Health. 2020 Jun 23;20(1):988. doi: 10.1186/s12889-020-09109-2.PMID: 32576168

II Krogstad H, Brunelli C, Sand K, Andersen E, Garresori H, Halvorsen T, Haukland EC, Jordal F, Kaasa S, Loge JH, Løhre ET, Raj SX, Hjermstad MJ.

Development of EirV3: A Computer-Based Tool for Patient-Reported Outcome Measures in Cancer

JCO Clin Cancer Inform. 2017 Nov; 1:1-14 doi: 10.1200/CCI.17.00051

III Krogstad H, Sundt-Hansen SM, Hjermstad MJ, Hågensen LÅ, Kaasa S, Loge JH, Raj SX, Steinsbekk A, Sand K.

Usability testing of EirV3: a computer-based tool for patient-reported outcome measures in cancer

Support Care Cancer. 2019 May; 27(5):1835-1844 doi: 10.1007/s00520-018-4435-3

1. Introduction

Cancer is one of the leading causes of death globally [1]. Worldwide, 18.1 million new cases and 9.6 million deaths from cancer were estimated in 2018 [2]. Mortality rates for most types of cancer decreased in the decade 2007─2017 [1]. In 2018, 34,190 new cancer cases were reported in Norway [3]. The most common cancer sites are prostate, breast, colon and lung [3, 4]. The CONCORD-3 study showed that survival rates for most cancers are generally increasing worldwide, and in Norway the 5-year net survival remains among the highest in the world [5]. By the end of 2018 a total of 283,894 Norwegians were alive after having had at least one cancer diagnosis [3].

Anticancer treatments include surgery and radiotherapy, and systemic treatments with chemotherapy that also includes hormonal, biological and targeted therapies. Systemic treatment might be given before or after surgery and/or radiotherapy as part of multimodal, curative treatment. Targeted therapies including immunotherapy are used to a much greater extent than before, which have led to more treatment options for several cancer diagnoses. The new treatment agents have side effects that differ from those of traditional chemotherapy, and attention must also be directed at new side effects and symptoms [6]. Many patients will live with incurable cancer for many years, with large variations in the need for palliative care [7-10]. According to the WHO definition, palliative care focuses on patients with a life-threatening disease [6]. Nevertheless, the focus in palliative care is the patient living with the disease or with the side-effects after the treatment, or both. As such, palliative care is applicable in all phases of the disease trajectory, also from early on, irrespective of treatment intention [6, 11]. A key element within palliative care is to improve or maintain best Quality of Life (QoL) for patients by early identification and treatment of symptoms. Optimal symptom assessment and management, and acknowledging the patients' perspectives on health, i.e. patient-centered care is essential to achieve this.

A **symptom** is defined as a feeling of disease or physical disturbance observed by the patient [12]. Symptoms are subjective by nature, thus best recognized by the individual experiencing them. In contrast, a "sign" is a finding identified by health care personnel through different methods such as clinical observations, biomarkers, imaging etc. [12]. Symptoms may be evidence of a disease or a disorder, but they also reflect the normal variation in the physical or psychological states as experienced by most individuals. Symptoms are common in the general population and are found to be associated with factors such as chronic conditions, gender, age, employment status, living situation, psychiatric disorders and functional status [13-15]. In

patients with cancer, symptoms are related to the disease itself or side effects of treatment. Valid reference data are particularly relevant in studies on cancer survivorship which may go beyond decades post-treatment, as common age-related conditions and life events may influence which symptoms the cancer survivors experience.

Systematic assessment of the patients´ self-evaluation on health and disease have been labelled differently in the last decades. Quality of Life and Health-Related Quality of Life were for long the most frequently used terms. In 2006, the U.S. Food and Drug Administration (FDA) proposed the term **Patient Reported Outcome Measures (PROMs)** for all measures that can best or only be assessed by asking the patients themselves [16]. By that, the FDA also formally recognized the importance and clinical utility of PROMs by releasing a new Guidance for Industry on this issue [16, 17]. Today the abbreviations PROMs and PROs are frequently used as interchangeable abbreviations in the literature, but this is actually not correct. Whereas PROMs denote the tools or measures that are used to elicit patients´ perceptions and self-report, be it on paper or electronically, PROs are the actual outcomes (patient-reported outcomes) that the patients report, e.g. pain, problems concentrating and QoL. The recognition of patients´ perspectives as valid outcomes in clinical medicine has been endorsed by the National Institute of Health consensus conference [18]. This shift has occurred in cancer care as well, and the term PROMs encompasses all instruments covering the patient´s self-reported perspective on their physical and psychological well-being, level of functioning, symptom intensity and symptom impact or severity as well as perceptions related to treatment effects and side effects [19, 20]. As the patient is the primary source of information, PROMs supplement clinical observations and objective findings with individual patient information and play an essential part in systematic follow-up in cancer care. Moreover, PROs are emphasized as primary or secondary endpoints in oncology trials, and may facilitate and improve the integration of patients' perspectives into clinical research [21, 22].

Symptom assessment tools provide the basis for detecting symptoms, grading their severity, and assessing the effectiveness of treatment [23]. Traditional symptom assessment tools were developed for the paper and pencil format, in the form of checklists, semi-structured interviews, or specific or generic tools of different lengths, formats and intents. Here, assessment by interviews will not be further described, even if the medical interview per se is strongly focused on symptom assessment. The advantage of paper-based tools is that most persons are familiar with use of paper and pencil. Nevertheless, paper tools may be cumbersome to use if they are lengthy, or if the questions are perceived as irrelevant or

repetitive [24]. A paper tool by itself is only a piece of paper. It will only be meaningful if it is used and interpreted as part of a clearly defined process. New advances in health information technology have promoted the development of electronic symptom assessment tools (e-PROMs) for use on different platforms, e.g. cell phones, computers or tablets. Such new tools allow an effective transfer and integration of patient related data from various sources, e.g. patients' self-report, physicians' objective reports of signs of disease and data from the electronic patient records over time. The use of electronic devices for symptom assessment offers several advantages compared with the traditional paper-based method with respect to data collection and storage, fewer missing data during the data entry process, by making certain items mandatory before proceeding, immediate presentation of scores, transfer of questionnaires, data completed outside the hospital setting and automatic reminders to patients and triggers for doctors if scores are high or worse compared with previous assessments [25-27]. Furthermore, if the tool is used repeatedly health care providers can easily review the development of symptoms over time. Electronic assessment may be targeted to the individual patient by tailoring the questions based on the patient's diagnosis, the specific treatment or the patient's previous responses, thereby reducing patient burden by skipping irrelevant questions.

Researchers at the Norwegian University of Science and Technology (NTNU) and St. Olavs hospital, Trondheim University Hospital have validated and used PROMs in clinical research and in cancer care for at least two decades, and initiated development of the electronic solution named **Eir**. Eir is an electronic PROM and may be completed on tablets or computers by the patients at home, in the hospital, nursing home or any other place where the patients are. When the patient has scored his/her symptoms in Eir, the data are immediately transferred to a database that can be accessed by health care providers (HCPs) on any computer, also displaying the patient's symptom development over time. By using the results available in Eir, HCPs could pay specific attention to areas that are particularly bothersome or have deteriorated over time, facilitate communication and changes in medication etc. based on needs, and initiate a more frequent schedule of symptom assessments if perceived necessary. The overall objective of Eir is to improve symptom management by introducing a systematic, standardized way of assessing PROs, thereby making patients´ self-report of symptoms, problems and level of functioning immediately available in clinical consultations.

2. Background

2.1 Patient-centered care

Patient-centered care is defined as "care that is respectful of, and responsive to, individual patient preferences, needs and values, and ensuring that patient values guide all clinical decisions" [28]. A patient-centered approach is recommended in medicine in general, not only in cancer care [6, 29, 30]. The main message is to focus on the needs, support and treatment that is perceived as most important by the patient [31, 32]. This makes it easier for people to make informed decisions about their own health and health care. Self-reporting of symptoms also engages patients as active participants [33]. Patient-centered care is improving the quality of care by promoting appropriate use of services [34].

Assessment of symptoms is a prerequisite for gaining insight into how patients perceive their own physical and psychological health. Studies have demonstrated potential benefits of routine symptom assessments in clinical practice, such as improved patient-physician communication [35], increased awareness of patients´ physical and psychosocial functioning [36], improved patient wellbeing [33], a more efficient and focused use of time [37], and even increased survival [38-40]. Systematic use of PROMs may also make the health care professionals (HCP) aware of symptoms that they did not know bothered their patients [41]. Regular symptom assessment during treatment with routine-based rapid feedback of results to clinicians has been reported as an efficient way of informing clinicians and patients about treatment effects and potential side effects, thereby guiding treatment decisions and follow-up [42, 43]. This in turn should result in a common agreement on what is the best treatment approach – a key feature of patient-centered care [6, 23].

PROMs can assess a complex construct like Health-Related Quality of Life (HRQoL) as well as a more focused, one-dimensional construct such as symptom impact or severity [44]. HRQoL is defined as a multidimensional concept that represents the patient´s perception of physical, psychological, social aspects, overall health and quality of life [17]. The use of PROMs in this respect has often been in the form of specifically developed tools for e.g. a given cancer type or treatment, to provide detailed clinical information about specific symptoms, disease or treatment-related effects on functioning and HRQoL in subgroups of patients. Hence, this knowledge may prompt relevant and efficient interventions.

2.1.1 Development of HRQoL instruments

There are several types of questionnaires for measuring HRQoL: generic, disease-specific and domain-specific. Generic measures are developed to collect data regardless of clinical diagnosis or specific population characteristics. Generic measures make comparisons across populations and conditions possible, and they are applicable for patients with more than one condition. The first generic instruments were launched in the 1970s and 1980s and were generally lengthy and time-consuming to fill in [45]. The Short-Form Health Survey 36 (SF-36) [46] is a second generation generic instrument, i.e. is not specific for any population or disease. The SF-36 assesses HRQoL by eight different scales covering aspects of mental and physical health and social functioning [46]. Reference values for the SF-36 have been collected and published in many countries [47-49]. The World Health Organization has developed the WHO Quality of Life Assessment Instrument (WHOQOL) which was developed simultaneously in several languages at 15 international centers [50]. This instrument encompasses five domains: physical health, psychological health, level of independence, social relationship, spirituality and environment. An abbreviated scale containing 26 items and primarily for use in epidemiological studies has also been published [51].

Disease-specific measures are developed to collect data, brought about by specific disease entities, e.g. rheumatoid arthritis, cancer, diabetes or other chronic diseases, and can be used to describe pre- and post-treatment health status [52]. A frequently used cancer-specific questionnaire, the EORTC QLQ-C30 [53] was finalized in 1993. The questionnaire covers physical, role, cognitive, emotional and social function, overall quality of life, three symptom scales and six single items. The single items assess common cancer symptoms such as dyspnea, loss of appetite, insomnia, constipation and diarrhea. However, these symptoms are common in the general population as well, and has led to the collection of reference values in multiple languages [54, 55]. This 30-item core questionnaire can be supplemented with additional questionnaires designed for specific cancer sites or patient subgroups. The Functional Assessment of Cancer-General Version (the FACT-G) was first published in 1993 [56]. The FACT-G includes 27 items covering physical well-being, social well-being, emotional well-being and functional well-being [56]. Reference values for the values FACT-G have been published for a sample of the general population and a sample of adult cancer patients [57].

In addition to generic and disease specific instruments, so-called domain-specific instruments have been designed to assess specific symptoms, such as pain, fatigue and anxiety [45]. For example, the Hospital Anxiety and Depression Scale (HADS) was a first-generation instrument constructed in 1983 [58]. HADS is one of the most frequently used instruments for measuring symptoms of anxiety and depression in oncology. When measuring anxiety and depression, it is important to examine whether the instrument includes somatic items. Symptoms such as fatigue, weight loss, loss of appetite etc. are strongly affected by underlying somatic disease. The Patient Health Questionnaire 9 (PHQ-9) was introduced in 1999 [59]. The PHQ-9 assesses all the nine diagnostic criteria for a major depressive disorder including duration and functional consequences. The Fatigue Questionnaire (FQ) measures physical and mental fatigue [60], and is commonly used in studies of cancer-related fatigue [61, 62]. The Brief Pain Inventory (BPI) [63] measures the impact of pain on physical functioning in addition to measuring pain intensity.

2.2 Symptoms

2.2.1 Symptoms in the general adult population

Symptoms are common in the general population [13, 14, 64] but do not necessarily indicate a disease or a disorder. A large Danish nationwide cohort study with 49,706 respondents and a response rate of 52% demonstrated that symptoms were common in the general population. A total of 44 symptoms were assessed using a web-based questionnaire, covering a wide area of clinically relevant symptoms and frequently occurring symptoms which are often presented to the general practitioner (GP). Subjects were asked if they had experienced any of the symptoms in the preceding four weeks. Prevalence estimates of self-reported symptoms varied from 49.4% reporting tiredness to 3.4% reporting difficulty swallowing. About 9 out of 10 respondents reported at least one symptom within the preceding four weeks. In total, 37% contacted the GP with at least one symptom [64]. Studies from general populations have shown associations between a high number of somatic symptoms and impaired health status [14, 65].

A Norwegian community-based study from year 2004 included 3325 subjects in a postal survey, yielding a response rate of 54.4% [15]. The questionnaire included 23 different symptoms, health, demographic and lifestyle factors. At least one symptom was reported by 91.9% of the participants, 47% reported six or more, and 17% reported 10 or more symptoms. Symptom reporting was more frequent among women, those reporting poor

health, unemployment, low education or obesity [15]. Self-reported overall health explained 28.2% of the variance in the number of symptoms [15].

From 1995 to 1997, the Health Study of Nord-Trøndelag County 2 (HUNT2) invited all inhabitants aged 20 years and above in this region to have their health examined. The inhabitants received a postal questionnaire asking about physical symptoms, demographic factors, lifestyle, and somatic diseases. Among those invited, 62,651 participants completed the questionnaire, yielding a response rate of 71.3% [66]. The questionnaire included 22 symptoms, formatted like: "Have you been bothered by pain or discomfort during the last year from the back, neck, stomach etc.?" Anxiety and depression were assessed by HADS. Results showed that the prevalence of all somatic symptoms except breathlessness and functional impairment were significantly higher in women than in men ($p<0.05$) [66]. The most prevalent symptom in females were headache (40%), pain in the shoulders (29%) and pain in the neck (27%). For men, the most prevalent symptoms were heartburn (31%), headache (24%) and low back pain (20%). The authors reported a statistically significant relationship between anxiety, depression and functional somatic symptoms, independent of age and gender [66].

The HUNT3 study in 2006–2008 invited a random sample of 6,419 participants to report pain every three months over a 12-month period with 4,782 (75%) accepting the invitation [67]. Five questionnaires were sent and included the one-week recall version of the SF-8 health survey. This is a shortened version of the SF-36 with one item representing each of the following eight subscales: bodily pain, general health, mental health, vitality, physical functioning, social functioning and work limitations [68]. The response categories range from e.g. no pain to very severe pain. Results showed that the total one-year prevalence of chronic pain, defined as reporting of moderate to severe pain in at least three of five measurements, was 31%. Estimates were 36% among women and 25% among men. The prevalence was higher in women (36%) than in men (25%), was higher among people with high body mass index, and in people with low income or low educational level and increased with age.

2.2.2 Use of population data as reference values

In medicine, reference values are needed for interpretation of results. When using PROMs in clinical follow-up studies, clinicians or researchers often request reference data to facilitate the interpretation of results [69]. Reference values make it possible to perform comparisons among groups of patients and should be adjusted for age, sex, education, other sociodemographic variables and comorbidities, as these affect the subjects´ self-report of

symptoms and functioning levels [70, 71]. The relevance of valid reference data is particularly well illustrated in follow-up studies on cancer survivorship, which often go beyond decades post-treatment. Common age-related conditions, somatic or psychological problems, lifestyle patterns and life events may influence how people perceive their health and quality of life, and how chronic diseases affect the general population independent of previous diagnoses and treatment. This in turn influences how people in general as well as cancer survivors experience and cope with symptoms and changes in level of functioning. Hence, poor PROM scores in different groups of cancer survivors should not automatically be interpreted as late effects from the disease or treatment. It is only by comparing with data from the general population that we can ascertain if these groups are at excess risk for specific somatic or psychological symptoms, or if the prevalence of these matches findings in the general population when adjusted for age, gender and common and well-known risk factors.

Clinical research communities are increasingly interested in the broader interpretations of PROMs scores for purposes of comparison across studies and populations and to allow for contextual interpretation of disease impact. This is particularly relevant for cancer research because there is a wide range of symptoms and functional deficits by cancer type, stage, and treatment. Reference values have been collected for several commonly used PROMs in different countries , e.g. the FACT-G [57], the SF-36 and the EORTC QLQ-C30 [70, 72]. Norwegian reference values for the SF-36 were collected in 2015 and compared with similar surveys in 1996 and 2002 [47]. The stability of scores on all HRQoL domains across the three surveys was high, indicating a relatively stable HRQoL in the Norwegian population during this period despite the fact that there has been a substantial decline in the response rates in surveys like this in the last decades [47]. Several examinations of reference values of the EORTC QLQ-C30 have been performed in Norway [55] and other countries [69, 72-74], and comparisons of samples have shown differences in mean scores in these countries, also among the European studies. In 2012, Hinz et al. [70] collected reference values for the EORTC QLQ-C30 based on a representative sample of the German adult population. They found that QoL decreased with age, but that there were only small gender differences. The mean scores were compared with age and gender adjusted scores of previous normative studies from Sweden, the Netherlands, Norway and Germany. The data of these studies were combined to arrive at common European normative values for the scales and the symptom items of the questionnaire.

For persons with cancer, where deficits relative to the general population are expected especially during treatment or rehabilitation, additional context regarding scores may be necessary to identify meaningful differences within and across cancer populations. Disease-specific reference values for PROMs may be used for these purposes and to evaluate the relative burden of one disease compared with other diseases [75], but up until now, few such studies have collected these types of reference values. Patient-Reported Outcomes Measurement Information System (PROMIS) is a PRO measurement system that can be used across chronic diseases and in the general population [76]. US-specific PROMIS cancer reference values were collected using a large, population-based cohort of persons recently diagnosed with cancer [75]. Data showed that cancer patients reported increased pain and fatigue and reduced physical function after end of treatment when compared to the general population. These findings suggest that meaningful, distinct symptom trends exist for patients with cancer by cancer stage, age at diagnosis and cancer type and support the necessity of reference values tailored to specific clinical information to ensure relevant interpretation across research and clinical settings [75].

2.2.3 Symptoms in adult cancer patients

In patients with cancer, symptoms of varying intensity and severity are prevalent, depending on the cancer diagnosis, stage, treatment, age, gender and other factors such as comorbid conditions and mental health state before, during and after disease and treatment [77]. The health problems of cancer survivors may be related to serious illness such as secondary malignancies or organ pathology like anthracycline-related cardiomyopathy but may also be subjective such as pain or fatigue. A nation-wide cross-sectional study examined the prevalence of chronic fatigue and associated factors among Norwegian long-term survivors of cancers in young adulthood (N=1088) [78]. Chronic fatigue was assessed by the Fatigue Questionnaire (FQ) [60]. Results showed that 25% of the young adult cancer survivors reported chronic fatigue at a median of 14 years from diagnosis [78]. Systemic treatment combined with surgery and/or radiotherapy, comorbidity, pain, numbness in hands, feet, and depressive symptoms were associated with chronic fatigue [78]. Similar findings have been reported from a study assessing fatigue among lymphoma survivors [79].

Patients with advanced cancer are often polysymptomatic, with fatigue, pain, loss of physical function and appetite, nausea/vomiting, dyspnea and depression being among the most distressing [80-82]. Pain is the second most prevalent symptom in palliative care. Pain intensity

is most often the main target for assessment of pain, but additional aspects such as variation of pain over time, pain triggered by physical activity or breakthrough pain might also be relevant in order to optimize symptom management [45]. As such, instruments must be evaluated for their properties in measuring these aspects of pain. Another challenge is how high a symptom score must be to be defined as a symptom, and further to prompt an intervention. The reported prevalence of symptoms varies due to the huge heterogeneity regarding patient samples, study designs and assessment methods. A review from 2007 covering 40 years of published articles, investigated the self-reported prevalence of pain in cancer patients. When pain severity was reported on visual analog scales (VAS) or numerical rating scales (NRS), the scores were converted into none (0), mild (1–4), moderate (5–6) or severe (≥7). The reported prevalence of pain was 64% in patients with advanced cancer, 59% in patients on anticancer treatment and 33% in patients who had been cured of cancer [83]. More than 33% of the patients graded their pain as moderate or severe [83]. An updated systematic review and meta-analysis reported pain prevalence rates of 39% after curative treatment, 55% during anticancer treatment and 66% in patients with advanced cancer [84], corresponding to another systematic review documenting that under-treatment of cancer pain is still an issue, probably affecting 30-40% of patients with advanced cancer [85]. These high prevalence rates exist despite great medical, pharmacological and technological advances, supplemented by the increased interest in pain assessment methods. However, the development of several assessment tools through the last two decades does not seem to overcome the single most important barrier to optimal pain management; inadequate and unsystematic pain assessments [8, 9].

The term depression is used invariably to describe both depressive symptoms and the disorder. A depressive disorder is one of the most common psychiatric conditions in patients with cancer, but frequently goes unnoticed [86]. This may be because of low awareness among health care providers, because it is difficult to distinguish between mild depressive symptoms as opposed to a depressive disorder, or because the physical symptoms receive more attention in clinical practice. A meta-analysis reported equal pooled prevalence rates of depression of 24% in cancer patients in palliative care settings and 16% in oncological and hematological settings [87].

In clinical trials, changes in symptoms may indicate a treatment benefit or toxicity. For example, a reduction of symptoms may be a primary or secondary endpoint in trials where comparison of treatments with similar anticancer effect may indicate that one of the treatments was associated with fewer adverse events, or where worsening of symptoms might

be considered [45]. Osoba et al. randomized 161 patients with castration-resistant prostate cancer in two groups receiving either mitoxantrone intravenously plus prednisolone or prednisolone alone [88]. They observed no difference in overall survival, but the patients receiving mitoxantrone had superior global QoL and pain control and improvements in several areas including physical functioning [88, 89]. In addition, the data allowed investigators to show that mitoxantrone reduced costs by preventing hospitalizations [90]. Based on this trial and a similar one [91], the FDA approved the drug. The Nordic Myeloma Study Group investigated the effect of interferon on HRQoL in multiple myeloma in a randomized controlled trial [92]. The EORTC QLQ-C30 questionnaire and 11 supplementary items relating to interferon toxicity were used. Results showed that the administration of interferon during induction treatment with melphalan and prednisone caused increased symptom and toxicity scores, and lower scores for global health and QoL with no superior effect on overall survival [92].

2.3 Symptom assessment

2.3.1 Symptom checklists/inventories

Optimal management of symptoms relies on frequent and accurate symptom assessment. Several symptom assessment tools (PROMs) have been developed in the form of checklists or disease-specific or generic questionnaires of different lengths and formats.

2.3.1.1 M.D. Anderson Symptom Inventory (MDASI)

The M.D. Anderson Symptom Inventory (MDASI) was developed at the M.D. Anderson Cancer Center at the University of Texas [23]. Following the development of the Brief Pain Inventory (BPI) [93] and the Brief Fatigue Inventory [94], the more complex MDASI was designed. The development process was an iterative process involving (1) selecting items for the initial scale by systematic evaluation of previous work in multiple-symptom assessment, (2) consulting with oncology health care professionals, and (3) modifying the list of potential items based on data obtained from patients with cancer-related symptoms [23]. The MDASI assesses the severity of 13 frequently experienced symptoms by patients with various cancer diagnoses and in different types of treatment. In addition, the MDASI assesses how much all symptoms interferes with the following domains: walking, work, general activity, mood, relations with others, and enjoyment of life. The response alternatives are 0–10 on a numerical rating scale (NRS), with 0 meaning *"not present"* and 10 meaning *"as bad as you can imagine"*. The MDASI was developed using common elements of test validity (content,

criterion, and construct) and reliability (internal consistency and test-retest/stability). These standards are similar to those proposed by the FDA's Draft Guidance for Industry [16]. Cognitive debriefing was added to the instrument development process. A total of 60 patients with thyroid cancer were asked to evaluate their understanding and ease of comprehension of the symptom items to which they were being asked to respond. Most of the patients reported that the MDASI items were relevant and easy to understand [95]. Use of recall period (e.g. the past week, the past 24 hours, or currently) may provide a more accurate picture of a patient´s symptom status. In clinical research, the choice of a suitable recall period depends on the specific purpose of the trial, the characteristics of the disease, and the treatment to be tested. As with the BPI, the MDASI can be used within a 24-hour recall period or a past-week recall period.

The reasons for using the NRS are that 0–10 ratings of symptoms are easy to adapt to both clinical and research needs and can be used by patients at home using a numeric keypad. The NRS was chosen from among several widely used options, including verbal descriptor scales (VDS), which use word descriptors such as *"none", "mild", "moderate", "severe",* and *"excruciating"* to describe severity, and visual analog scales (VAS), in which the patient indicates what portion of the line anchored by *"none"* and *"as bad as you can imagine"* is equivalent to the severity of symptoms. A comparative study showed a high degree of association between the VDS, VAS and NRS [96]. Further, the NRS has been found to be more reliable and easier to complete than the VAS [97]. The test-retest reliability of the MDASI has been examined in several studies. For example, in a sample of 33 patients with multiple myeloma, non-Hodgkin's lymphoma, or breast cancer who underwent autologous transplantations, test-retest reliability over a 30-day period was calculated. Coefficients ranged from 0.75 to 0.96, indicating that the MDASI is a reliable and sensitive symptom assessment tool [98]. Internal consistency reliability reflects whether the items in a domain are intercorrelated, as evidenced by an internal consistency statistic (e.g. a Cronbach coefficient alpha > 0.7). The internal consistency of the MDASI was demonstrated in the initial validation sample by Cronbach coefficient alphas of 0.85 for the general symptom severity items and 0.91 for the interference items [23]. The smallest difference that is considered clinically important (the minimum important difference (MID)) for the MDASI is following guidelines for HRQoL instruments and is set to be about half a standard deviation [98]. For example, given a standard deviation value of 1.95 for the 13 core items [23], the MID is estimated to be 0.98.

2.3.1.2 Edmonton Symptom Assessment System (ESAS)

The Edmonton Symptom Assessment System (ESAS) was developed in 1991 as one of the first symptom assessment instruments. ESAS has been psychometrically validated and translated into several languages [99]. ESAS is now commonly used for symptom screening and longitudinal symptom monitoring in patients seen by palliative care, oncology, nephrology and other disciplines, in both clinical practice and research [99]. The initial version consisted of eight horizontal 0-100 mm visual analog scales and has later evolved to 11-point numerical rating scale ranging from 0 (no symptom) to 10 (worst possible symptom). Several studies have examined patients´ perceptions of ESAS, and reported that the items of appetite and sleep could be misinterpreted, and that some patients had difficulty understanding for example the term depression, anxiety and wellbeing [100, 101]. Findings led to the proposal of a revised ESAS (ESAS-r) consisting of 9 core symptoms and an optional 10th symptom. The time frame of symptom assessment was stated as "now" [102].

The EAPC basic dataset is a minimum dataset for reporting patient characteristics and medical variables in a palliative care cancer population, developed by an international Delphi process. In the first round, ESAS-r was included. After five rounds, a consensus on content was reached. The EAPC basic dataset contain 31 variables, including assessment of 12 common symptoms using NRS scales from 0 to 10 [103]. All ESAS-r symptoms are included in the EAPC basic dataset: Pain, tiredness, drowsiness, nausea, lack of appetite, shortness of breath, depression, anxiety and wellbeing. In addition, the EAPC basic dataset includes sleep, constipation and vomiting.

Cut-off values are used to determine if PRO scores represent clinically relevant symptom burden. Selby et al. included 400 cancer patients in a prospective study to examine the relationship between the numerical and verbal scores using the ESAS and to identify a cut-off for severe symptom intensity. Findings suggest that a score of 7 or higher represents a severe symptom intensity across all ESAS symptoms [104]. Generally, ESAS scores of 0, 1–3, 4–6 and 7–10 are considered as none, mild, moderate and severe in clinical practice [105], although there may be significant variations in how the individual patient interprets the scores [106]. A prospective multicenter study was conducted by Hui et al. to identify the minimally clinically important difference for each of the 10 ESAS symptoms [107]. In total, 796 cancer patients were included. The patients were asked about their average symptom intensity over the past 24 hours at the first visit and a subsequent visit three weeks later, and they were asked to provide the assessment of change (better, same, or worse). Results showed that a

change of 1 point was the optimal cut-off for both improvement and deterioration for each of the 10 symptoms, with sensitivities of 59% to 85% and specificities of 69% to 85% [108]. The minimum important difference of a PRO measure represents the smallest improvement or deterioration that patients perceive as important and which would lead clinicians to consider a change in care [109]. By representing the smallest clinically significant score changes, MIDs facilitate interpretation of patients responses to treatment and other changes over time [110].

2.3.1.3 Barriers to use of PROMs

Patient-reported data have often not been collected in clinical cancer care, and unfortunately systematic collection and use is still not part of daily clinical routine [111, 112]. Despite the increased focus on the patient perspective and PROs, symptoms still go undetected by clinicians [113, 114], and the correlation between patient-reported symptoms and clinicians' reports is poor [115]. A cross-sectional study (N=194) found that most patients (93%) reported some degree of side or late effects after treatment for breast cancer, with significantly more side effects or late effects reported by the women than registered by the oncologists ($p<0.001$) [116]. The most commonly cited barriers for routine use in clinical practice are logistical problems, cumbersome use, time constraints, resistance to change and difficulties related to interpretation [24, 36, 117, 118]. However, implementation of systematic symptom assessment in clinical practice is important to increase the quality of diagnostics and treatment [119, 120].

2.4 Psychometric considerations

Regardless of format, PROMs must conform to accepted standards, i.e. validity, reliability and sensitivity to change (table 1) [52]. Validity testing examine whether the instrument measure the concepts they are intended to measure. Content validity, or the related, "face validity", is the extent to which the instrument measures the concept of interest, i.e. that the items and domains of an instrument are specific to the population, condition, and treatment to be studied [17]. There is no specific technique for testing content validity; the judgement is usually based on a review and consensus by an expert panel of health-care workers and/or patients [45]. The FDA suggests that the following development and instrument attributes should be evaluated in all PROMs to assess content validity: 1) item generation should include input from the population of interest, 2) data collection methods or administration modes should be specified, 3) the most suitable recall period should be chosen, 4) the

response options for each item should be consistent with its purpose and intended use (e.g. VAS, NRS), 5) patients' understanding should be examined, and 6) the summary score should be appropriate and the respondent burden should be tolerable [17].

In addition, construct validity, reliability and the instrument's ability to detect changes are required. Construct validity is defined as "evidence that relationships among items, domains and concepts conform to *a priori* hypotheses concerning logical relationships that should exist with other measures or characteristics of patients and patient groups" [17]. Reliability is defined as "the ability of the questionnaire to yield consistent, reproducible estimates of true treatment effect" [17]. Tests of reliability seek to determine whether a PROM reliably measures the concept it was designed to measure, and to establish the quality of the evidence of reliability.

Table 1: Psychometric properties

Validity: Whether the instrument measures what it is intended to measure	Content validity: the extent to which the instrument is sensible and reflect the concept of interest Construct validity: the degree to which a test measures what it claims to be measuring, i.e. forming a hypothetical model describing the constructs being assessed (e.g. pain, fatigue, anxiety) and postulate their relationships (the extent of consistency in a multiple-item measurement) Criterion validity: the extent to which the scale has empirical association with external criteria such as other established instruments ("gold standard")
Reliability: Whether the instrument reliably measures the concepts it was designed to measure (consistency, precision, repeatability, trustworthiness)	Test-retest reliability measures the stability of a test score over time when no change has occurred in the concept of interest Internal consistency reliability reflects whether the items in a domain are intercorrelated, as evidenced by an internal consistency statistic (e.g. a Cronbach coefficient alpha > 0.7)

	Parallel forms reliability measures the correlation between two equivalent versions of a test
Ability to detect change: Whether the instrument detects differences in scores over time with respect to the measurement concept	

2.4.1 Reporting issues

PROs are subjective and require completion by patients, and as such they present a range of scientific and logistical challenges for researchers [121]. Robust methodology and accurate reporting of data are crucial when evaluating PROs in clinical trials [122]. A systematic review assessed the quality of patient-reported outcome data in advanced breast cancer randomized controlled trials (RCTs) between 2001 and 2017. Only 12% reported a PRO specific hypothesis, and 73% did not report how missing data were handled [123]. Systematic reviews on cancer RCTs have also shown a heterogeneity of statistical methods were used to evaluate PRO data [123-125]. The variety of statistical methods makes it challenging to compare findings across trials and to build on previous work to make results more generalizable. Further, missing data is a common problem in analyses of PRO data in trials. If the amount of missing data is substantial, this may bias the analysis and critically influence the conclusions that can be drawn [126, 127]. Guidelines exist to improve reporting of PROs in protocols (Standard Protocol Items: Recommendations for Interventional Trials-PRO extension (SPIRIT-PRO)) [128] and publications (Consolidated Standards of Reporting Trials Statement-PRO extension (CONSORT-PRO)) [129]. The Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data for Cancer Clinical trials (SISAQOL) Consortium [130] was established to provide recommendations on how to standardize the analysis of PRO data in cancer randomized trials [130, 131]. The SISAQOL provides a framework of well-defined PRO research objectives and appropriate statistical methods developed through literature reviews and expert discussions [132].

2.5 Development of electronic symptom assessment tools (e-PROMs)

New advances in health information technology have promoted the development of electronic tools for collection of PROMs the last two decades. Some are simply a paper PROMs in its original paper form adapted for a screen. Other PROMs have been adapted or developed and programmed in a way that opens a whole new array of options regarding flexibility and adaptation to the individual user, immediate transfer of data, and compatibility with other clinical resources like laboratory tests, imaging and journals. The requirements of a PROM are that it is acceptable to patients, easily accessible for clinicians for interpretation of results, and that the measurement properties are acceptable. Confidentiality issues and adherence to regulations regarding handling of data must be ensured, as is the case with paper tools as well, but more complicated than locking a cabinet. A review by Jensen et al. [25] identified 33 electronic patient-reported outcome (e-PRO) systems implemented in cancer care. The systems were generally developed to improve symptom management, identify psychosocial problems, and facilitate communication. Most of them (63%) were intended for use during treatment, most commonly chemotherapy, while 40% were also used in follow-up care. The majority (85%) of the systems sent real-time alerts based on patient responses directed to clinicians, 44% were integrated into the electronic health record (EHR) while some were flexible regarding location of administration. Almost all systems (96%) provided summaries of patient-reported data to pre-specified providers, and 93% provided summaries of PROs over repeated assessments. Findings suggested that usability and integration in clinical care were important system characteristics.

Riis et al. conducted a randomized trial to evaluate patients' satisfaction with individualized follow-up care after treatment for hormone receptor positive early breast cancer (N=134) [133]. Patients were randomized to receive standard follow up care with prescheduled consultations every six months or individualized follow up care reporting PROs electronically every third month over two years. E-PROs were used both as a screening tool for patients' problems and as a dialogue tool. Consultation were planned according to the urgency of the reported problems. The questionnaire included the EORTC QLQ-C30, and the EORTC breast cancer module [53, 134]. Results showed that women in standard care group attended twice as many consultations during the follow-up period as women in individualized care (4.3 vs. 2.1, 95% CI: 1.6–2.6, $p < 0.001$). There were no statistically significant differences reported in relation to unmet needs, QoL or adherence to treatment [133].

A single-center study randomly assigned patients receiving chemotherapy for metastatic solid tumors to self-report 12 common symptoms via tablet computers, or to usual care (N=766).

Treating physicians received symptom printouts at visits and nurses received email alerts when participants had reported severe or worsening symptoms. Patients in the e-PRO group reported improved quality of life (34% vs 18%), were less frequently admitted to the emergency room (34% vs 41%; p=0.02) or hospitalized (45% vs 49%; p=0.08), and remained on chemotherapy longer (mean 8.2 vs 6.3 months; p=0.002 [33]. Systematic reporting of PROs even showed results suggestive of lengthened median overall survival (31.2 vs 26.0 months; p=0.03) [40]. A multicenter randomized clinical trial compared web-based electronic symptom monitoring vs standard scheduled imaging to detect symptomatic recurrence in patients with lung cancer following initial treatment [39, 135]. In the e-PRO group, patients were invited to complete weekly self-reports of 13 common symptoms online between clinical consultations. The PRO system automatically triggered an alert email to the oncologist when patient-reported symptoms matched predefined criteria for severity and worsening. The questionnaire included common cancer-related symptoms and symptoms indicating progression of lung cancer. Symptoms were scored from 0 (no symptom) to 3 (major symptoms). Results showed that median overall survival was 22.5 months in the intervention group and 14.9 months in the control group [38]. A potential mechanism for this huge difference is that symptoms suggesting adverse events or recurrence were detected earlier on.

2.5.1 Eir

Our research group in the European Association of Palliative Care Research Network (EAPC RN) [136] and European Palliative Care Research Centre (PRC) [137] has developed several prototypes of electronic symptom assessment tools during the past decades. PAT-C was a first prototype tested in a pilot study and national clinical studies showing that the majority of palliative care cancer patients were able to report symptoms directly on a touchscreen computer [138]. In 2008–2009, a more sophisticated software version for tablets was tested in 1017 patients with advanced cancer in the EPCRC-CSA, an international cross-sectional study involving 17 centers in eight countries [26]. The patients responded to questions about symptoms, nutritional intake, and physical and emotional functions. The software contained several skip sessions to reduce patient burden; if the patient had no pain, further questions on pain were omitted. The completion rate was high (95%), with more missing information and need for assistance associated with higher age and lower performance status. The software was feasible, even for patients with little digital experience. A later version that also provided treatment recommendations for pain and depression, was developed in 2012 and tested by 143 patients in a controlled before-and after study [139]. Findings suggested that this

symptom assessment and decision support system did not improve pain intensity. The development and iterative tests of a computerized pain body map clearly demonstrated the need to optimize the design and user-friendliness for use in the frailest patients [140, 141].

Introducing changes in an organization, be it technology-driven or more conventional ones such as new routines, often leads to resistance and poses several challenges. In that respect, a user-centered approach involving the end-users all along the process of the development of an electronic symptom assessment tool, is important [26]. Based on our former experience with e-PROMs it became evident that this is an important criterion for a probable success for clinical use. Our experiences led to the development of Eir, an electronic symptom assessment tool, for use in treatment of adult patients with cancer in all stages of the disease trajectory and in different clinical settings. The overall objective of Eir is to improve symptom management by introducing a systematic, standardized way of assessing PROs that is well-perceived by all end-users and that makes patients' self-report of symptoms and problems immediately available in clinical consultations.

3. Aims of the thesis

Overall aims:

The overall aim of this thesis is to develop a new e-PROM solution; EirV3, built on a better understanding of patient-reported outcomes to improve patient-centered care.

More specifically, I will answer the following research questions:

- What is the prevalence of symptoms assessed by the MDASI in the general Norwegian adult population?
- Which factors are associated with a high symptom burden?
- What is the rationale for the use of e-PROMs from a technical and patient-centered perspective?
- What is the optimal content and technical format of an electronic tool for patient-reported outcomes in clinical practice?
- How do patients and health care providers evaluate the content and format of EirV3?

4. Materials and methods

This thesis is based on three studies, presented in table 2.

Table 2: Overview of study designs and populations

Paper	**Localization**	**Study design**	**Population**	**PROMs**	**Number of participants (total/analyzed)**
I	National population study	Cross-sectional study	General population	MDASI	N=2116/2021
II	I. Single center study Regional II. National multicenter study Six centers	I. Descriptive study II. Randomized, comparative study	I. Cancer patients Physicians at the Cancer Clinic II. Cancer patients	 EirV1 EirV2 EirV3 PROs on paper PROs on tablet	I. Patients: N=75 Physicians: N=8 II. N=114/110
III	Single center study Regional	Usability evaluation Observations and interviews	Cancer patients	EirV3	Patients: N=37 Physicians: N=17

4.1 Materials and methods paper I

4.1.1 Subjects and data collection

In 2015, a total of 6,165 subjects were randomly drawn by Bring Dialog. The sample was representative of the Norwegian adult population with respect to age (18–80 years), sex and place of residence. All subjects received a postal questionnaire packet on paper containing the Short-Form Health Survey (SF-36), version 1 [46, 47], the M.D. Anderson Symptom Inventory (MDASI) [23], the Fatigue Questionnaire (FQ) [60] and the Patient Health Questionnaire 9 (PHQ-9) [59, 142]. The questionnaire packet also included questions covering comorbidities, and 14 questions on sociodemographic background, physical activity, general health and contact with health care providers. Background variables included year of birth, sex and level of education. Level of education was divided into three groups: elementary and/or primary school; second level, (high school); and third level (university or university college). Medical comorbidities were self-reported on the Self-Administered Comorbidity Questionnaire (SCQ) [143]. In this study, data on age, gender, education, the SCQ, the PHQ-9 and the MDASI were used.

4.1.2 Measurement tools

4.1.2.1 The M.D. Anderson Symptom Inventory (MDASI)

The M.D. Anderson Symptom Inventory (MDASI) is a validated multi-symptom patient-reported outcome measure for use in symptom surveys, clinical trials and patient follow-up care [144-146]. The MDASI is designed for use in general cancer populations [23] as it assesses the severity of 13 frequently experienced symptoms by patients with cancer. These symptoms are pain, fatigue, nausea, disturbed sleep, distress/feeling upset, shortness of breath, difficulty remembering, lack of appetite, drowsiness, dry mouth, sadness, vomiting and numbness/tingling. The response alternatives for each symptom are 0–10 on a numerical rating scale with 0 meaning *"not present"* and 10 meaning *"as bad as you can imagine"* in the last 24 hours. The responses can be used as single items or as an added sum score. All 13 symptoms included in the MDASI are also experienced by people without cancer and represent common reasons for contact with the health care system [23]. In addition, the MDASI includes another six questions about how much the symptoms interfere with general activity, mood, work, relations with other people, walking and enjoyment of life. The response alternatives are 0–10 with 0 meaning *"did not interfere"* and 10 meaning *"interfered completely"*.

The first introductory sentence in the MDASI refers to people with cancer. For the purpose of our general population survey, the word cancer was removed and was not used anywhere else in the questionnaire. The translation of MDASI into Norwegian followed the multi-step, well-established procedures developed by the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group, according to the 2009 procedure [147]. Two independent forward translations from English to Norwegian were done by native speakers of the target language, then the translations were reviewed to reach a reconciled version, prior to two independent back translations into English. When comparing the original and the back-translated English versions, no translation problems became apparent. The Norwegian version of the MDASI was proof-read and pilot-tested in six persons who found the comprehensibility and clarity satisfactory according to the EORTC debriefing interviews. Permission to translate and use the MDASI was obtained from MD Anderson, TX, USA.

4.1.2.2 The Patient Health Questionnaire-9 (PHQ-9)

The PHQ-9 is a nine-item questionnaire designed to screen for depression [59]. The nine PHQ-9 items correspond to the DSM-5 diagnostic criteria for major depressive disorder and include anhedonia, depressed mood, sleep-problems, fatigue, weight/appetite change, feelings of worthlessness/guilt, poor concentration, psychomotor retardation/agitation and thoughts of self-harm/suicidal ideations [148]. The response alternatives assess the frequency to which these symptoms have been bothersome during the past two weeks and include four categories: 0= "*not at all"*, 1= "*several days"*, 2= "*more than half of the days"* and 3= "*nearly every day"*. Major depression is diagnosed if five or more of the symptoms have been present at least "more than half the days" in the past two weeks with one of these being item 1 (depressed mood) or item 2 (anhedonia). As a severity measure, the PHQ-9 sum score ranges from 0–27, since each item can be scored from 0 to 3.

4.1.2.3 The Self-Administered Comorbidity Questionnaire (SCQ)

The SCQ is a brief, comprehensive, self-administered questionnaire to assess comorbidities [143]. The questionnaire includes 12 common conditions. The questions "*Do you have any of the following problems*" are asked in relation to heart disease, high blood pressure, lung disease, diabetes, ulcer or stomach disease, kidney disease, liver disease, anemia or other blood disease, cancer, depression, arthritis and back pain. The response alternatives are "*yes*" or "*no*".

4.1.3 Analyses

Basic descriptive analyses were used to analyze baseline characteristics of the sample and the number and intensity of MDASI symptoms. The total MDASI sum score for the 13

symptoms was calculated (possible range 0–130, i.e. the sum scores for the 13 individual symptoms on the NRS 0–10). In the analyses, the four somatic depression symptoms in the PHQ-9 (sleep-problems, fatigue, weight/appetite change and psychomotor retardation/agitation) were excluded as these overlaps with the following MDASI symptoms: sleep-problems, fatigue, weight/appetite change and psychomotor retardation.

Associations between the MDASI sum score as the dependent variable, and age, sex, education, comorbidity and depression assessed by the PHQ-9 as independent variables were analyzed using univariable linear regression. Variables with a p-value ≤0.10 were included in a multivariable regression model, which also included sex and age regardless of the significance in the univariable analyses. The interference items were used as dependent variables in separate analyses. The corresponding effect sizes were reported as unstandardized coefficients and 95% confidence interval (CI). A p-value of <0.05 was used to denote statistical significance.

The statistical software applied was IBM SPSS Statistics for Windows, version 25.0, (IBM Corporation, USA).

4.1.4 Ethical considerations

The study was performed according to the rules of the Helsinki declaration [149]. All respondents received written information about the study. Return of the questionnaires was taken to indicate written, informed consent. The Regional Committee for Medical and Health Research Ethics (REC) South East Norway approved the survey (REK-2014/1172).

4.2 Materials and methods paper II and III

4.2.1 Paper II: Participants and data collection

Eir has been designed following expert-driven and user-driven approaches and an iterative development process (from 2013 to 2016), with literature reviews on traditional and electronic assessment and classification methods and usability testing. Between 2013 and 2015, regular meetings were held by the international expert panel and the core Norwegian working group, in addition to two national workshops.

The core Norwegian working group consisted of a total of 15 oncologists, palliative care physicians, researchers, interaction designers, graphic designers and software developers.

Members of the core working group were the first to test each new feature of Eir, and patients at the Cancer Clinic were regularly involved in testing suggestions for functionality and user interface. Regular meetings were organized with members of the core working group. Functionality and features were discussed based on findings and observations from patient testing, and necessary refinements were agreed upon.

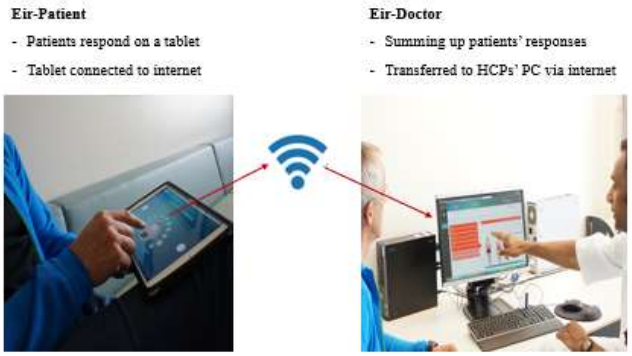
Prior to the software development, two national workshops were conducted assessing the needs and preferences of the end-users: patients, health care personnel and researchers. The first workshop (2013) presented the overall idea and intention of Eir to the participating oncologists, nurses, designers and cancer patients (N=20). Eir has two modules: Eir-Patient and Eir-Doctor. Eir-Patient is for patient self-report on tablets or computers. When the physicians log on to Eir-Doctor on their computer, the PRO registrations have been wirelessly transferred and transformed to a special format designed for immediate use in clinical consultations. The intended features of Eir-Patient such as content, layout and functionalities were presented, and feedback suggestions from the participants were collected. The second workshop in 2013 was conducted with five oncologists who suggested two different ways of presenting the patient-reported outcomes when opening Eir-Doctor, either with as much information as possible on the opening screen, or to highlight only the most relevant information.

The first international meeting was organized in 2013. Here, 26 oncologists and palliative care physicians were recruited from Italy, England, Scotland, Germany, Denmark and Spain as part of the European Partnership for Action Against Cancer [150]. Electronic symptom assessment and development of Eir was discussed. The meeting was organized with short introductions about the objectives of Eir and symptom assessment followed by plenary discussions and two workshops. Decisions were made regarding content, structure, concept and design:

- The content of Eir should be based on evidence or consensus assessment methods
- Eir should have a hierarchical structure, with an introductory question about the patient's well-being today prior to a screening section on symptoms, followed by a section on symptom intensity and another section for characterization and more detailed assessments of the endorsed symptoms
- Registrations in Eir should be immediately transferred and visually presented (figure 1)

- Eir should be user-friendly and relevant for heterogeneous cancer populations
- Eir should be easy to adapt to other languages and cultural and clinical preferences

Figure 1: Structure of Eir



The second international meeting (2013; n=9) was arranged after user-testing of the first version of Eir-Patient. Here, feedback regarding content and layout was collected and summarized. Subsequent discussions resulted in consensus on which symptoms to include in Eir and how the included items should be structured and presented. The third meeting (2014) consisted of five experts in neuropathic and breakthrough pain and focused on pain assessment in Eir.

In 2013, the first advanced prototype of Eir-Patient (Eir version1) was systematically tested by cancer patients at an outpatient clinic. Ten patients tested the pain body map within Eir for marking pain location, and seven patients tested the first complete version of Eir-Patient. In 2014, a pilot test of the revised Eir-Patient (version 2) was tested by seven cancer patients at the outpatient clinic. The initial testing was performed by patients only. Further, the Eir-Patient and Eir-Doctor modules were tested in 42 patients and eight physicians at the Cancer Clinic. Patients were recruited from the cancer outpatient clinic and differed by age, sex, cancer diagnosis, and treatment intent (curative and palliative). Patients completed Eir-Patient on tablets in the waiting room, and physicians used Eir-Doctor in the consultations. Usability data were collected through interviews and observations. Findings from this test led to ample amendments of Eir, including a redesign of the user interface. In 2015, nine cancer patients pilot tested Eir-Patient version 3 at the outpatient clinic.

The comparative study

In 2016, a comparative study was carried out among 114 patients with cancer at six Norwegian hospitals to examine agreement between PRO assessments on tablets and paper and to assess patients´ preferences for either method. A simplified, shortened version of Eir was used to assess intensity of 19 common cancer-related symptoms on a numerical rating scale from 0–10. The questions were similar in the electronic and the paper version. Patients were randomly assigned to complete the questions either electronically or on paper first. After a waiting period of at least 30 minutes, patients completed the other version. Sociodemographic data and medical information were registered by health care personnel. Eligible patients were palliative cancer patients aged 18 years or older coming for a scheduled appointment at the cancer outpatient clinics. Patients with obviously impaired cognitive function as judged by the treating clinician according to common criteria (e.g. problems with orientation, coherent speech, memory, and attention span) were not included.

4.2.2 Paper III: Subjects and data collection

A usability evaluation using observations, think-aloud sessions and interviews with patients and physicians was conducted. The study was performed from September 2015 to September 2017. Patients were included at a local hospital and at a university hospital. The focus was to gain new information about usability issues regarding barriers experienced by the end-users using the system, e.g. patients and health care professionals.

Patients: Recruitment of patients was done by purposive sampling to ensure variation in age, gender, diagnosis and anticipated symptom burden. Eligible patients were diagnosed with cancer in all phases of the disease trajectory, i.e. both curative and palliative settings. Participants were above 18 years of age with no upper age limit. Patients with obvious cognitive impairment as judged by the physician according to established criteria were not included.

Physicians: All participating hospital physicians were oncologists, and they were recruited from cancer outpatient units at a university hospital and at a collaborating local hospital. General practitioners (GPs) were recruited from two nearby municipalities.

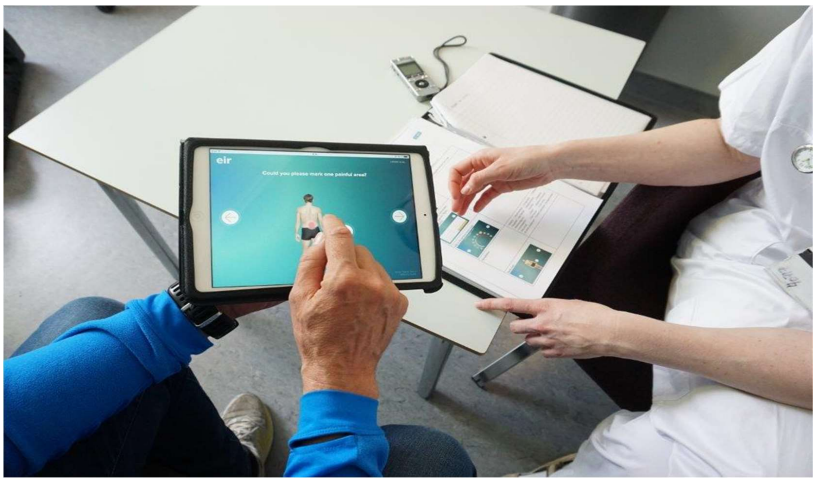
EirV3 was completed by the patients either (1) while waiting for a scheduled consultation at the cancer outpatient clinic, (2) at home between consultations, or (3) prior to a scheduled consultation at the GP's office. The patients who had their GP in the two nearby municipalities

were encouraged to visit their GP. Before a scheduled consultation at the GP`s office, the patients were observed while completing EirV3 at home.

Patients were briefly introduced to EirV3, and a researcher provided the login information before the patients started to use the tool. The think-aloud method [151] and patient interviews were used to collect data on the patients' practical use of the tool, and to provide insight in their immediate reactions and experiences when using it [152]. Patients were encouraged to think-aloud, i.e. to constantly verbalize their thoughts while completing the Eir-questions [151]. If they were unsure on how to proceed, they were encouraged to do what they found most intuitive, before being assisted by the researcher if needed. The patients were observed by the researcher as they used EirV3 (figure 2). Field-notes were made based on a predefined observation template covering navigation errors, ease of use, apparent misunderstandings or technical difficulties. The patients were interviewed by the researcher after completion. The interview was structured, following an interview guide designed specifically for this study. The content of the interview guide was based on previous usability studies of electronic symptom assessment tools [153, 154], with standardized, open questions about potential difficulties regarding understandability, practical use, design, layout and time expenditure. In addition, the participants were asked if they had any suggestions for change. Specific usability issues that had been observed were also addressed in the interview. The whole session was audio-recorded. Patients who used EirV3 more than once, were asked to take part in a second interview following the same interview guide.

Physicians were involved in the testing of EirV3 in clinical consultations to observe the flow of information and the actual use of the data gathered in Eir. Prior to study start, physicians received a quick introduction on how the patient's responses would appear in Eir-Doctor. A thorough instruction was not given, since the intention was for *them* to decide how to use EirV3 in their clinical work. The physicians' use of Eir-Doctor during the consultations was observed by the researchers. Field notes were made based on a predefined observation template. By the end of the study, physicians were invited to attend individual interviews to summarize their experience with the tool. The physicians were asked if and how they used EirV3 before and during consultations, difficulties regarding the use of EirV3, their perceived potential benefits or disadvantages of using Eir and whether they had suggestions for changes.

*Figure 2: Patients were observed as they used Eir-Patient**



* The picture is arranged, i.e. not from a real patient

4.2.3 Eir

Eir gives the patient the opportunity to report symptoms by using a tablet, connected to an internet source, e.g. Wi-Fi. The patient answers the questions by ticking the appropriate alternative or scoring the symptom intensity prior to a consultation or at home in between consultations.

The third version of Eir, EirV3, has been developed for and by end-users through iterative development including regular user testing and continuous amendments [155]. The Eir-Patient module includes items assessing 19 of the most common cancer-related symptoms, and items related to level of functioning and nutritional status (figure 1) [155]. The selection of the content of Eir was based on literature reviews, expert opinions, clinical experience and evidence-based guidelines for symptom management [156-162]. The symptoms and follow-up questions in Eir-Patient are based on well-validated PROMs, i.e. EAPC basic dataset [103], Patient-Generated Global Assessment (PG-SGA) [163], Patient Health Questionnaire-9 (PHQ-9) [142], General Anxiety Disorder-2 (GAD-2) [164], Insomnia Severity Index [165] and Common Toxicity Criteria for Adverse Events (CTCAE 4.0) [166].

Technical specifications and data safety

EirV3 is developed as a web site using standard HTML5, CSS3 and Javascript, and is designed for ease of use and touch based navigation. This allows the system to run on any platform with

a modern web browser: tablets, cell phones, laptops, workstations and public terminals. It is designed to run on Windows Server using IIS, but it also supports running on Windows Azure. For data storage, Microsoft SQL Server is the default database system. However, other storage methods such as Azure Blob storage and document databases are supported as well.

4.2.4 Analyses

Usability testing

Feedback from end-users, i.e. patients and physicians were perceived as crucial before and during the development process. Thus, testing and retesting of Eir with continuous feedback from patients and health care providers have been a major endeavor during the development of Eir, towards the clinical objective to improve symptom management. The aim of the usability testing of Eir was to assess patients' and physicians' evaluation of ease of navigation, clarity of instruction and relevance of the content of Eir-Patient and Eir-Doctor. Formative usability tests [167] on separate sections, e.g. general pain and breakthrough pain, as well as on more complete versions of Eir prototype modules have been performed by patients at the Cancer Clinic, St. Olavs hospital, Trondheim University Hospital during the entire development process. Usability issues in the first versions led to immediate modifications of the system. For example, patients were invited to test the pain body map by marking painful areas, or to navigate between questions in the program in a real clinical setting. The patients gave feedback on whether they understood how to use it, if they could mark all sites with pain and whether they had suggestions for changes in functionality or design. The feedback from end-users were presented for the working group and improvements were made based on the feedback while taking technical and professional requirements into account.

In study III, all audio-recorded material (think aloud-sessions and interviews) was transcribed verbatim by one of the first authors and was analyzed together with the field notes. Usability issues were identified by use of simple content analysis, categorised and rated by the authors (SSH, HK, KS), guided by the approach by Rubin and Chisnell [152]. Identified usability issues were categorised as follows: Understandability, visibility, workflow, content, navigation and bugs. The number of participants experiencing each issue was registered. Each issue was graded on a scale from 1 to 4 (1=irritant, 2=moderate, 3=severe, and 4=unusable) [152], based on the severity of the problem, frequency and potential for affecting treatment. The grading was done by the authors (SSH, HK, KS) independently and

subsequently compared and discussed until consensus was reached on each issue of divergence.

Usability is commonly defined as the extent to which a product can be used to achieve specific goals with effectiveness, efficiency and satisfaction in a defined context of use [168]. Repeated testing of usability during development of any electronic medical system is the preferred method for identifying and solving usability issues [17, 154, 169]. Usability testing provide information about the ease of use and guide further development. It is essential that end-users are actively involved in the development and testing of electronic health devices, to identify strengths and limitations of content or functionality [154, 170]. Further, the electronic devices must be perceived as useful by the end-users if they are going to be used in daily clinical practice [26]. Formative usability testing is done early in the product development to help form the product´s shape and design and identifies why something does not work as intended. Summative usability testing is evaluating a product at the end of the development process through defined measures relative to usability.

Statistical analyses

Participant characteristics and mean symptom intensity scores were summarized by applying means, proportions and standard deviations (SD). In the comparative study, intraclass correlation coefficients (ICCs) based on a two-way mixed effect analysis of variance, single measure and absolute agreement were used to examine agreement of tablet and paper scores. According to interpretation guidelines [171], an ICC > 75 indicates excellent agreement.

4.2.5 Ethical considerations

Confidentiality issues and adherence to all regulations regarding the transfer, handling and storage of data was a major issue during the development process. Data communication between the device on which the system is running, and the site where the data are stored, was secured using HTTP over SSL. Verification of the patient's identity was ensured using token-based authentication, with support for numerous authentication protocols including OAuth, OpenId, and SAML 2.0. Patient data were stored anonymously on secure servers hosted by each clinic or institution using Eir according to approval by each hospital´s data protection supervisors and IT departments. The data was also encrypted using AES encryption to ensure that access to the database required an encryption key. No data were stored on the tablets. Physicians using Eir must log into Eir-Doctor (password-protected) to retrieve the registrations the patient made in Eir-Patient. Accordingly, only the patients' study IDs were presented to the

physicians. The patients logged on to EirV3 using a randomly generated study ID. Patients received oral and written information about the study. Written consent was signed before inclusion.

In study II, the Regional Committee for Medical and Health Research Ethics, Central Norway, was consulted, and they concluded that the usability tests did not require ethical approval while the comparative study was approved (REK-2015/185).

In study III, approval was obtained from the Regional Committee for Medical and Health Research Ethics, Central Norway (REK-2014/212 and REK-2015/185).

5. Results and summary of papers

5.1 Paper I

Symptoms in the general Norwegian population-prevalence and associated factors

The response rate was 36%. Of these, 1101 (54%) were female and 920 (46%) were male. Mean age was 55 years with a standard deviation (SD) of 14, ranging from 18 to 79 years. Forty-six percent of the respondents had university or university college education. The most frequent comorbidities overall were hypertension, arthrosis and depression. The most frequent symptoms (cut off ≥ 1) were fatigue (60%), drowsiness (56%) and pain (56%). When using a cut off ≥3, the prevalence was 34.8% for fatigue, 34.2% for pain and 26.7% for drowsiness. The symptoms fatigue, pain and disturbed sleep had the highest mean scores overall. Linear regression analyses showed positive significant associations between the MDASI sum score, depression on the PHQ-9 ($p<0.001$) and the presence of one or more comorbidities ($p<0.001$). Participants with the highest education level had significantly lower MDASI sum scores than respondents with education in levels one ($p=0.006$) and two ($p=0.003$). Comorbidities, PHQ score and MDASI sum score were significantly associated with the interference items general activity and work as dependent variables ($p\leq0.001$). More comorbidities and higher MDASI sum score were significantly associated with higher scores on the interference item walking ($p<0.001$). Further, the multivariable regression analyses showed that PHQ score and MDASI sum score were significantly associated ($p<0.001$) with the interference items mood, relations and enjoyment of life as dependent variables.

5.2 Paper II

Development of EirV3: A Computer-Based Tool for Patient-Reported Outcome Measures in Cancer

Eir-Patient has a hierarchical structure starting with dichotomous assessments (yes/no) regarding the presence of 19 initial symptoms frequently experienced by cancer patients. It includes the 12 symptoms in the EAPC Basic Dataset [103] (i.e. pain, tiredness, drowsiness, nausea, reduced appetite, breathlessness, depression, anxiety, well-being, sleep, constipation, vomiting), supplemented by four items of particular relevance for patients undergoing chemotherapy based on the Common Terminology Criteria for Adverse Events (CTCAE) [166] (i.e. numbness in hands or feet, diarrhea, mouth sores, dry mouth). Another four items adapted

from the Patient-Generated Subjective Global Assessment (PG-SGA) [163] for assessment of nutritional status (i.e. altered sense of taste, altered sense of smell, problems swallowing, early satiety) were also included, as well a question on physical activity.

The first question in Eir-Patient resembles the usual start of a real clinical situation asking a general question about the patient's wellbeing. Then there is a symptom screening section organized into three levels:

- Level 0: Symptom screening (figure 3)
- Level 1: Symptom intensity (figure 4)
- Level 2: Symptom characterization (figure 5)

Level 0 is an initial screening for symptoms occurrence, where patients tick the symptoms that they have experienced in the past week from a predefined list on the screen. Patients are then routed to level 1 (symptom intensity) where they rate the intensity of those symptoms marked at Level 0. The intensity is assessed on a 0–10 numerical rating scale (NRS-11) where 0= "*no symptom"* and 10= "*worst possible symptom"*, in line with expert recommendations [162]. In Level 1, the symptoms are presented to the patient one at a time. Patients who score a symptom above a predefined threshold (≥1) on the intensity score in Level 1, are presented with additional questions exploring that symptom in more detail for the following symptoms included in Eir-Patient: pain, breathlessness, depression, anxiety, insomnia, constipation, vomiting and diarrhoea (Level 2). For symptoms scored 0, no level 2 follow-up questions are presented. Consequently, patients with more symptoms get more questions. Two final sections regarding nutritional intake are filled in by all users, on height, weight and food intake and a performance status section with questions about ability to perform physical activities.

PROs reported by the patient on the tablet are immediately available in the Eir-Doctor module (figure 6). This makes it possible for the physician to prepare a subsequent consultation and guide the communication in the consultation by focusing on the most bothersome symptoms and problems reported by the patient. The endorsed symptoms are displayed with scores in descending order of intensity. Scores ≥ 3 are marked with pink, indicating clinical relevance (figure 4), while scores ≥ 7 are marked with darker red. Eir-Doctor also includes a graph displaying the symptom intensity over time, and detailed symptom information (i.e. patient´s answers to follow-up questions).

Overall, usability results showed that patients and health care providers found EirV3 to be intuitive, easy to use and relevant. The patients appreciated that the physician received updated information about their clinical condition. Some patients with related symptoms (e.g. tiredness, lack of appetite, depression) found some of the follow-up questions to be overlapping. Using EirV3 was more demanding for patients with advanced cancer and a high symptom burden compared to patients with less symptoms.

When testing Eir-Doctor, physicians defined the graphical presentation of symptom trajectories as a key factor to monitor effects of treatments. They also reported that they became aware of symptoms that they had not known troubled the patient.

Comparative study

Of the 114 included patients, 110 patients (97%) completed both the electronic and paper versions, 54% on tablets first and 46% on paper first. When comparing PROM assessment on paper versus tablets (N=114), the median intraclass ICC was 0.815, ≥0.75 for 13 items. Overall, 41% of the patients preferred assessment on tablets, 19% preferred paper while 40% had no preference.

Figure 3: Eir-Patient, level 0: Symptom screening

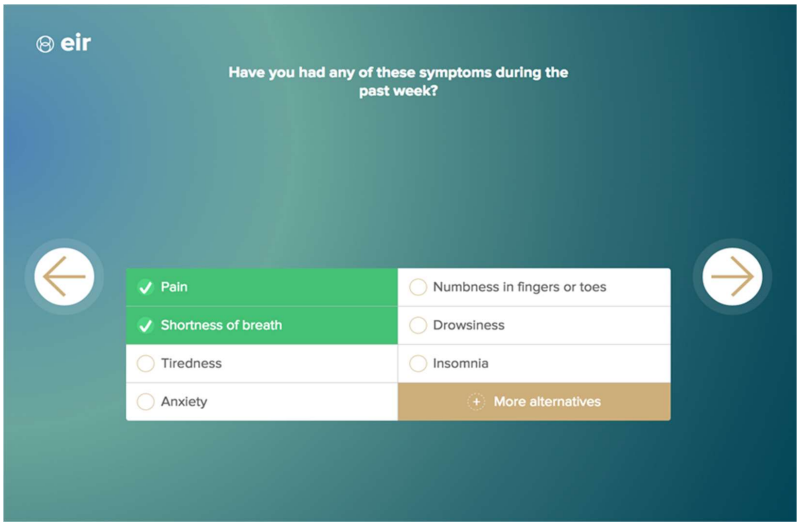


Figure 4: Eir-Patient, level 1: Symptom intensity

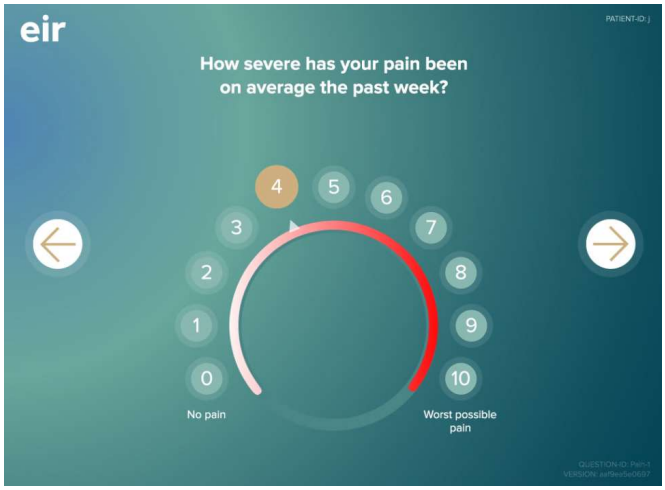
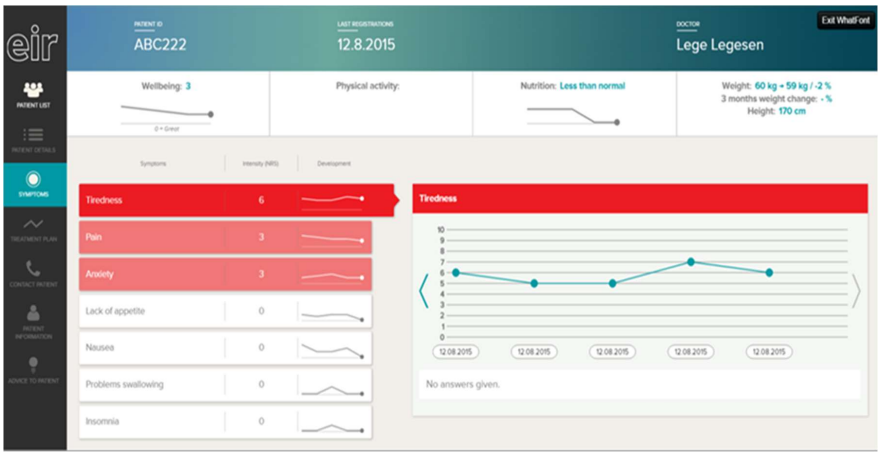


Figure 5: Eir-Patient, level 2: Symptom characterization

*Figure 6: Eir-Doctor: The patient`s present symptom intensity to the left, and a graphical presentation of symptom intensity to the right**



*this is not data from a real patient

5.3 Paper III

Usability testing of EirV3 – a computer-based tool for patient-reported outcome measures in cancer

Thirty-seven cancer patients were enrolled. Mean age was 64 years (SD 11.3) for patients and median Karnofsky score was 80 (range 50–100). Fifty-one percent of the patients used tablets daily, while 35% had never used a tablet prior to the study. Seventeen physicians were enrolled. Five worked as GPs and 12 as oncologists. Mean age was 48 years (SD 11.7) for the physicians.

A total of 73 Eir registrations were completed by patients in different settings, at the outpatient clinics, at home and at the participating GPs' offices. The physicians used EirV3 in a total of 59 consultations. No technological difficulties appeared in any of the settings. All patients were able to complete the Eir-Patient symptom registration. In total, 72 usability issues were identified in Eir-Patient and Eir-Doctor. None were graded as unusable. For Eir-Patient, 62% of the identified usability issues were graded as irritant (grade 1), 18% as moderate (grade 2), and 20% as severe (grade 3). An example of usability issue graded as moderate was missing questions on urinary problems. Examples of usability issues grade 3 are described in table 3. For Eir-Doctor, 46% of the identified usability issues were graded as irritant, 36% as moderate and 18% as severe.

Based on the identified usability issues, EirV3 has been improved. Summarized, the most important changes were 1) making the questions in the symptom screening section in Eir-Patient mandatory, 2) adding text to describe anchors, e.g. 0=*"great"* and 10=*"worst imaginable",* and 3) improving the accuracy of the pain body map in Eir-Doctor, to prevent overlap of marked pain areas.

Table 3: Examples of usability issues graded at level 3 (severe) in Eir-Patient, and possible resolutions

Usability issues	**Quote/observation**	**Resolution/suggestion**
In the well-being question it was confusing that 0 equals best well-being, while 10 equals worst	"In my head it just gets a bit confusing to read this. It should be the opposite"	Resolution: Added text to number: 0= great, 10= worst imaginable well-being
A risk that patients unintentionally skipped one of the three symptom screening pages if they double-clicked the screen	One patient, who had pain, had not marked pain: "Were there any questions about pain then?"	Resolution: Patients must either mark one or more symptoms, or tick off "neither of these"
Problems with the touch screen not responding to the patients' taps due to wrong technique	"I have to use my fingers, not the stylus. There you see. How could I have used it at home, it's impossible"	Resolution: use a stylus of high quality, and spend more time on instructions and training for unexperienced tablet-users

6. Discussion

6.1 Discussion of main findings

A crucial element of patient-centered care is to ask the patients systematically about their symptoms, functions and preferences for care. Today we have electronic patient records, thus it should be obvious to perform electronic assessment of patient-reported outcomes. In this thesis, I have been a part of a team that have developed an e-PROM (EirV3). In this context, I have investigated patient preferences for paper versus electronic assessments and tested the usability of EirV3 in different settings. These studies have contributed to improve the relevance and usability of Eir and show that electronic PRO assessments are well perceived. These results led to a continuous process to optimize the usability of e-PROMs in clinical practice. Clinicians and researchers often request reference data to facilitate interpretation of PRO scores. Reference data from the general Norwegian adult population can be used as reference against which patient scores can be compared for interpretative reasons. E.g. if the patient's score is significantly higher or lower than found in the general population after adjusting for known variables that affect the outcomes, follow-up of potential disease or side effects of treatment may be indicated. In total, these studies may contribute to improved clinical practice and an increased patient-centered focus. The current Eir version is designed for use in both clinical practice and in research.

In paper I, Norwegian reference values for the MDASI from the general population were presented. The most frequently reported symptoms were fatigue (35%), pain (34%) and drowsiness (27%), when using a cut off ≥3. The mean scores for fatigue were highest in the youngest age group (18-29 years). In this respect, these findings may indicate an unhealthy bias among the youngest participants due to the low response rate in this age group. Our findings suggest that a higher symptom score is dependent on the presence of one or more comorbidities, higher levels of depressive symptoms and lower level of education. Furthermore, our findings suggest that a high symptom score and increased levels of depressive symptoms interfere with functional status.

In paper II, the rationale behind the development and processes towards EirV3 was presented. Eir has been developed as a result of long-term, iterative development process with regular end-user testing, in multi-professional workshops and meetings both nationally and internationally. The end-users found that EirV3 was a dynamic, user-friendly tool for symptom assessment in clinical cancer care. Many patients preferred electronic PRO assessments (41%) or had no preference (40%) compared to the paper-and pencil method.

In paper III, further testing of EirV3 led to improvements based on the identified usability issues. Findings suggest that EirV3 is usable by a heterogeneous population of cancer patients and physicians in different settings. Overall, EirV3 was found easy to use in multiple settings and the content was perceived relevant across cancer diagnoses, and treatment intent, e.g. curative vs palliative. The automatic ranking of symptom intensity was actively used in all consultations.

6.1.1 Paper I

Symptoms in the general population

Here we collected the first reference data for the MDASI in the general Norwegian adult population. Reference values are useful for comparing diseased against healthy samples, thereby facilitate the interpretation of patient scores. The MDASI is developed to capture the presence and severity of a patient's symptoms, supplemented with items assessing how they affect functioning as perceived by the patients. General population studies suggest that a high number of symptoms have been associated with poor self-reported health status and increased use of health care [65, 172], as well as lower functional ability [14, 23, 65, 173]. As such, symptoms are an important aspect of the overall health status in the general population. Studies have shown that women generally report a higher number of symptoms than men [13, 15, 66, 174]. Hjermstad et al. documented that the reduction in HRQoL scores in disease groups compared with values from the general population was smaller when adjusting for age and gender [71]. In paper I, we found that increasing symptom scores were dependent on the presence of one or more comorbidities, higher levels of depressive symptoms and lower level of education. This finding is in line with another Norwegian population study that found a statistically significant relationship between anxiety, depression and functional symptoms [66]. This means that there is no golden standard for a given symptom score, as there is for most laboratory results as an example. It is important to adjust for variables that significantly affect the symptom level. Although a cross-sectional study does not allow for investigation on causations, this study indicates that symptom scores are influenced by several covariates. By controlling for relevant associated factors, potential biases are likely to be reduced. Such adjustments apply to comparisons between patient samples with similar age, gender and other background variables.

Use of population data as reference values

The use of PROs as relevant outcomes in clinical research and as part of the decision making in the clinical encounter is increasing [129]. As such, there has been increasing interest in the broader interpretation of PRO scores for purposes of comparison across studies and populations. A review of electronic PROMs in cancer care [25] found that inclusion of reference values is used in approximately 50% of published reports. By using reference values to compare PROs between groups by age, gender, cancer stage and other relevant covariates, more meaningful information about symptoms may be provided, and as such assist health care providers in identifying and monitoring symptoms [75]. With cancer survivors as an example, the reference values might be used to investigate if the PRO scores reported by cancer survivors are above or below those of the general population when adjusted for age, gender and other relevant covariates. Reference values may also indicate whether they return to their normal physical function level by comparing them to their own "baseline" scores, or with other patients with the same age and diagnosis after end of cancer treatment. As such, reference values are particularly relevant in studies on cancer survivorship. In a Norwegian cross-sectional study of young adult cancer survivors, 25% reported chronic fatigue assessed by the FQ, at a median of 14 years from diagnosis [175]. Chronic fatigue was associated with systemic treatment in combination with surgery and/or radiotherapy, comorbidity, pain, numbness and depressive symptoms. Chronic fatigue was most prevalent among survivors of breast (29%) and colorectal cancer (29%) and in non-Hodgkin lymphoma (27%). Survivors of localized malignant melanoma treated with localized surgery had the lowest prevalence of chronic fatigue (15%) [175]. Reference values may increase knowledge of trajectory of fatigue and other symptoms after different cancer treatment, and whether the symptom prevalence is increased compared to the general population with the same background characteristics.

6.1.2 Paper II and III

Patient-centered care

Patient-centered care in oncology is an approach that includes focusing on those elements of disease and treatment that matter most to the patient. Studies have shown that systematic electronic PRO assessments may have clinical benefits, including more frequent discussion between the patients and HCPs discussions about symptoms in the clinical consultations [37, 105] and improved symptom management in response to patient reports [105, 176, 177]. In contrast, undetected symptoms may continue to worsen the symptom burden, thereby leading to more serious complications and functional impairment. Therefore, systematic assessment

of PROs enhances physician's awareness to intensify management of symptoms as an important component of patient-centered care. Systematic use of EirV3 gives the patient the opportunity to report information directly to the physician's computer which may direct the discussion of the subsequent consultation to issues that are most bothersome for the patient. This may lead to changes in treatment based on the patient's needs. Moreover, the graphical presentation of symptom development over time might be valuable to evaluate the effect of interventions.

Development of e-PROMs

Symptom assessment by computers is only effective if it provides valid results, is perceived as useful, if they are implemented in daily routine and is the preferred assessment method by patients and health care providers [26]. The ideal symptom assessment tool, regardless of format, should include the symptoms that occur most frequently and are the most distressing for patients, but it should also be short, easy to understand, and applicable to both clinical and research settings [23]. The content in EirV3 covers common cancer-related symptoms. The clinical relevance of the symptoms may be discussed. Theoretically however, the relevance of symptoms in the electronic format should not differ from the ones on well-validated paper tools. Also, it is reason to believe that the algorithms provided by an electronic format increases the relevance by tailoring questions based on individual scores. In Eir, patients' symptoms with individually tailored in-depth questions are presented on the physician´s computer, supplemented with graphs for symptom development over time. These properties are not achievable in the paper-and-pencil format. As such, the content of Eir does by no means represent a paper tool that is pasted into an electronic format, as unfortunately is often seen [25, 105, 178]. This is an important distinction. For example, on paper, all questions must be completed by the patient. In Eir-Patient, the questions in the symptom screening section was made mandatory to avoid missing items. Further, if the patients do not report pain over a specified, low cut-off score in Eir, they get no further questions on pain. Tailored questions make the assessment more targeted and probably less cumbersome for the patient. However, using EirV3 is more demanding for patients with a high symptom burden as an increased number of symptoms trigger several follow up questions

Usability testing of EirV3 in different settings identified a total of 72 usability issues, however, most of them were classified as irritant or moderate. Not all identified usability issues were resolved during the test period due to hardware, software, funding or resource

constraints. This is also reported in similar studies [179, 180]. For example, EirV3 has not been integrated into electronic medical records yet, due to resource and funding constraints and confidentiality issues. Also, a new digital health platform including electronic records is currently under development in our health region, so integration into the existing electronic journal was not prioritized, even if we knew this was not optimal. It may be inconvenient for the physician to shift attention from EirV3 to other programs like the medical record, imaging, laboratory tests etc. in separate programs. Separate internet addresses and passwords may represent a barrier to the use of EirV3. On the other side, accessibility must not compromise patient confidentiality and security issues regarding registration, transfer, handling and storage of data which are major important issues.

An important finding from the usability testing was that patients wanted access to see their own assessments and specifically so, the graph that showed the development of PROs over time. However, in this study, we were more concerned about privacy issues and we were not allowed to store the data on the tablets. Consequently, no previous entered results were available for the patient the next time he/she was asked to use Eir. During initial testing of Eir, the patients were only presented the Eir-Patient (and not Eir-Doctor), and they used Eir only once. They were thus not given the opportunity to reflect of their need to see the visualization of their own data in Eir-Doctor. In later phases the patients often looked at the screen together with the physician during the consultation. When they became aware of the storage and presentations of their own answers, they said they preferred to have access to their own data, for instance to follow their own symptom development over time. This illustrates that the patients want to be involved in decisions regarding their own health. The physicians defined graphical presentation of symptom development over time as a key factor to evaluate the effect of treatment. Physicians found it useful to start the consultations with the list of symptoms ordered by intensity scores (figure 6). They also appreciated that fact that using Eir made them aware of symptoms they did not know had troubled the patient. In this way, electronic PRO assessments with Eir made the communication easier and led to more focused consultations by addressing issues that the patients have reported as most troublesome.

In a single-center randomized, controlled trial, Basch et al. [33, 40] randomly assigned patients receiving chemotherapy for metastatic solid tumors to self-report 12 common symptoms via tablet computers, or to usual care. Patients in the e-PRO group reported improved quality of life, were less frequently admitted to the emergency room or

hospitalized, and remained on chemotherapy longer [33, 40]. Results were even suggestive of prolonged survival [40, 181]. A multicenter randomized clinical trial by Denis et al. [39] compared web-based reporting of PROs (experimental arm) vs. routine follow-up with CT imaging (control arm) to detect symptomatic recurrence in lung cancer patients following initial treatment. The median overall survival (OS) was 19.0 months in the experimental arm and 12.0 months in the control arm. In the PRO group, the patients were invited to complete weekly self-report of 11 common symptoms and weight online between visits. Symptom severity was graded from 0 (no symptom) to 3 (major symptoms) for appetite loss, fatigue, pain, cough, depression and breathlessness. The five other symptoms were specific for lung cancer and included fever, face swelling, lump under the skin, voice changing and blood in sputum and were assessed by yes or no answers. Alerts were automatically sent to the oncologist when predefined criteria were present. A graphical presentation showed scores over time with major symptoms (score = 3) marked in red. The same principle is used in EirV3. Here, a score ≥ 3 on a 0-10 NRS is pink while scores ≥7 are red, both indicating a clinically significant symptom, calling for attention. EirV3 is designed for use in a heterogeneous population of cancer patients, in contrast to this randomized trial where five of the symptoms were specific for lung cancer progressive disease. Electronic PRO systems should ideally integrate both treatment-and patient-centered perspectives [25]. Future development of Eir might include disease-specific modules to be able to detect symptomatic recurrence of tumor following treatment.

As the primary aim of symptom assessment is to improve symptom management, the cut-off levels in Eir decides if the symptom is followed by an in-depth question. If the cut off level is too high, there is a risk of suboptimal treatment because the problem goes unnoticed. If the cut-off is too low, patients will be exposed to follow-up questions that may seem irrelevant. Whether a change of one point on a scale from 0 to 10 is of clinical significance is not a statistical question. The clinician must understand the clinical importance of the measure to regard it as "clinically significant"[45], and also talk to the patient to clarify the significance of the symptom severity.

Traditionally, scores on the EORTC QLQ-C30 are transformed from the 4-point response scales (i.e. *"not at all", "a little", "quite a bit", and "very much")* to 0-100 scales in studies and in clinical practice, as are the EORTC reference values [182]. However, one suggested approach to define clinical thresholds for the different symptoms and problems is simply to use the responses on the 4-point scales. Then, a patient is classified as having a clinically

important problem if he or she responds with at least “a little” for any given item [183, 184]. Another approach is based on score distributions and makes use of statistics from reference populations, most often the general population. For example, the general population mean may be used as a threshold as was done with data from two prospective Nordic Myeloma Study Group trials using the EORTC QLQ-C30 questionnaire [185]. The results were compared with the scores of an age and gender adjusted Norwegian reference population (n=3000). Findings suggested that the most distressing problems were pain and fatigue, reduced physical and role functioning and reduced overall QoL. These differences from the reference population scores were statistically significant ($p<0.001$). Comparison with a reference population facilitates the interpretation of QoL and prevents overestimation of symptoms and underestimation of subjective treatment response [185]. Giesinger et al. [183] estimated thresholds for clinical importance for four EORTC QLQ-C30 scales. Patients who rated their symptom as “quite a bit” or “very much” for any anchor item were classified as having a problem of “clinical importance”. Using the definition of “clinical importance”, reflecting a higher degree of burden, prevalence rates were 41.7% for physical functioning, 39.2% for fatigue, 28.0% for emotional functioning, and 24.1% for pain. [183].

Assessment of symptoms may lead to better symptom management by identification of clinically important problems, but also because the patient preferences are taken into account in a more systematic way. A high score does not automatically indicate that this is what the patient perceives as most bothersome. In paper I, we found that the prevalence of pain in the adult Norwegian population corresponded to results from the Health Study of Nord-Trøndelag County (HUNT 3) [67] when using a cut off ≥3. The cut-off for follow-up questions on specific symptoms in Eir-Patient is >1 and was purposefully set to be this low to avoid overlooking symptoms. Cut-offs for high values are important, and based on common clinical practice, systematic reviews [186] and clinical studies [105], values ≥ 3 were flagged to alert physicians. Use of reference data may improve measurement precision by relating individual patient scores to reference data adjusted for associated variables.

Patient preferences

The comparative study that examined the equivalence between electronic and paper-based PROMs, showed excellent agreement between the two methods. The majority of patients (81%) preferred to respond on the tablets or had no preference. In a large, international study (N=965) comparing symptom assessment electronically vs on paper, 52% of patients

responded that they would have preferred electronic assessment and one third had no preference, even though most patients had little experience with computers [26]. In our study, preference for electronic assessment was more frequent among patients with previous digital experience. In Norway, at least 98% of all households has access to internet [187]. Over time, most people will be familiar with use of computers and other electronic devices. However, the mean symptom intensity scores in the comparative study were low, which may indicate that more patients had a good performance status, and as such not may not have been representative of cancer patients at all stages of the disease trajectory. In future studies, purposive sampling must be considered to examine use in frailer patients.

6.2 Methodological considerations

Methodological considerations include evaluation of internal and external validity. Internal validity concerns whether data are collected, analyzed and interpreted without bias [188]. External validity concerns whether the results from the study are generalizable to other subjects than those included in the study sample [188].

6.2.1 Study design

Cross-sectional design

A cross-sectional study design in the form of a survey was chosen to collect data on the symptom prevalence in the general population in study I. This design is appropriate to examine prevalence rates and makes it possible to investigate associations between variables [188]. A cross-sectional design was considered appropriate according to the aim of study I, which was to investigate the prevalence of symptoms and associated factors in the general Norwegian adult population. Advantages of the cross-sectional design are that a study may be conducted relatively fast and is less expensive and time-consuming than other designs. However, cross-sectional designs do not make it possible to distinguish whether the exposure preceded or followed the disease or the event of interest, and thus cause and effect relationships are not certain. Potential associations must be interpreted carefully but raises hypothesis that may be investigated in further research [188]. In this thesis, we found significant associations between the dependent and independent variables, but no conclusion about cause and effect can be drawn. For example, we found a statistically significant association between increased levels of depressive symptoms and higher MDASI sum scores, but whether increased MDASI sum score may cause increased levels of depressive symptoms

or vice versa remains unknown. A prospective design may answer the question of possible bidirectional associations of MDASI sum score and PHQ score. Also, the use of self-report to measure both depressive symptoms and MDASI sum score may have caused overlap between the variables. However, this problem may have been reduced by excluding the four somatic symptoms of PHQ-9, as is also done in other studies [189, 190].

When it comes to the use of paper-based questionnaires in this study, an electronic system for nationwide surveys according to our preferences was not available at a reasonable cost at the time. Also, we have good experience with previous paper-based surveys [55].

Randomized comparative design

Equivalence testing is designed to evaluate the comparability of data obtained via the original and adapted administration mode [191]. In study II, PROM scores from electronic administration and paper-and pencil administration were compared, with the intention to ensure that the scores from the electronic questionnaire do not vary from scores on a paper questionnaire [191]. Patients were randomly assigned to complete PROMs either on paper or tablet for the first administration, and then the other mode for the second administration. This study design is recommended by the ISPOR ePRO Task Force [191]. Moreover, to minimize carryover effects from the first testing, adequate time should be allowed between administrations. In our study, the patients waited at least 30 minutes between each testing without performing any interventions in the meantime. A within-patient design provides greater statistical power and decreases sample size requirements, and intraclass correlation coefficient is useful to measure equivalence [191]. According to interpretation guidelines [171], an ICC > 0.75 indicates excellent agreement. A randomized parallel group design could have been considered. In this design, patients in one study arm completes PROMs on tablets while the other arm complete the paper version. A comparative design was chosen because the primary outcome was to assess patient preferences for paper versus electronic assessments.

User involvement

User-based testing involves observation of end-users to evaluate the ease of navigation, interaction with application features, ability to perform essential functions, and satisfaction with task flow, and guide further development [192, 193]. The involvement of several multi-professional experts both nationally and internationally as well as usability testing by the end-users were an indispensable resource in the development of Eir, as emphasized by others

[194]. This is first and foremost related to the importance of developing a program that has clinical utility, is evidence-based and perceived as relevant and user-friendly for those who will be using it. Our research group in the European Association of Palliative Care (EAPC) Research Network [195] and the European Palliative Care Research Centre (PRC) [196] has developed several electronic symptom assessment tools over the past decades [26, 138-140], and experiences from these projects constitute the basis for the choice of method in this project.

The Computer Based Assessment and Treatment (COMBAT) study performed by our research group, aimed to evaluate the impact of a computerized clinical decision support tool on pain management in cancer out-patients [139]. The study was designed as a prospective, controlled study comparing pain intensity and opioid dose in two different patient cohorts before (N=103) and after (N=153) implementation of COMBAT. Results showed that the COMBAT intervention did not improve the management of pain in cancer patients. Possible reasons might be that the software was not integrated within the electronic medical records that made it cumbersome to use. Further, the experiences when using COMBAT as perceived by the end-users were not examined. Based on previous experiences, regular usability testing, by means of a combination of observations and semi-structured interviews, was the chosen method during the entire Eir development process. We believe this may provide a better understanding of the end-users´ experiences, barriers and perspectives [197], to ensure a comprehensive understanding of the end-users' experiences of using Eir. By focusing on ease of use and clinically relevant issues for both clinicians and patients, the intention was to reduce the barrier for systematic symptom assessment in daily clinical practice. Lack of assessment is an important reason for inadequate symptom management. However, assessment per se does not necessarily improve symptom management, the PROs must be evaluated and discussed with the patients. Hence, a crucial feature of e-PROMs is the immediate transfer of results to the clinician, preferable as part of the electronic patient record. Whether EirV3 improves symptom management is an empirical question and must be examined specifically as done in the COMBAT study.

Involvement of patients in the development may increase their understanding and motivation for using the tool. Patients and clinicians have been invited to test Eir and give their feedback from the initial ideas on paper and post it notes, through the first digital version of Eir-Patient with each question written in speech bubbles as asked by a nurse avatar, to the so far final version with a refined and modern user interface and a dynamic item order. In the

development of Eir, patients had some impact on the wording of the questions, but not regarding which symptoms should be included. Navigation and layout, however, were to a large extent based on users´ preferences and adjusted to users´ skills. In fact, involvement of end-users is not as common as one might think, and definitely not through the entire process. For example, some systems are developed without involving end-users. An internet-delivered program for cognitive behavioral therapy for people with depression and anxiety was developed by the Scarborough Hospital in Canada [198]. Six months after the clinical implementation of the program, the dropout rate was as high as 90%. To understand why the dropout rate was so high, the authors collected feedback from the patients. Results showed that the patients were not satisfied with the solution as the content was confusing and difficult to understand, and technical difficulties occurred. Thus, patients were included in the further development of the tool, resulting in a dropout rate of 33% percent in the next round. Involvement of end-users may result in reduced dropout, but is probably still not sufficient to solve the dropout challenges alone [199].

It was considered an advantage that the initiative to develop Eir came from health care personnel and researchers, and not from technology companies. A different approach than the one chosen for developing Eir, is to start the development process by identifying patients´ needs and wishes. Two reasons why a symptom assessment tool cannot be entirely based on patients´ preferences are that health care professionals are at least as important users as patients, and that the content must be determined based on what is clinically relevant and based on evidence and consensus. As such, there are at least three needs to consider: 1) Eir or any other PROM, e-PROM or paper tool, for that matter, must be clinically relevant and perceived useful by patients and clinicians, 2) symptom assessment must be done in ways that are based on evidence and consensus, and 3) users´ need for ease of use.

The usability testing in the Eir development process was entirety qualitative in design. In the study presented in paper III, our intention was to include the System Usability Scale (SUS) to combine qualitative and quantitative methods [200]. SUS is a reliable, valid tool for measuring the usability and is often used to evaluate software, websites etc. as it can effectively differentiate between usable and unusable systems. It consists of a 10-item questionnaire with five response options for respondents, from strongly agree to strongly disagree. However, the scoring system is somewhat complicated as the scoring of the response alternatives are reversed for every other question, and the frailest patients found the form difficult to understand. Results from qualitative studies supplement results of

quantitative studies as they may provide a more profound understanding of the meaning and implications of the subject under study, using the patients' own words and descriptions [197]. SUS might have provided additional information about usability, but as it represented an additional burden for the patients and a possible source of bias, it was decided to discontinue use of SUS in this study.

6.2.2 Selection bias

Selection bias occurs if study participants are systematically different from the population of interest [188]. In paper I, the randomly drawn sample was assumed to be representative of the general Norwegian adult population with respect to age, gender and place of living. However, only 36% of the sample responded to the survey. Compared to similar surveys in Norway in 1996 and 2002, this response rate was low [47]. Unfortunately, no information on non-responders in study I was available. Reference values for the SF-36 has previously been collected in Norway, with a response rate of 67% in 1996 and 36% in 2015. Participation rates for epidemiologic studies have declined steadily over the past decades [201], although there is substantial variability in participation rates between studies. The response rate in a Norwegian population-based study assessing self-reported symptoms on a paper-based questionnaire in 2004 was 54% [15].The response rates in the population-based health surveys in Nord-Trøndelag county were 71% in the HUNT2 in 1997 and 54% in HUNT3 in 2008. Another Norwegian study found that HRQoL measured with EORTC QLQ-C30 was relatively stable in two cross-sectional studies over an eight year period despite the much lower response rate in the second study, 68% versus 35% [202]. A large Danish population study from 2015 [64] used web-based questionnaires to estimate the prevalence of self-reported symptoms and got a response rate of 52%. Balter et al. compared the use of web-questionnaires with a similar printed questionnaire in a population-based study [203]. They found that the initial response rate was lower for the web-based questionnaire than for the printed questionnaire. However, the willingness to answer a second questionnaire was higher when using a web-questionnaire instead of a printed questionnaire. An electronic questionnaire might have increased the response rate, especially among the youngest age group. Increased attention about the study in the general population could have been achieved by promoting the study in newspapers and social media. Also, sending more than one reminder to non-responders could have increased the response rate. In study I, the high mean symptom scores for the youngest age group are probably not representative for the general population as the response rate in this age group was low. However, the opposite pattern was

seen in the oldest age group. The relatively high symptom scores in the youngest age group compared to the older may indicate an unhealthy bias among the youngest and a healthy bias in the older age groups.

The Cross-sectional Cambridge Centre for Ageing and Neuroscience study was developed to examine associations between epidemiological and cognitive neuroscience data across the adult lifespan [204-206]. The researchers investigated the pattern of response at different adult ages within the cohort and found an association between age and participation. Individuals in middle age groups were more likely to participate. The highest participation was in the 58–67 years age group. The main reasons for active refusals were time constraints. Overall, there was no difference between men and women. In the younger age groups, women were more likely to participate than men, while the opposite was seen in older age groups [204]. This is in line with the responses in our study, where the response rate is highest in the older age groups. In 2015, 15% of the Norwegian population was 67 years or above. In our study, 27% of the responders were in the same age group. About 21% of the population was between 18 and 29 years, while only 5% of this age group responded. For the older population, the opposite pattern was seen. This suggest that the reference values might be biased due to the characteristics of the participants in the youngest age group. Also, the fact that a large proportion of the sample had university level education may reduce the representativity. According to Statistics Norway 32% of the Norwegian population had a university level education in 2015, 41% had completed high school and 27% had finished elementary school, relative to 46%, 37% and 17% in paper 1. This should be considered when performing comparisons for subjects with low education. The response rate was slightly higher in women. Findings from other studies have shown that women are more willing to participate than men [207, 208]. There are probably several reasons for the growing refusal to participate. It has been shown that the more a study requests of a potential participant, the more likely he or she is to decline participation [207]. The increasing almost incessant number of requests, particularly so in marketing surveys may have the effect that persons refuse to participate [209].

Bias undermines the validity of research. All observational studies have bias, and the challenge is to identify the biases and evaluate how they might have influenced the results, and how to handle this in the analyses and interpretation of results [210]. The major concern about study nonparticipation is whether it introduces nonresponse bias, as non-respondents may differ from the responders on the main factors of interest. The non-response rate

produces its effects through the difference between non-respondents and respondents to the survey [209].

6.2.3 Psychometric considerations

Information bias occurs from errors related to collection and/or measurement of data [188]. In order to obtain valid results from PROMs, the scales must demonstrate good measurement properties, i.e. validity and reliability (table 1).

The MDASI measures the severity of cancer-related symptoms and their impact on function. Twenty-six symptoms and six interference items were rated by a validation sample of 527 outpatients, 30 inpatients, and a cross-validation sample of 113 outpatients [23]. Clinical judgement and cluster analysis were used to eliminate similar items. Validation demonstrated that the items account for the majority of distress in patients with different malignancies at various stages and that these items are sensitive to expected differences in symptoms and side effects. Internal consistency (reliability) was examined by calculating the coefficient alpha values for both the validation and the cross-validation samples. The values were ≥ 0.82 for the validation sample and ≥ 0.87 for the cross-validation sample indicating a high level of reliability. To examine the sensitivity of severity of disease, patients were divided into two groups based on ECOG performance status. There was a significant difference in mean symptom severity (2.36 vs. 3.62, $p<0.001$) and mean symptom interference (2.95 vs.5.31, $p<0.001$), between patients with a good performance status and those with a poor performance status [23].

The MDASI has been translated and psychometrically validated in several languages [23, 211, 212]. A full validation study as is common when developing or translating a questionnaire for use in disease specific populations was not performed. The common approach is to translate the questionnaire according to well-established, consensus-based guidelines, and to perform a pilot study with debriefing questions regarding relevance, understanding etc. The MDASI was translated into Norwegian in 2011 according to the 2009 procedures developed by the EORTC Quality of Life Group [147]. This did not lead to any changes in wording of the questions, and most subjects perceived the content as highly relevant. A full psychometric validation of the MDASI in a Norwegian cancer population is probably a next step, for example along with commonly used questionnaires like the QLQ-C30 and the BPI to test whether the items are corresponding. On the other hand, given that

the MDASI symptoms are common among cancer patients, it is reasonable to assume that the Norwegian MDSAI has high content and construct validity.

Specific precautions were taken in the Eir development process to ensure validity and reliability, and to develop a user-friendly tool that is not too comprehensive. An underlying premise was to assess symptoms as should be done in clinical consultations by focusing on the common symptoms and particularly so those perceived as most bothersome by the individual patient. Further, the content was based on well-validated PROMs. Relevant and validated symptom assessment tools for the choice of specific items were identified and presented to researchers and clinicians to reach consensus regarding relevance and importance. Overall, the validity of EirV3 is considered satisfactory as the content is based on reviews and consensus by experts, and perceived relevant by end-users. Parallel forms reliability was measured in the comparative study. Median intraclass correlation coefficient was 0.82, indicating high parallel forms reliability between symptom assessment on tablets and on paper.

The internal consistency (e.g. Cronbachs alpha) of Eir has not been tested. This can be viewed as a criticism or a limitation. On the other hand, there is little reason to believe that the internal consistency of scales differs largely by changing from paper to computer. Eir is primarily designed for symptom assessment and use in clinical consultations. However, testing internal consistency should be considered in further validation of Eir to ensure that observed variance in the measurements can be attributed to real differences in scores.

6.2.4 External validity

External validity is dependent on internal validity, and concerns whether the study results are generalizable to other populations [188]. In this thesis, external validity in paper I refers to whether the results can be generalized to the general Norwegian adult population, and for paper II and III, it refers to whether results can be deemed valid for cancer patients in different phases of the disease trajectory, and to physicians.

In study I, the subjects were drawn by a professional agency Bring [213], to ensure a representative sample with respect to age, sex and geographical spread, according to common procedures for population surveys. The age range of 18–80 was set because the reference values was collected for the adult population. The response rate was highest in the older age groups. About 21% of the population was between 18 and 29 years, while only 5% of this age group responded. For the older population, the opposite pattern was seen. This may suggest

that the reference values might be biased due to the characteristics of the participants in the youngest age group. Evidence suggests that persons with higher socioeconomic status and more education are more likely to participate in studies [208, 214]. In Paper 1, 46% of the respondents had university or university college education which is higher than in the general population (32%). This may be considered a potential bias regarding representativity of the sample and should be considered when using the reference values in subjects with low education.

Cancer patients constitute a very heterogeneous group with respect to age, diagnoses, symptom burden, functional status, prognosis and survival time [103]. In paper II and III, the findings from the usability testing of Eir are applicable in most subgroups of cancer patients. In study II, patients were recruited by convenience sampling due to practical reasons. In study III, recruitment of patients was done by purposive sampling to ensure variation in age, gender, diagnosis and anticipated symptom burden. The included physicians had a variability in age, gender and clinical experience. Nevertheless, the patients and physicians who participated might be more positive to using electronic devices than others. None of the patients were in the youngest age group (18–35 years). However, there is no reason to believe that these patients should find Eir more difficult to use than other patients, given the frequent use of electronic devices in this age group. Eir was only assessed in outpatients, so it is not known whether the usability is comparable in hospitalized patients. Also, participants with visual, auditory, or tactile impairments that might restrict their use of computer hardware were not included in the usability testing. For these participants, assessment will also be demanding on a paper-based questionnaire. In Eir, the number of elements on each screen has been reduced and placed in the middle to make it easier to read the text. Tactile impairments might be overcome by using a stylus to register taps on the screen.

In the comparative study (paper II), the mean symptom intensity scores were low. The median Karnofsky performance score was 90. This indicates that the included patients were in a good general condition. In forthcoming Eir studies, purposive or stratified sampling in further testing of Eir must be considered to examine use in frailer patients. Inclusion of hospitalized patients should also be considered. As the cancer population is very heterogenous it is important that trials report the study characteristics and the study population adequately, to allow the readers to evaluate the generalizability of results and to compare findings with other studies [215].

7. Summary and conclusion

The overall aim of this thesis is to develop a new e-PROM solution; EirV3, built on a better understanding of patient-reported outcomes to improve patient-centered care in clinical practice. To improve interpretation of PROs in clinical encounters and in clinical studies, reference values from the general Norwegian population were collected.

Paper I

1. What is the prevalence of symptoms assessed by the MDASI in the general Norwegian adult population?
   The most frequently reported symptoms were fatigue (35%), pain (34%) and drowsiness (27%), when using a cut off ≥3.
2. Which factors are associated with a high symptom burden?
   The presence of one or more comorbidities, higher level of depressive symptoms and lower level of education were significantly associated with higher MDASI sum score. Background variables must be controlled for when using the reference values.

Paper II

1. What is the optimal content and technical format of an electronic tool for patient-reported outcomes in clinical practice?
   EirV3 has the following two modules: Eir-Patient for registration on tablets and Eir-Doctor for presentation of patient scores in a user-friendly interface. EirV3 was found to be intuitive and easy to use and perceived as relevant for patients and health care providers.
2. What is the rationale for the use of e-PROMs from a technical and patient-centered perspective?
   In the comparative study, the equivalence between electronic and paper assessment of PROs-scores was good. The majority of patients preferred to use tablets or expressed no preference.

Paper III

1. How do patients and health care providers evaluate the content and format of EirV3?
   Seventy-two usability issues were identified. None of them were graded as unusable. Overall, EirV3 was found easy to use in multiple settings and the content was

perceived relevant across cancer diagnoses, and treatment intent, e.g. curative vs palliative. EirV3 has been improved based on the identified usability issues to optimize the usability of using real-time PROMs in clinical practice.

8. Implications for clinical practice and future research

- The first Norwegian reference values for the MDASI have been published and are available for use by clinicians and researchers. Reference values facilitate comparison of symptom scores across populations. However, the findings should be used and interpreted with caution for the youngest age group due to the low response rate in this group.
- Reference values for PROs adjusted for age, cancer stage and type, and other relevant covariates, allow for more tailored interpretation. Incorporation of cancer specific reference values in Eir could help clinicians to better identify and monitor symptoms and should be considered in the future development.
- In Norway, 97% of households have internet access, and the use of smartphones and tablets is increasing. As such, use of electronic tools is feasible for most of the general population. The majority of patients in the comparative study preferred symptom assessment on tablets or had no preference. Considering the many advantages of electronic symptom assessment, this finding is promising.
- Assessment of PROs promote patient-centered care. Integration of PRO data to guidelines for practice and clinical pathways as well as engaging health-care professionals might improve the acceptability and usefulness of routine assessment of PROs. Randomized trials are needed in future research to show the effect of PROs on symptom management. EirV3 is still in development, and is currently implemented into the patient care pathways and clinical practice in a Norwegian cluster randomized trial on early integration of palliative care in oncology [216].
- The fact that EirV3 is not yet integrated into electronic patient records makes it cumbersome to use in clinical consultations. Implementation of EirV3 in daily clinical practice was beyond the scope of this study. Also, additional refinements are necessary before EirV3 becomes a complete, integrated part of the administrative and clinical hospital systems. Future work should address this, to enable a successful integration and use of EirV3 in daily clinical practice. This is related to compatibility with the existing systems, security and confidentiality issues to facilitate its use. This work needs to be a multidisciplinary process to enable successful integration and use.

References

1. GBD Collaborators, *Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017.* Lancet, 2018. **392**(10159): p. 1736-1788.
2. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA: A Cancer Journal for Clinicians, 2018. **68**(6): p. 394-424.
3. Cancer Registry of Norway, *Cancer in Norway 2018- Cancer incidence, mortality, survival and prevalence in Norway.* 2019: Oslo: Cancer Registry of Norway.
4. Cancer Registry, *Cancer in Norway 2017. Cancer incidence, mortality, survival and prevalence in Norway*, I.o.p.-b.c. reserach, Editor. 2017: Oslo.
5. Allemani, C., et al., *Global surveillance of trends in cancer survival 2000-2014 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries.* The Lancet, 2018. **391**(10125): p. 1023-1075.
6. Kaasa, S., et al., *Integration of oncology and palliative care: a Lancet Oncology Commission.* Lancet Oncol, 2018. **19**(11): p. e588-e653.
7. Temel, J.S., et al., *Early palliative care for patients with metastatic non-small-cell lung cancer*. N Engl J Med, 2010. **363**(8): p. 733-42.
8. Zimmermann, C., et al., *Early palliative care for patients with advanced cancer: a cluster-randomised controlled trial.* Lancet, 2014. **383**(9930): p. 1721-30.
9. Gomes, B., *Palliative care: if it makes a difference, why wait?* J Clin Oncol, 2015. **33**(13): p. 1420-1.
10. Hui, D. and E. Bruera, *Integrating palliative care into the trajectory of cancer care.* Nat Rev Clin Oncol, 2016. **13**(3): p. 159-71.
11. Ferrell, B.R., et al., *Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update.* Journal of Clinical Oncology, 2017. **35**(1): p. 96-112.
12. Merriam-Webster, in *Merriam-Webster.com Dictionary*. 2019: Springfield, Massachusetts, USA.
13. McAteer, A., A.M. Elliott, and P.C. Hannaford, *Ascertaining the size of the symptom iceberg in a UK-wide community-based survey.* Br J Gen Pract, 2011. **61**(582): p. e1-11.
14. Bruusgaard, D., et al., *Symptom load and functional status: results from the Ullensaker population study.* BMC Public Health, 2012. **12**: p. 1085.
15. Kjeldsberg, M., et al., *Symptom reporting in a general population in Norway: results from the Ullensaker study.* Scand J Prim Health Care, 2013. **31**(1): p. 36-42.
16. Food and Drug Administration FDA, *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance.* Health Qual Life Outcomes, 2006. **4**: p. 79.
17. Food and Drug Administration. *Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical product Development to Support Labeling Claims*. 2009; Available from: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm193282.pdf.
18. Patrick, D.L., et al., *National Institutes of Health State-of-the-Science Conference Statement: Symptom management in cancer: pain, depression, and fatigue, July 15-17, 2002.* Journal of the National Cancer Institute. Monographs, 2004(32): p. 9-16.
19. Doward, L.C. and S.P. McKenna, *Defining patient-reported outcomes.* Value Health, 2004. **7 Suppl 1**: p. S4-8.

20. Efficace, F., et al., *Health-related quality of life and symptom assessment in clinical research of patients with hematologic malignancies: where are we now and where do we go from here?* Haematologica, 2007. **92**(12): p. 1596-8.
21. Vodicka, E., et al., *Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007–2013).* Contemporary Clinical Trials, 2015. **43**: p. 1-9.
22. Porter, M.E., *What is value in health care?* N Engl J Med, 2010. **363**(26): p. 2477-81.
23. Cleeland, C.S., et al., *Assessing symptom distress in cancer patients: the M.D. Anderson Symptom Inventory.* Cancer, 2000. **89**(7): p. 1634-46.
24. Velikova, G., et al., *The clinical value of quality of life assessment in oncology practice-a qualitative study of patient and physician views.* Psychooncology, 2008. **17**(7): p. 690-8.
25. Jensen, R.E., et al., *Review of electronic patient-reported outcomes systems used in cancer clinical care.* J Oncol Pract, 2014. **10**(4): p. e215-22.
26. Hjermstad, M.J., et al., *Computer-based symptom assessment is feasible in patients with advanced cancer: results from an international multicenter study, the EPCRC-CSA.* J Pain Symptom Manage, 2012. **44**(5): p. 639-54.
27. Benze, G., et al., *PROutine: a feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer.* Ann Palliat Med, 2017.
28. America, I., *Crossing the quality chasm: A new health system for the 21st century. 2001*. National Academies Press.
29. Levit, L.A., et al., *Delivering high-quality cancer care: charting a new course for a system in crisis*. 2013: National Academies Press Washington, DC.
30. Richards, T., A. Coulter, and P. Wicks, *Time to deliver patient centred care.* BMJ 2015. **350**: p. h530.
31. Epstein, R.M. and R.L. Street, *The Values and Value of Patient-Centered Care.* The Annals of Family Medicine, 2011. **9**(2): p. 100.
32. Epstein, R.M. and R.L. Street Jr, *Patient-centered communication in cancer care: promoting healing and reducing suffering.* 2007.
33. Basch, E., et al., *Symptom Monitoring With Patient-Reported Outcomes During Routine Cancer Treatment: A Randomized Controlled Trial.* J Clin Oncol, 2016. **34**(6): p. 557-65.
34. Stacey, D., et al., *Decision aids for people facing health treatment or screening decisions.* Cochrane Database Syst Rev, 2014(1): p. Cd001431.
35. Velikova, G., et al., *Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial.* J Clin Oncol, 2004. **22**(4): p. 714-24.
36. Rose, M. and A. Bezjak, *Logistics of collecting patient-reported outcomes (PROs) in clinical practice: an overview and practical examples.* Qual Life Res, 2009. **18**(1): p. 125-36.
37. Kotronoulas, G., et al., *What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials.* J Clin Oncol, 2014. **32**(14): p. 1480-501.
38. Denis, F., et al., *Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer.* JAMA, 2019. **321**(3): p. 306-307.
39. Denis, F., et al., *Randomized Trial Comparing a Web-Mediated Follow-up With Routine Surveillance in Lung Cancer Patients.* J Natl Cancer Inst, 2017. **109**(9).
40. Basch, E., et al., *Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment.* Jama, 2017. **318**(2): p. 197-198.
41. Strasser, F., et al., *The effect of real-time electronic monitoring of patient-reported symptoms and clinical syndromes in outpatient workflow of medical oncologists: E-MOSAIC, a multicenter cluster-randomized phase III study (SAKK 95/06).* Ann Oncol, 2016. **27**(2): p. 324-32.

42. Au, H.J., et al., *Added value of health-related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG.* Expert Rev Pharmacoecon Outcomes Res, 2010. **10**(2): p. 119-28.
43. Kortteisto, T., et al., *Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care.* BMC health services research, 2012. **12**(1): p. 349.
44. Doward, L.C., A. Gnanasakthy, and M.G. Baker, *Patient reported outcomes: looking beyond the label claim.* Health Qual Life Outcomes, 2010. **8**: p. 89.
45. Hanks, G., et al., *Oxford Textbook of Palliative Medicine*. Oxford Textbook. 2011: OUP Oxford.
46. Ware, J., *The SF-36 health survey*. 2 ed. Quality of Life and Pharmaeconomics in Clinical Trials. 1996, Philadelphia: Lippincott Raven.
47. Jacobsen, E.L., et al., *Norwegian reference values for the Short-Form Health Survey 36: development over time.* Qual Life Res, 2018. **27**(5): p. 1201-1212.
48. Scott, K.M., et al., *SF-36 health survey reliability, validity and norms for New Zealand.* Aust N Z J Public Health, 1999. **23**(4): p. 401-6.
49. Sullivan, M. and J. Karlsson, *The Swedish SF-36 Health Survey III. Evaluation of criterion-based validity: results from normative population.* J Clin Epidemiol, 1998. **51**(11): p. 1105-13.
50. *The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties.* Soc Sci Med, 1998. **46**(12): p. 1569-85.
51. *Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group.* Psychol Med, 1998. **28**(3): p. 551-8.
52. Coulter, A., et al., *Cancer PROMs: a scoping study*. 2015.
53. Aaronson, N.K., et al., *The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology.* J Natl Cancer Inst, 1993. **85**(5): p. 365-76.
54. Nolte, S., et al., *General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the Unites States.* Eur J Cancer, 2019. **107**: p. 153-163.
55. Hjermstad, M.J., et al., *Health-related quality of life in the general Norwegian population assessed by the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire: the QLQ=C30 (+ 3).* J Clin Oncol, 1998. **16**(3): p. 1188-96.
56. Cella, D.F., et al., *The Functional Assessment of Cancer Therapy scale: development and validation of the general measure.* J Clin Oncol, 1993. **11**(3): p. 570-9.
57. Brucker, P.S., et al., *General population and cancer patient norms for the Functional Assessment of Cancer Therapy-General (FACT-G).* Eval Health Prof, 2005. **28**(2): p. 192-211.
58. Zigmond, A.S. and R.P. Snaith, *The hospital anxiety and depression scale.* Acta Psychiatr Scand, 1983. **67**(6): p. 361-70.
59. Spitzer, R.L., K. Kroenke, and J.B. Williams, *Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire.* Jama, 1999. **282**(18): p. 1737-44.
60. Chalder, T., et al., *Development of a fatigue scale.* J Psychosom Res, 1993. **37**(2): p. 147-53.
61. Hjermstad, M.J., et al., *Fatigue in long-term Hodgkin's Disease survivors: a follow-up study.* J Clin Oncol, 2005. **23**(27): p. 6587-95.
62. Orre, I.J., et al., *Chronic cancer-related fatigue in long-term survivors of testicular cancer.* J Psychosom Res, 2008. **64**(4): p. 363-71.
63. Daut, R.L., C.S. Cleeland, and R.C. Flanery, *Development of the Wisconsin Brief Pain Questionnaire to assess pain in cancer and other diseases.* Pain, 1983. **17**(2): p. 197-210.
64. Elnegaard, S., et al., *Self-reported symptoms and healthcare seeking in the general population--exploring "The Symptom Iceberg".* BMC Public Health, 2015. **15**: p. 685.

65. Creed, F.H., et al., *The epidemiology of multiple somatic symptoms.* J Psychosom Res, 2012. **72**(4): p. 311-7.
66. Haug, T.T., A. Mykletun, and A.A. Dahl, *The association between anxiety, depression, and somatic symptoms in a large population: the HUNT-II study.* Psychosom Med, 2004. **66**(6): p. 845-51.
67. Landmark, T., et al., *Chronic pain: One year prevalence and associated characteristics (the HUNT pain study).* Scand J Pain, 2013. **4**(4): p. 182-187.
68. Turner-Bowker, D.M., et al., *Usefulness of the SF-8™ Health Survey for comparing the impact of migraine and other conditions.* Quality of Life Research, 2003. **12**(8): p. 1003-1012.
69. van de Poll-Franse, L.V., et al., *Normative data for the EORTC QLQ-C30 and EORTC-sexuality items in the general Dutch population.* Eur J Cancer, 2011. **47**(5): p. 667-75.
70. Hinz, A., S. Singer, and E. Brahler, *European reference values for the quality of life questionnaire EORTC QLQ-C30: Results of a German investigation and a summarizing analysis of six European general population normative studies.* Acta Oncol, 2014. **53**(7): p. 958-65.
71. Hjermstad, M.J., et al., *Using reference data on quality of life--the importance of adjusting for age and gender, exemplified by the EORTC QLQ-C30 (+3).* Eur J Cancer, 1998. **34**(9): p. 1381-9.
72. Schwarz, R. and A. Hinz, *Reference data for the quality of life questionnaire EORTC QLQ-C30 in the general German population.* Eur J Cancer, 2001. **37**(11): p. 1345-51.
73. Yun, Y.H., et al., *Age, sex, and comorbidities were considered in comparing reference data for health-related quality of life in the general and cancer populations.* J Clin Epidemiol, 2007. **60**(11): p. 1164-75.
74. Derogar, M., M. van der Schaaf, and P. Lagergren, *Reference values for the EORTC QLQ-C30 quality of life questionnaire in a random sample of the Swedish population.* Acta Oncol, 2012. **51**(1): p. 10-6.
75. Jensen, R.E., et al., *United States Population-Based Estimates of Patient-Reported Outcomes Measurement Information System Symptom and Functional Status Reference Values for Individuals With Cancer.* J Clin Oncol, 2017. **35**(17): p. 1913-1920.
76. Cella, D., et al., *The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008.* J Clin Epidemiol, 2010. **63**(11): p. 1179-94.
77. Reilly, C.M., et al., *A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment.* Support Care Cancer, 2013. **21**(6): p. 1525-50.
78. Bohn, S.H., et al., *Chronic fatigue and associated factors among long-term survivors of cancers in young adulthood.* Acta Oncol, 2019. **58**(5): p. 753-762.
79. Kiserud, C.E., et al., *Fatigue in male lymphoma survivors differs between diagnostic groups and is associated with latent hypothyroidism.* Acta Oncologica, 2015. **54**(1): p. 49-59.
80. Stromgren, A.S., et al., *Symptom priority and course of symptomatology in specialized palliative care.* J Pain Symptom Manage, 2006. **31**(3): p. 199-206.
81. Homsi, J., et al., *Symptom evaluation in palliative medicine: patient report vs systematic assessment.* Support Care Cancer, 2006. **14**(5): p. 444-53.
82. Walsh, D., S. Donnelly, and L. Rybicki, *The symptoms of advanced cancer: relationship to age, gender, and performance status in 1,000 patients.* Support Care Cancer, 2000. **8**(3): p. 175-9.
83. van den Beuken-van Everdingen, M.H., et al., *Prevalence of pain in patients with cancer: a systematic review of the past 40 years.* Ann Oncol, 2007. **18**(9): p. 1437-49.
84. van den Beuken-van Everdingen, M.H., et al., *Update on Prevalence of Pain in Patients With Cancer: Systematic Review and Meta-Analysis.* J Pain Symptom Manage, 2016. **51**(6): p. 1070-1090 e9.
85. Greco, M.T., et al., *Quality of cancer pain management: an update of a systematic review of undertreatment of patients with cancer.* J Clin Oncol, 2014. **32**(36): p. 4149-54.

86. Rayner, L., et al., *The detection of depression in palliative care.* Curr Opin Support Palliat Care, 2009. **3**(1): p. 55-60.
87. Mitchell, A.J., et al., *Prevalence of depression, anxiety, and adjustment disorder in oncological, haematological, and palliative-care settings: a meta-analysis of 94 interview-based studies.* Lancet Oncol, 2011. **12**(2): p. 160-74.
88. Osoba, D., et al., *Health-related quality of life in men with metastatic prostate cancer treated with prednisone alone or mitoxantrone and prednisone.* J Clin Oncol, 1999. **17**(6): p. 1654-63.
89. Tannock, I.F., et al., *Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian randomized trial with palliative end points.* J Clin Oncol, 1996. **14**(6): p. 1756-64.
90. Bloomfield, D.J., et al., *Economic evaluation of chemotherapy with mitoxantrone plus prednisone for symptomatic hormone-resistant prostate cancer: based on a Canadian randomized trial with palliative end points.* J Clin Oncol, 1998. **16**(6): p. 2272-9.
91. Kantoff, P.W., et al., *Hydrocortisone with or without mitoxantrone in men with hormone-refractory prostate cancer: results of the cancer and leukemia group B 9182 study.* J Clin Oncol, 1999. **17**(8): p. 2506-13.
92. Gulbrandsen, F.W.N., *Health-related Quality of Life and Patients' Perceptions in Interferon-treated Multiple Myeloma Patients.* Acta Oncologica, 2000. **39**(7): p. 809-813.
93. Cleeland, C.S. and K.M. Ryan, *Pain assessment: global use of the Brief Pain Inventory.* Ann Acad Med Singapore, 1994. **23**(2): p. 129-38.
94. Mendoza, T.R., et al., *The rapid assessment of fatigue severity in cancer patients: use of the Brief Fatigue Inventory.* Cancer, 1999. **85**(5): p. 1186-96.
95. Gning, I., et al., *Development and Initial Validation of the Thyroid Cancer Module of the M. D. Anderson Symptom Inventory.* Oncology, 2009. **76**(1): p. 59-68.
96. De Conno, F., et al., *Pain measurement in cancer patients: a comparison of six methods.* Pain, 1994. **57**(2): p. 161-6.
97. Ferraz, M.B., et al., *Reliability of pain scales in the assessment of literate and illiterate patients with rheumatoid arthritis.* J Rheumatol, 1990. **17**(8): p. 1022-4.
98. Cleeland, C.S. *The M.D. Anderson Symptom Inventory User Guide. Version 1*. 2009 November 19, 2009; Available from: https://www.mdanderson.org/content/dam/mdanderson/documents/Departments-and-Divisions/Symptom-Research/MDASI_userguide.pdf.
99. Hui, D. and E. Bruera, *The Edmonton Symptom Assessment System 25 Years Later: Past, Present, and Future Developments.* J Pain Symptom Manage, 2017. **53**(3): p. 630-643.
100. Watanabe, S., et al., *The Edmonton symptom assessment system—what do patients think?* Supportive Care in Cancer, 2009. **17**(6): p. 675-683.
101. Garyali, A., et al., *Errors in symptom intensity self-assessment by patients receiving outpatient palliative care.* Journal of Palliative Medicine, 2006. **9**(5): p. 1059-1065.
102. Watanabe, S.M., et al., *A Multicenter Study Comparing Two Numerical Versions of the Edmonton Symptom Assessment System in Palliative Care Patients.* Journal of Pain and Symptom Management, 2011. **41**(2): p. 456-468.
103. Sigurdardottir, K.R., et al., *The European Association for Palliative Care basic dataset to describe a palliative care cancer population: Results from an international Delphi process.* Palliat Med, 2014. **28**(6): p. 463-473.
104. Selby, D., et al., *A Single Set of Numerical Cutpoints to Define Moderate and Severe Symptoms for the Edmonton Symptom Assessment System.* Journal of Pain and Symptom Management, 2010. **39**(2): p. 241-249.
105. Seow, H., et al., *Do High Symptom Scores Trigger Clinical Actions? An Audit After Implementing Electronic Symptom Screening.* Journal of Oncology Practice, 2012. **8**(6): p. e142-e148.

106. Hui, D., et al., *Personalized symptom goals and response in patients with advanced cancer.* Cancer, 2016. **122**(11): p. 1774-1781.
107. Hui, D., et al., *Minimal clinically important differences in the Edmonton Symptom Assessment Scale in cancer patients: A prospective, multicenter study.* Cancer, 2015. **121**(17): p. 3027-3035.
108. Greenberg, D., et al., *When is cancer care cost-effective? A systematic overview of cost-utility analyses in oncology.* J Natl Cancer Inst, 2010. **102**(2): p. 82-8.
109. Guyatt, G.H., et al., *Methods to Explain the Clinical Significance of Health Status Measures.* Mayo Clinic Proceedings, 2002. **77**(4): p. 371-383.
110. Garcia, S.F., et al., *Standardizing patient-reported outcomes assessment in cancer clinical trials: a patient-reported outcomes measurement information system initiative.* J Clin Oncol, 2007. **25**(32): p. 5106-12.
111. Snyder, C.F., et al., *Relevant content for a patient-reported outcomes questionnaire for use in oncology clinical practice: Putting doctors and patients on the same page.* Qual Life Res, 2010. **19**(7): p. 1045-55.
112. Basch, E. and A.P. Abernethy, *Supporting clinical practice decisions with real-time patient-reported outcomes.* J Clin Oncol, 2011. **29**(8): p. 954-6.
113. Fisch, M.J., et al., *Prospective, observational study of pain and analgesic prescribing in medical oncology outpatients with breast, colorectal, lung, or prostate cancer.* J Clin Oncol, 2012. **30**(16): p. 1980-8.
114. Atkinson, T.M., et al., *Reliability of adverse symptom event reporting by clinicians.* Qual Life Res, 2012. **21**(7): p. 1159-64.
115. Xiao, C., R. Polomano, and D.W. Bruner, *Comparison between patient-reported and clinician-observed symptoms in oncology.* Cancer Nurs, 2013. **36**(6): p. E1-e16.
116. Ellegaard, M.B., et al., *Women with breast cancer report substantially more disease- and treatment-related side or late effects than registered by clinical oncologists: a cross-sectional study of a standard follow-up program in an oncological department.* Breast Cancer Res Treat, 2017. **164**(3): p. 727-736.
117. Greenhalgh, J., *The applications of PROs in clinical practice: what are they, do they work, and why?* Quality of Life Research, 2009. **18**(1): p. 115-123.
118. Greenhalgh, J., *The applications of PROs in clinical practice: what are they, do they work, and why?* Qual Life Res, 2009. **18**(1): p. 115-23.
119. Zikos, E., et al., *Health-related quality of life in small-cell lung cancer: a systematic review on reporting of methods and clinical issues in randomised controlled trials.* Lancet Oncol, 2014. **15**(2): p. e78-89.
120. Gravis, G., et al., *Patients' self-assessment versus investigators' evaluation in a phase III trial in non-castrate metastatic prostate cancer (GETUG-AFU 15).* Eur J Cancer, 2014. **50**(5): p. 953-62.
121. Kyte, D., et al., *Current practices in patient-reported outcome (PRO) data collection in clinical trials: a cross-sectional survey of UK trial staff and management.* BMJ Open, 2016. **6**(10): p. e012281.
122. Calvert, M., et al., *The CONSORT Patient-Reported Outcome (PRO) extension: implications for clinical trials and practice.* Health Qual Life Outcomes, 2013. **11**: p. 184.
123. Pe, M., et al., *Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review.* Lancet Oncol, 2018. **19**(9): p. e459-e469.
124. Fiteni, F., et al., *Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review.* BMC Cancer, 2016. **16**: p. 122.
125. Bottomley, A., et al., *A review of the quality of statistical methods employed for analyzing quality of life data in cancer RCTs.* Journal of Clinical Oncology, 2016. **34**(15_suppl): p. 10058-10058.

126. Little, R.J., et al., *The prevention and treatment of missing data in clinical trials.* N Engl J Med, 2012. **367**(14): p. 1355-60.
127. Little, R.J., et al., *The design and conduct of clinical trials to limit missing data.* Stat Med, 2012. **31**(28): p. 3433-43.
128. Calvert, M., et al., *Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension.* Jama, 2018. **319**(5): p. 483-494.
129. Calvert, M., et al., *Reporting of Patient-Reported Outcomes in Randomized Trials: The CONSORT PRO ExtensionPatient Reported Outcomes in Randomized Trials.* JAMA, 2013. **309**(8): p. 814-822.
130. Bottomley, A., et al., *Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards.* The Lancet Oncology, 2016. **17**(11): p. e510-e514.
131. Bottomley, A., et al., *Moving forward toward standardizing analysis of quality of life data in randomized cancer clinical trials.* Clin Trials, 2018. **15**(6): p. 624-630.
132. Coens, C., et al., *International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium.* Lancet Oncol, 2020. **21**(2): p. e83-e96.
133. Riis, C.L., et al., *Satisfaction with care and adherence to treatment when using patient reported outcomes to individualize follow-up care for women with early breast cancer - a pilot randomized controlled trial.* Acta Oncol, 2020. **59**(4): p. 444-452.
134. Fayers, P. and A. Bottomley, *Quality of life research within the EORTC-the EORTC QLQ-C30. European Organisation for Research and Treatment of Cancer.* Eur J Cancer, 2002. **38 Suppl 4**: p. S125-33.
135. Denis, F., et al., *Improving Survival in Patients Treated for a Lung Cancer Using Self-Evaluated Symptoms Reported Through a Web Application.* Am J Clin Oncol, 2017. **40**(5): p. 464-469.
136. European Association for Palliative Care Research Network. Available from: http://www.eapcnet.eu/Themes/Research.aspx.
137. European Palliative Care Research Centre (PRC). Available from: www.ntnu.edu/prc.
138. Fyllingen, E.H., et al., *Computer-based assessment of symptoms and mobility in palliative care: feasibility and challenges.* J Pain Symptom Manage, 2009. **38**(6): p. 827-36.
139. Raj, S.X., et al., *COMBAT study - Computer based assessment and treatment - A clinical trial evaluating impact of a computerized clinical decision support tool on pain in cancer patients.* Scand J Pain, 2017. **17**: p. 99-106.
140. Andreassen Jaatun, E.A., et al., *Development and testing of a computerized pain body map in patients with advanced cancer.* J Pain Symptom Manage, 2014. **47**(1): p. 45-56.
141. Andreassen Jaatun, E.A., et al., *Designing a reliable pain drawing tool: avoiding interaction flaws by better tailoring to patients' impairments.* Pers Ubiquit Comput 2015. **19**: p. 635–648.
142. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure.* J Gen Intern Med, 2001. **16**(9): p. 606-13.
143. Sangha, O., et al., *The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research.* Arthritis Rheum, 2003. **49**(2): p. 156-63.
144. Kirkova, J., et al., *Cancer symptom assessment instruments: a systematic review.* J Clin Oncol, 2006. **24**(9): p. 1459-73.
145. Mendoza, T.R., et al., *Measuring the symptom burden of lung cancer: the validity and utility of the lung cancer module of the M. D. Anderson Symptom Inventory.* Oncologist, 2011. **16**(2): p. 217-27.
146. Wang, X.S., et al., *Validation and application of a module of the M. D. Anderson Symptom Inventory for measuring multiple symptoms in patients with gastrointestinal cancer (the MDASI-GI).* Cancer, 2010. **116**(8): p. 2053-63.
147. Dewolf, L., et al., *EORTC Quality of Life Group Translation Procedure*. Third ed. 2009.

148. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. 2013: American Psychiatric Pub.
149. *World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects.* Jama, 2013. **310**(20): p. 2191-4.
150. Martin-Moreno, J., T. Albreht, and K. SR., *Boosting Innovation and Cooperation in European Cancer Control. Key findings from the European Partnership for Action Against Cancer*. 2013, Ljubjana, Slovenia: National Institute of Public Health of the Republic of Slovenia.
151. Lewis C, *Using the "thinking- aloud" method in cognitive interface design*. 1982: RC 9265, Yorktown Heights, New York.
152. Rubin J, C.D., *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests* Second ed. 2008: Wiley.
153. Stinson, J., et al., *Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis.* J Med Internet Res, 2010. **12**(3): p. e30.
154. Stinson, J., et al., *Usability testing of an online self-management program for adolescents with cancer.* J Pediatr Oncol Nurs, 2015. **32**(2): p. 70-82.
155. Krogstad, H., et al., *Development of EirV3: A Computer-Based Tool for Patient-Reported Outcome Measures in Cancer.* JCO Clinical Cancer Informatics, 2017.
156. Knudsen, A.K., et al., *Which domains should be included in a cancer pain classification system? Analyses of longitudinal data.* Pain, 2012. **153**(3): p. 696-703.
157. Knudsen, A.K., et al., *Classification of pain in cancer patients--a systematic literature review.* Palliat Med, 2009. **23**(4): p. 295-308.
158. Caraceni, A., et al., *Use of opioid analgesics in the treatment of cancer pain: evidence-based recommendations from the EAPC.* Lancet Oncol, 2012. **13**(2): p. e58-68.
159. Rayner, L., et al., *The development of evidence-based European guidelines on the management of depression in palliative cancer care.* Eur J Cancer, 2011. **47**(5): p. 702-12.
160. Sigurdardottir, K.R., et al., *The European Association for Palliative Care basic dataset to describe a palliative care cancer population: Results from an international Delphi process.* Palliat Med, 2014.
161. Fearon, K., et al., *Definition and classification of cancer cachexia: an international consensus.* Lancet Oncol, 2011. **12**(5): p. 489-95.
162. Kaasa, S., et al., *Expert conference on cancer pain assessment and classification—the need for international consensus: working proposals on international standards.* BMJ Supportive & Palliative Care, 2011. **1**: p. 281-287.
163. Bauer, J., S. Capra, and M. Ferguson, *Use of the scored Patient-Generated Subjective Global Assessment (PG-SGA) as a nutrition assessment tool in patients with cancer.* Eur J Clin Nutr, 2002. **56**(8): p. 779-85.
164. Kroenke, K., et al., *Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection.* Ann Intern Med, 2007. **146**(5): p. 317-25.
165. Morin, C.M., *Insomnia: Psychological Assessment and Management*. 1996.
166. National Cancer Institute *Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0*. 2009.
167. Theofanos, M.F. and W. Quesenbery, *Towards the Design of Effective Formative Test Reports.* Journal of Usability Studies, 2005. **1**(1): p. 27-45.
168. International Organization for Standardization/International Electrotehcnical Commission (ISO/IEC) 9241-11, *Ergonomic requirements for office work with visual display terminals (VDTs)-Part 11- Guidance on usability.* 1998.
169. Corrao, N.J., et al., *Importance of Testing for Usability When Selecting and Implementing an Electronic Health or Medical Record System.* Journal of Oncology Practice, 2010. **6**(3): p. 120-124.
170. Stinson, J.N., et al., *Perspectives on quality and content of information on the internet for adolescents with cancer.* Pediatr Blood Cancer, 2011. **57**(1): p. 97-104.

171. Cicchetti, D.V., *Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.* Psychological Assessment, 1994. **6**(4): p. 284-290.
172. Tomenson, B., et al., *Total somatic symptom score as a predictor of health outcome in somatic symptom disorders.* Br J Psychiatry, 2013. **203**(5): p. 373-80.
173. Rokstad, K., J. Straand, and H. Sandvik, *[Patient encounters in general practice. An epidemiological survey in More and Romsdal].* Tidsskr Nor Laegeforen, 1997. **117**(5): p. 659-64.
174. Bardel, A., et al., *Age and sex related self-reported symptoms in a general population across 30 years: Patterns of reporting and secular trend.* PLoS One, 2019. **14**(2): p. e0211532.
175. Bøhn, S.-K.H., et al., *Chronic fatigue and associated factors among long-term survivors of cancers in young adulthood.* Acta Oncologica, 2019. **58**(5): p. 753-762.
176. Cleeland, C.S., et al., *Automated symptom alerts reduce postoperative symptom severity after cancer surgery: a randomized controlled clinical trial.* Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2011. **29**(8): p. 994-1000.
177. Kroenke, K., et al., *Telecare Collaborative Management of Chronic Pain in Primary Care: A Randomized Clinical Trial.* JAMA, 2014. **312**(3): p. 240-248.
178. Holzner, B., et al., *The Computer-based Health Evaluation Software (CHES): a software for electronic patient-reported outcome monitoring.* BMC Medical Informatics and Decision Making, 2012. **12**: p. 126.
179. Breakey, V.R., et al., *The value of usability testing for Internet-based adolescent self-management interventions: "Managing Hemophilia Online".* BMC Med Inform Decis Mak, 2013. **13**: p. 113.
180. Wolpin, S.E., et al., *Development and usability testing of a web-based cancer symptom and quality-of-life support intervention.* Health Informatics J, 2015. **21**(1): p. 10-23.
181. Basch, E.M., et al., *Overall survival results of a randomized trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment.* Journal of Clinical Oncology, 2017. **35**(18_suppl): p. LBA2-LBA2.
182. Scott, N., et al., *EORTC QLQ-C30 Reference Values Manual*. 2 ed. 2008, Brussels, Belgium: EORTC Quality of Life Group. 427.
183. Giesinger, J.M., et al., *Thresholds for clinical importance for four key domains of the EORTC QLQ-C30: physical functioning, emotional functioning, fatigue and pain.* Health Qual Life Outcomes, 2016. **14**: p. 87.
184. Johnsen, A.T., et al., *Symptoms and problems in a nationally representative sample of advanced cancer patients.* Palliat Med, 2009. **23**(6): p. 491-501.
185. Gulbrandsen, N., M.J. Hjermstad, and F. Wisloff, *Interpretation of quality of life scores in multiple myeloma by comparison with a reference population and assessment of the clinical importance of score differences.* Eur J Haematol, 2004. **72**(3): p. 172-80.
186. Oldenmenger, W.H., et al., *Cut points on 0-10 numeric rating scales for symptoms included in the Edmonton Symptom Assessment Scale in cancer patients: a systematic review.* J Pain Symptom Manage, 2013. **45**(6): p. 1083-93.
187. Norway, S. *Statistics Norway*. 2020 [cited 2020; Available from: https://www.ssb.no/befolkning.
188. M, V., L. S, and L. P, *Medical Statistics in clinical and epidemiological research*. Vol. 1. 2012: Gyldendal.
189. Grotmol, K.S., et al., *Depression-A Major Contributor to Poor Quality of Life in Patients With Advanced Cancer.* J Pain Symptom Manage, 2017. **54**(6): p. 889-897.
190. Grotmol, K.S., et al., *Patients with advanced cancer and depression report a significantly higher symptom burden than non-depressed patients.* Palliative and Supportive Care, 2019. **17**(2): p. 143-149.

191. Coons, S.J., et al., *Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report.* Value Health, 2009. **12**(4): p. 419-29.
192. Bastien, J.M., *Usability testing: a review of some methodological and technical aspects of the method.* Int J Med Inform, 2010. **79**(4): p. e18-23.
193. Marzorati, C., et al., *Telemedicine Use Among Caregivers of Cancer Patients: Systematic Review.* J Med Internet Res, 2018. **20**(6): p. e223.
194. Schoen, M.W., et al., *Software for Administering the National Cancer Institute's Patient-Reported Outcomes Version of the Common Terminology Criteria for Adverse Events: Usability Study.* JMIR Hum Factors, 2018. **5**(3): p. e10070.
195. European Association for Palliative Care Research, *European Association for Palliative Care Research Network.*
196. *European Palliative Care Research Centre (PRC),* . Available from: www.ntnu.edu/PRC.
197. Malterud, K., *Qualitative research: standards, challenges, and guidelines.* Lancet, 2001. **358**(9280): p. 483-8.
198. Gratzer, D., S. Khalid-Khan, and S. Balasingham, *The Internet and CBT: A New Clinical Application of an Effective Therapy*. 2018.
199. Rollman, B.L., et al., *Effectiveness of Online Collaborative Care for Treating Mood and Anxiety Disorders in Primary Care: A Randomized Clinical Trial.* JAMA Psychiatry, 2018. **75**(1): p. 56-64.
200. Brooke, J., *SUS: a "quick and dirty" usability scale, in Usability Evaluation in Industry. Jordan P.W., Thomas B., Weerdmeester B.A., McClelland I.L. (editor).* Taylor & Francis: London., 1996: p. 189-194.
201. Curtin, R., S. Presser, and E. Singer, *Changes in Telephone Survey Nonresponse over the Past Quarter Century.* Public Opinion Quarterly, 2005. **69**(1): p. 87-98.
202. Fossa, S.D., et al., *Stability of health-related quality of life in the Norwegian general population and impact of chronic morbidity in individuals with and without a cancer diagnosis.* Acta Oncol, 2007. **46**(4): p. 452-61.
203. Balter, K.A., et al., *Web-based and mailed questionnaires: a comparison of response rates and compliance.* Epidemiology, 2005. **16**(4): p. 577-9.
204. Green, E., et al., *Exploring patterns of response across the lifespan: the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study.* BMC Public Health, 2018. **18**(1): p. 760.
205. Taylor, J.R., et al., *The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample.* Neuroimage, 2017. **144**(Pt B): p. 262-269.
206. Shafto, M.A., et al., *The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing.* BMC Neurol, 2014. **14**: p. 204.
207. Galea, S. and M. Tracy, *Participation Rates in Epidemiologic Studies.* Annals of Epidemiology, 2007. **17**(9): p. 643-653.
208. Hille, E.T., et al., *Nonresponse bias in a follow-up study of 19-year-old adolescents born as preterm infants.* Pediatrics, 2005. **116**(5): p. e662-6.
209. Groves, R.M., *Survey Errors and Survey Costs*. 2004, USA: John Wiley & Sons, inc.
210. Grimes, D.A. and K.F. Schulz, *Bias and causal associations in observational research.* The Lancet, 2002. **359**(9302): p. 248-252.
211. Guirimand, F., et al., *Cancer-related symptom assessment in France: validation of the French M. D. Anderson Symptom Inventory.* J Pain Symptom Manage, 2010. **39**(4): p. 721-33.
212. Wang, X.S., et al., *Chinese version of the M. D. Anderson Symptom Inventory: validation and application of symptom measurement in cancer patients.* Cancer, 2004. **101**(8): p. 1890-901.
213. Bring. [cited 2019 October 31st]; Available from: https://www.bring.no/.

214. Partin, M.R., et al., *The impact of survey nonresponse bias on conclusions drawn from a mammography intervention trial.* J Clin Epidemiol, 2003. **56**(9): p. 867-73.
215. Kacha, A.K., et al., *Clinical Study Designs and Sources of Error in Medical Research.* J Cardiothorac Vasc Anesth, 2018. **32**(6): p. 2789-2801.
216. *Palliative Care Integrated in Oncology (PALLiON)*. 2017; Available from: https://pallion.no/.

Appendices

- MDASI (Norwegian)
- Paper I
- Paper II
- Paper III

27. Hvor sterke symptomer har du? Mange mennesker har symptomer på grunn av skader eller sykdom. Vi ber deg gradere hvor sterke de følgende symptomer har vært på sitt verste de siste 24 timene. Sett ett kryss for hver linje:

	Ikke hatt symptomet → *Det verste du kan forestille deg*										
	0	1	2	3	4	5	6	7	8	9	10
a. **Smerter** når de var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
b. **Tretthet (utmattelse)** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
c. **Kvalme** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
d. **Søvnforstyrrelser** når de var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
e. **Følelse av bekymring (uro)** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
f. **Tungpust** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
g. **Hukommelsesvansker** når de var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
h. **Appetittmangel** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
i. **Søvnighet (døsighet)** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
j. **Munntørrhet** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
k. **Tristhetsfølelse** når den var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
l. **Brekninger** når de var som VERST?.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
m. **Nummenhet/prikkinger i kroppen** når de var som VERST?	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Hvordan har symptomene påvirket livet ditt?

Symptomene påvirker ofte hvordan vi føler oss og hvordan vi fungerer. Hvor mye har symptomer påvirket de følgende områdene av livet ditt de siste 24 timene? Sett ett kryss på hver linje:

	Har ikke påvirket → *Har påvirket maksimalt*										
	0	1	2	3	4	5	6	7	8	9	10
n. **Generell aktivitet?**	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
o. **Humør?**	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
p. **Arbeid (inkludert husarbeid)?**.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
q. **Forhold til andre mennesker?**.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
r. **Evne til å gå?**.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
s. **Livsglede?**.....	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

28. Alt i alt, hvor stort problem har blærefunksjonen din vært for deg i løpet av **de siste 4 ukene**?

Ikke noe problem..... ☐
Svært lite problem..... ☐
Lite problem..... ☐
Moderat problem..... ☐
Stort problem ..... ☐

29. Alt i alt, hvor stort problem har tarmfunksjonen din vært for deg i løpet av **de siste 4 ukene?**

Ikke noe problem..... ☐
Svært lite problem..... ☐
Lite problem..... ☐
Moderat problem..... ☐
Stort problem ..... ☐

Paper I

Krogstad et al. BMC Public Health (2020) 20:988
https://doi.org/10.1186/s12889-020-09109-2

BMC Public Health

RESEARCH ARTICLE Open Access

Symptoms in the general Norwegian adult population - prevalence and associated factors

Hilde Krogstad[1,2*], Jon Håvard Loge[3,4,5], Kjersti S. Grotmol[3], Stein Kaasa[4], Cecilie E. Kiserud[6], Øyvind Salvesen[7] and Marianne Jensen Hjermstad[3,4]

Abstract

Background: Patients´ own perceptions and evaluations of symptoms, functioning and other health-related factors, i.e. Patient Reported Outcomes (PROs), are important elements for providing good patient care. Symptoms are subjective and best elicited by the patient orally or by using PRO measures (PROMs), be it on paper, or as electronic assessment tools. Reference values on frequently used PROMs facilitate the interpretation of scores for use in clinics and research settings, by comparing patient data with relevant samples from the general population. Study objectives were to (1) present reference values for the M.D. Anderson Symptom Inventory (MDASI) (2) examine the occurrence and intensity of symptoms assessed by the MDASI in a general Norwegian adult population sample, and (3) examine factors associated with higher symptom burden defined as the sum score of all symptoms, and factors associated with symptoms` interference on functions.

Methods: In 2015, MDASI was sent by mail as part of a larger survey, to a representative sample of the general Norwegian adult population ($N = 6165$). Medical comorbidities were assessed by the Self-Administered Comorbidity Questionnaire. Depression was self-reported on the Patient Health Questionnaire 9 (PHQ-9). Linear multivariable regression analysis was used to examine for factors associated with MDASI sum score and factors associated with symptoms' interference on functions.

Results: The response rate was 36%. More women (54%) than men (46%) responded. Mean age was 55 years (SD 14). The most frequent symptoms were fatigue (59.7%), drowsiness (56.2%) and pain (56.1%). Fatigue, pain and disturbed sleep had the highest mean scores. The presence of one or more comorbidities, increasing PHQ-9 score and lower level of education were associated with higher MDASI sum score ($p < 0.001$). The MDASI sum score and the PHQ-9 score were positively associated with all interference items ($p < 0.001$) except for walking ($p = 0.22$).

(Continued on next page)

* Correspondence: hilde.krogstad@stolav.no
[1]European Palliative Care Research Centre (PRC), Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, and St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
[2]Cancer Clinic, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway
Full list of author information is available at the end of the article

(Continued from previous page)
Conclusion: This study provides the first Norwegian reference values for MDASI. The presence of one or more comorbidities, higher level of depressive symptoms and lower level of education were significantly associated with higher MDASI sum score. These covariates must be controlled for when using the reference values.

Keywords: Patient reported outcome measures, PROMS, MDASI, Reference values

Background

Patient Reported Outcomes (PROs) are patients´ own perceptions and evaluations of symptoms, functioning and other health-related factors, and are important elements for providing good patient care [1]. A symptom is defined as any subjective evidence of a disease, health condition, or treatment-related effect that can be noticed and known only by the patient [1]. In contrast, a "sign" is any objective evidence of disease that can be identified by health care personnel by observations, examinations, biomarkers, imaging etc. or may be noticed and reported by the patient [1]. Symptoms may indicate the presence of a disease or a disorder but may also reflect normal variations in physical or psychological states as commonly experienced by most individuals. Symptoms are common in the general population [2–5]. A large Danish nationwide cohort study with 49, 706 respondents representative of the general population demonstrated that symptoms were common; about 9 out of 10 respondents reported at least one symptom within the preceding 4 weeks [2]. Other population studies have reported that 75 and 90% had experienced at least one symptom in the previous 2 weeks and 30 days respectively [3, 5]. Some symptoms have low positive predictive value for disease while others are stronger predictors [6]. As this may vary for different symptoms across patient populations, reference values from the general population provide important information about the predictive values of symptoms for disease. The prevalence of symptoms in the general population is found to be associated with factors such as chronic conditions, age, employment status, living situation and psychiatric disorders [3, 7]. The number of symptoms is also documented to have a linear relationship with functional status [4].

Patient-Reported Outcome Measures (PROMs) denote any standardized measure of a PRO, i.e. a questionnaire, of a patient's health and quality of life (QoL) [8]. These questionnaires are intended for self-completion by patients, in the form of the traditional paper forms or more recently in electronic formats (e-PROMs) for use on different platforms, e.g. cell phones, computers, tablets etc. [9]. PROMs provide information that comes directly from the patient [8]. In clinical care, PROMs can be used alongside laboratory tests and imaging, if properly assessed and followed. Regular and systematic use of PROMs may improve communication between patients and health care providers [10] and be used to monitor treatment response and detect unrecognized problems or problems not reported spontaneously by the patient [11]. Beyond their clinical utility, PROMs are increasingly being used as outcomes in epidemiologic, health economic and clinical research [12]. PROMS are also central components of patient-centered care [13, 14]. Recent studies suggested that active use of PROMs during treatment for advanced cancer may even prolong survival [15–17].

Clinicians or researchers often request reference data to facilitate the interpretation of patient data or study results [18]. Reference values for PROMs facilitate the interpretation of PROMs scores both in clinics and research settings, by comparing patient data with relevant samples from the general population. Reference values may also be used to evaluate the relative symptom burden of a disease in a given diagnosis, when controlled after adjusting for relevant covariates [19]. Hence, a number of datasets with population-based reference data have been published and are frequently being used, e.g. the Patient-Reported Outcomes Measurement Information System [19], European Organisation for Research and Treatment of Cancer (EORTC) Core Quality of Life Questionnaire C30 [20, 21] and the Functional Assessment of Cancer Therapy-General [22]. Reference values make comparisons between samples possible, but this requires adjusting for known variables that affect the outcomes, e.g. age, sex, residence, education, comorbidities and other sociodemographic variables [20, 23]. As reference values are based on self-report, as are patient-reported outcomes, there is not and should not be, a golden standard for a given symptom score as is the case. In contrast to e.g. reference values for laboratory results, the principle of PROs as part of patient-centered care is to assess the patients' own perception of symptoms and QoL. As such, reference data provide information about the distribution of self-reported QoL scores for given reference populations. These scores can be used as reference against which patient scores can be compared. If the average score in a patient group is significantly higher or lower than expected after controlling for known covariates, follow-up of potential disease or treatment side effects may be indicated [24]. The relevance of valid reference data is illustrated in follow-up studies among cancer survivors, which may go beyond

20 years post-treatment [25, 26]. During such a long period, common age-related health problems and life events may influence which symptoms the cancer survivors experience and how they perceive their QoL and level of functioning. By comparing with data from the general population one can ascertain if cancer survivors are at excess risk for specific symptoms and health problems compared to individuals with similar age, sex and other background variables.

The M.D. Anderson Symptom Inventory (MDASI) is a brief, reliable and valid tool for self-report of symptoms commonly experienced by patients with cancer and also assesses their impact on daily functioning [27]. The MDASI is frequently used in clinical cancer care [28, 29]. Importantly, all MDASI symptoms are prevalent in the general population and how self-reported severity of symptoms interfere daily living is an important issue in all populations. Reference values for the MDASI from the general adult population therefore allow for interpretation of scores from patient samples and for comparison across studies and between relevant populations samples. Up until now, there are no reference values for the MDASI from the Norwegian population, nor have we found this from other countries.

On this background, study objectives were to (1) present reference values for the M.D. Anderson Symptom inventory (MDASI), (2) examine the occurrence and intensity of symptoms assessed by the MDASI in a general Norwegian adult population sample, and (3) examine factors associated with higher symptom burden defined as the sum score of all symptoms, and factors associated with symptoms` interference on functions.

Methods

Data collection

In the spring 2015, 6165 subjects, aged 18–80 years, and representative of the general Norwegian adult population with respect to age, gender and place of residence, were randomly drawn by Bring Dialog [30]. They received a mailed questionnaire packet on paper containing the Short-Form Health Survey-36 (SF-36), version 1 [31, 32], the M.D. Anderson Symptom Inventory (MDASI) [27], the Fatigue Questionnaire (FQ) [33] and the Patient Health Questionnaire-9 (PHQ-9) [34, 35]. The questionnaire packet also included questions covering 13 comorbidities [36] and 14 questions related to socio-demographic variables, physical activity, general health and contact with health care providers. Socio-demographic variables (see below), comorbidities, the MDASI and the PHQ-9 were used in this study.

Socio-demographic variables

Socio-demographic variables included year of birth, sex, and level of education. Education was divided into three groups referring to highest level of completed education: elementary and/or primary school; second level (high school); and third level (university college or university). Comorbidities were self-reported on a modified version of the Self-Administered Comorbidity Questionnaire (SCQ) [36]. The subjects were asked "do you have, or have you ever had, any of the following diseases/problems?"

Instruments

The M.D. Anderson symptom inventory (MDASI)

The M. D. Anderson Symptom Inventory (MDASI) was developed by the Pain Research Group at M. D. Anderson Cancer Center at the University of Texas. Validation studies have shown that the MDASI is useful for symptom surveys, clinical trials, and patient follow-up care [28, 37, 38]. MDASI is designed for use in cancer populations [27], hence applies to patients with various cancer diagnoses and types of treatment. MDASI assesses the severity of 13 frequent symptoms experienced during the last 24 h (pain, fatigue, nausea, sleep disturbance, distress, shortness of breath, difficulty remembering, lack of appetite, drowsiness, dry mouth, sadness, vomiting, numbness/tingling) in patients with cancer. The response alternatives are 0–10 on numerical rating scales, with 0 meaning "not present" and 10 meaning "as bad as you can imagine". In this study, a cut off ≥1 was chosen to denote any presence of a symptom. These 13 items not only account for the most frequently reported symptoms by cancer patients, but they are also common reasons for contact with the health care system in the general population [27, 39]. In addition, the MDASI includes another six questions on how much the symptoms interfere with general activity, mood, work, relations with other people, walking and enjoyment of life. The interference items are also measured on 0–10 scales, with 0 meaning "did not interfere," and 10 meaning "interfered completely". The first introductory sentence in the MDASI refers to people with cancer *"people with cancer frequently have symptoms that are caused by their disease or by their treatment"*. For the purpose of this survey, the sentence was changed to: "*many people often have symptoms due to injuries or disease*". Thus, the word cancer was omitted from the questionnaire.

The translation of MDASI into Norwegian followed the multi-step, well-established 2009 procedures developed by the EORTC Quality of Life Group [40]. This includes two independent forward translations from English to Norwegian by native speakers of Norwegian language with good knowledge of English. A third person fluent in both languages merged the translations into a reconciled version, that was back-translated by two persons having a very good command of English. When comparing the original and the back-translated

English versions, no translation problems became apparent. The Norwegian version of the MDASI was proofread and pilot-tested by six persons who found the comprehensibility and clarity satisfactory according to the EORTC debriefing interviews (length, relevance, confusing, upsetting and intrusive items, unclear wording) [40]. Permission to translate and use the MDASI was obtained from MD Anderson, TX, USA.

The patient health Questionnaire-9 (PHQ-9)

PHQ-9 is a nine-item questionnaire designed to screen for depression [35]. The nine items correspond to the DSM-5 diagnostic criteria for major depressive disorder [41]. The response alternatives are the frequency to which these symptoms have been bothersome during the past 2 weeks, divided in four categories: 0 = *not at all*, 1 = *several days*, 2 = *more than half of the days* and 3 = *nearly every day*. "Major depression" is diagnosed if five or more of the symptoms have been present at least "more than half the days" in the past 2 weeks provided that one of these is item 1 (depressed mood) or item 2 (anhedonia). As a severity measure, the PHQ-9 score ranges from 0 to 27, since each item can be scored from 0 to 3. In the present study, the four somatic depression symptoms in the PHQ-9 are excluded to avoid overlap with MDASI (sleep-problems, fatigue, weight/appetite change and psychomotor retardation). The instrument will hereafter be referred to as the PHQ. Here, the score ranges from 0 to 15. We have previously shown that the agreement between the 9 - and 5 - item versions in detecting depression was excellent [42].

Statistical analysis

The returned questionnaires that were blank, had no data on sex or missed more than half of the individual MDASI symptoms were excluded from analysis. Standard descriptive analyses were used with the baseline characteristics. Variables examined included age, gender and education. The number of age groups was limited to six: 18–29, 30–39, 40–49, 50–59, 60–69, and 70–80 years. The number of comorbidities were grouped as follows: Category 0 (no comorbidity), category 1 (1–2 comorbidities) and category 2 (≥ 3 comorbidities). Basic descriptive analyses were used for the number and intensity of MDASI symptoms. The total MDASI sum score for the 13 symptoms was calculated (possible range 0–130; the sum of scores for the 13 individual symptoms).

Associations between the MDASI sum score as the dependent variable, and age, sex, education, comorbidity and depression as independent variables were analyzed using linear multivariable regression. Univariable linear regression was used to examine for factors associated with MDASI sum score. Variables from the univariable analyses with a p-value ≤0.10 were included in the multivariable regression model, which also included sex and age regardless of the significance in the univariable analyses. The six MDASI interference items were used as dependent variables in separate analyses. The corresponding effect sizes are reported as unstandardized coefficients and 95% confidence interval (CI). A p-value of < 0.05 was used to denote statistical significance.

The statistical software applied was IBM SPSS Statistics for Windows, version 25.0, (IBM Corporation, Armonk, NY, USA).

Ethical considerations

The study was performed according to the rules of the Helsinki declaration. All respondents received written information about the study. Return of the questionnaires was taken to indicate written, informed consent. The Regional Committee for Medical and Health Research Ethics (REC) South East Norway approved the survey (2014/1172).

Results

The overall response rate was 36%. Of the 2130 returned questionnaires, 23 were blank, 21 had no data on sex, and 65 had responded to less than half of the individual MDASI symptoms. All these respondents were omitted, giving a sample of 2021. Missing values of the MDASI ranged from 0.1% ($n = 3$, numbness) to 1.4% ($n = 28$, fatigue).

More women (54%) than men (46%) responded. As shown in a previous publication from the same material [32], the response rate for both men and women was 5% in the youngest age group (≤ 29 years) which was significantly lower compared to the other groups ($p < 0.001$). Mean age of the study sample was 55 years (SD 14) (Table 1). Forty-six% of the respondents had university college or university education.

Table 2 shows the frequency of comorbidities. Forty-two% reported no comorbidities, 45% reported one or two while 13% reported three comorbidities or more. The most frequent were hypertension, arthrosis and depression. Arthrosis and depression were more common in women (23.6 and 15.3% vs. 12.5 and 9.3%), while there was no difference regarding hypertension between men and women. Depression was more common among women in the youngest age group (23.1%) compared to women ≥70 years (15.3%).

The most frequent symptoms were fatigue (59.7%), drowsiness (56.2%) and pain (56.1%). When using a cut off ≥3, the prevalence was 34.8% for fatigue, 34.2% for pain and 26.7% for drowsiness (Table 3). The mean scores for the 13 symptoms by age and sex are presented in Table 4. Fatigue, pain and disturbed sleep had the highest mean scores overall (Fig. 1). Fatigue had the

Table 1 Socio-demographic characteristics, and mean MDASI sum score

Variables	Population ($N = 2021$)	Mean MDASI sum score (SD)[a]
Age		
Mean (±SD)	55 (14)	
Min.-Max.	18–79	
Age groups, N (%)		
≤ 29 years	101 (5.0)	18.78 (20.24)
30–39 years	197 (9.7)	15.76 (18.89)
40–49 years	390 (19.3)	14.68 (18.15)
50–59 years	467 (23.1)	15.46 (17.88)
60–69 years	499 (24.7)	15.13 (18.63)
≥ 70 years	367 (18.2)	15.84 (17.91)
Gender, N (%)		
Women	1101 (54)	16.71 (18.83)
Men	920 (46)	14.03 (17.65)
Education, N (%), Missing 10 (0.5)		
Elementary and/or primary school	344 (17.1)	18.63 (20.55)
Second level (high school)	751 (37.3)	16.98 (19.57)
Third level (university college or university)	916 (45.5)	12.98 (15.95)
Number of comorbidities, N (%)		
0	856 (42)	
1–2	912 (45)	
≥ 3	253 (13)	

[a]Min-max 0–130

Table 2 Comorbidities [a], overall and by sex

Comorbidity	All N (%)	Women N (%) $N = 1101$	Men N (%) $N = 920$
Heart disease	135 (6.7)	34 (3.1)	101 (11.0)
Hypertension	482 (23.8)	262 (23.8)	220 (23.9)
Chronic lung disease	205 (10.1)	116 (10.5)	89 (9.7)
Diabetes	113 (5.6)	44 (4.0)	69 (7.5)
Kidney disease	40 (2.0)	17 (1.5)	23 (2.5)
Liver disease	23 (1.1)	9 (0.8)	14 (1.5)
Stomach/Bowel disease	123 (6.1)	62 (5.6)	61 (6.6)
Rheumatic disease	145 (7.2)	100 (9.1)	45 (4.9)
Arthrosis	375 (18.6)	260 (23.6)	115 (12.5)
Epilepsy	22 (1.1)	16 (1.5)	6 (0.7)
Stroke	60 (3.0)	27 (2.5)	33 (3.6)
Depression	259 (12.8)	169 (15.3)	90 (9.8)
Other psychiatric conditions	155 (7.7)	99 (9.0)	56(6.1)

[a]The Self-Administered Comorbidity Questionnaire [36]

Table 3 Frequency of symptoms (MDASI score), N (%)

Symptom	MDASI score ≥ 1	MDASI score ≥ 3
Pain	1125 (56.1)	692 (34.5)
Fatigue (tiredness)	1190 (59.7)	704 (35.3)
Nausea	305 (15.3)	134 (1.3)
Disturbed sleep	913 (45.5)	507 (25.3)
Being distressed	913 (45.5)	433 (21.6)
Shortness of breath	600 (30.0)	289 (14.4)
Remembering	699 (34.9)	276 (13.8)
Lack of appetite	357 (17.8)	148 (7.4)
Drowsy	1127 (56.2)	540 (26.9)
Dry mouth	578 (28.8)	285 (14.2)
Sad	789 (39.2)	374 (18.6)
Vomiting	164 (8.1)	69 (3.4)
Numbness or tingling	503 (24.9)	265 (13.1)

highest mean score; 2.39 in women and 1.90 in men. The mean scores for fatigue were highest in the youngest age group (< 30 years), with higher score for women (3.45) than in men (2.36). Overall, the mean scores for pain were 2.24 in women and 1.94 in men, and the mean scores for disturbed sleep were 1.93 in women and 1.42 in men.

Univariable regression analysis showed a significant positive association between the presence of one or more comorbidities ($p < 0.001$) and PHQ- score and MDASI sum score ($p < 0.001$). Level of education was also associated with MDASI sum score ($p < 0.001$), while no association was found with age ($p = 0.5$). Further, because of the low response rate in youngest age group separate analyses were done without this age group yielding similar results.

Multivariable linear regressions (Table 5) showed positive significant associations between the MDASI sum score and depression on the PHQ ($p < 0.001$) and the presence of one or more comorbidities ($p < 0.001$). Participants with the highest education level had significantly lower MDASI sum score than respondents with education in elementary and/or primary school ($p = 0.006$) and second level (high school) ($p = 0.003$). Women had significantly higher MDASI sum score than men in univariable analyses ($p = 0.001$), but not in the multivariable regression model. The overall model fit was $R^2 = 0.45$.

Each interference item was used as the dependent variable in separate multivariable linear regression analyses (Table 6), with age, sex, education, comorbidity, PHQ score and MDASI sum score as independent variables. Comorbidities, PHQ score and MDASI sum score were significantly associated with both general activity and work as the dependent variables ($p \leq 0.001$). Increased

Table 4 Mean MDASI scores (SD)[a] by sex and age groups, $N = 2021$

Symptoms	Age groups												Total	
	18–29 years		30–39 years		40–49 years		50–59 years		60–69 years		70–80 years			
	W	M	W	M	W	M	W	M	W	M	W	M	W	M
	(*n* = 64–65)	(*n* = 36)	(*n* = 115–116)	(*n* = 80–81)	(*n* = 222–227)	(*n* = 162–163)	(*n* = 251–257)	(*n* = 207–210)	(*n* = 242–248)	(*n* = 248–251)	(*n* = 182–188)	(*n* = 174–179)	(*n* = 1084–1101)	(*n* = 914–920)
Core items														
Pain	1.69 (2.44)	1.17 (2.12)	1.69 (2.49)	1.64 (2.35)	2.22 (2.71)	1.90 (2.34)	2.39 (2.65)	1.97 (2.58)	2.53 (2.78)	2.04 (2.46)	2.21 (2.58)	2.10 (2.51)	2.24 (2.66)	1.94 (2.46)
Fatigue	3.45 (2.86)	2.36 (2.58)	2.87 (2.75)	1.96 (2.32)	2.53 (2.78)	1.94 (2.43)	2.57 (2.67)	1.87 (2.39)	2.03 (2.61)	1.90 (2.31)	1.78 (2.20)	1.80 (2.23)	2.39 (2.66)	1.90 (2.34)
Nausea	1.03 (2.03)	0.56 (1.59)	0.74 (1.62)	0.28 (0.97)	0.64 (1.70)	0.22 (0.83)	0.55 (1.61)	0.41 (1.51)	0.40 (1.40)	0.24 (0.84)	0.51 (1.37)	0.40 (1.28)	0.57 (1.58)	0.32 (1.16)
Disturbed sleep	2.47 (3.41)	1.92 (3.00)	1.70 (2.54)	1.57 (2.53)	1.87 (2.84)	1.26 (2.26)	2.05 (2.68)	1.56 (2.41)	1.83 (2.53)	1.44 (2.34)	1.95 (2.48)	1.23 (2.10)	1.93 (2.68)	1.42 (2.34)
Distress/feeling upset	2.55 (3.12)	1.72 (2.25)	2.00 (2.70)	1.84 (2.24)	1.63 (2.54)	1.09 (2.12)	1.58 (2.27)	1.39 (2.32)	1.58 (2.40)	1.18 (2.13)	1.54 (2.22)	1.09 (1.83)	1.68 (2.46)	1.27 (2.14)
Shortness of breath	0.88 (1.88)	0.58 (1.34)	0.64 (1.60)	0.99 (2.11)	0.66 (1.76)	0.72 (1.77)	0.67 (1.69)	0.85 (1.78)	1.20 (2.30)	1.19 (2.12)	1.45 (2.43)	1.36 (1.98)	0.93 (2.02)	1.02 (1.94)
Difficulty remembering	1.39 (2.64)	0.64 (1.48)	1.07 (2.12)	0.62 (1.72)	1.17 (2.24)	0.67 (1.52)	0.98 (1.83)	0.78 (1.57)	0.86 (1.65)	0.88 (1.50)	1.01 (1.63)	1.40 (2.11)	1.03 (1.94)	0.89 (1.69)
Lack of appetite	0.83 (1.92)	1.00 (1.97)	0.80 (2.07)	0.38 (1.34)	0.34 (1.12)	0.36 (1.22)	0.47 (1.35)	0.52 (1.60)	0.53 (1.62)	0.54 (1.49)	0.66 (1.61)	0.55 (1.34)	0.55 (1.55)	0.51 (1.46)
Drowsiness	2.66 (2.62)	2.17 (2.74)	2.26 (2.71)	1.89 (2.51)	1.96 (2.55)	1.84 (2.31)	1.98 (2.49)	1.75 (2.46)	1.57 (2.26)	1.73 (2.20)	1.63 (2.26)	1.56 (2.09)	1.89 (2.46)	1.75 (2.31)
Dry mouth	0.55 (1.23)	0.92 (2.03)	0.59 (1.72)	0.59 (1.74)	0.65 (1.86)	0.68 (1.69)	1.08 (2.22)	0.76 (1.80)	1.14 (2.22)	1.10 (2.18)	1.66 (2.54)	1.40 (2.32)	1.02 (2.14)	0.95 (2.02)
Sadness	2.15 (2.70)	1.67 (2.88)	1.84 (2.89)	1.16 (1.86)	1.48 (2.44)	1.03 (2.04)	1.34 (2.21)	1.19 (2.19)	1.29 (2.32)	1.02 (1.94)	1.22 (2.16)	1.14 (2.17)	1.44 (2.39)	1.12 (2.10)
Vomiting	0.58 (1.81)	0.17 (0.70)	0.30 (1.16)	0.31 (1.11)	0.26 (1.26)	0.11 (0.63)	0.26 (1.23)	0.23 (1.09)	0.19 (1.03)	0.26 (1.04)	0.35 (1.18)	0.25 (1.11)	0.28 (1.22)	0.22 (1.00)
Numbness/ tingling	0.57 (1.37)	0.50 (1.40)	0.71 (1.67)	0.59 (1.57)	0.92 (2.21)	0.74 (1.76)	0.85 (1.85)	0.82 (1.88)	1.01 (2.11)	0.80 (1.89)	1.11 (2.07)	0.85 (1.76)	0.91 (1.99)	0.77 (1.79)
Interference items														
General activity	2.17 (2.66)	1.44 (2.20)	1.75 (2.52)	1.65 (2.52)	1.94 (2.80)	1.65 (2.55)	1.83 (2.65)	1.54 (2.44)	1.95 (2.70)	1.49 (2.43)	1.78 (2.63)	1.39 (2.14)	1.88 (2.67)	1.52 (2.30)
Mood	2.69 (2.87)	2.08 (2.38)	2.30 (2.69)	1.73 (2.28)	1.96 (2.67)	1.77 (2.50)	1.56 (2.28)	1.25 (1.99)	1.38 (2.21)	1.15 (1.93)	1.34 (2.02)	1.20 (2.03)	1.71 (2.42)	1.38 (2.14)
Working	2.31 (2.92)	1.92 (2.68)	2.29 (3.00)	1.68 (2.38)	2.12 (2.96)	1.84 (2.70)	1.96 (2.75)	1.44 (2.43)	1.99 (2.78)	1.32 (2.30)	1.98 (2.74)	1.38 (2.20)	2.06 (2.83)	1.51 (2.41)
Relations with other people	2.20 (2.76)	1.89 (2.71)	1.89 (2.66)	1.41 (2.52)	1.61 (2.49)	1.46 (2.42)	1.30 (2.17)	1.00 (1.98)	1.38 (2.39)	0.96 (1.78)	1.04 (2.00)	1.02 (2.00)	1.45 (2.37)	1.15 (2.11)
Walking	0.45 (1.49)	0.36 (1.44)	0.34 (0.93)	0.67 (1.97)	0.77 (2.05)	0.61 (1.68)	1.05 (2.22)	0.86 (2.14)	1.51 (2.79)	1.02 (2.13)	1.79 (2.99)	1.32 (2.41)	1.11 (2.39)	0.91 (2.10)
Enjoyment of life	1.68 (2.50)	1.83 (2.71)	1.58 (2.56)	1.44 (2.42)	1.49 (2.52)	1.37 (2.32)	1.36 (2.22)	1.20 (2.25)	1.47 (2.65)	1.18 (2.16)	1.41 (2.31)	1.27 (2.38)	1.46 (2.45)	1.29 (2.30)

W women, *M* men

[a]0 ("not present") to 10 ("as bad as you can imagine") NRS scale

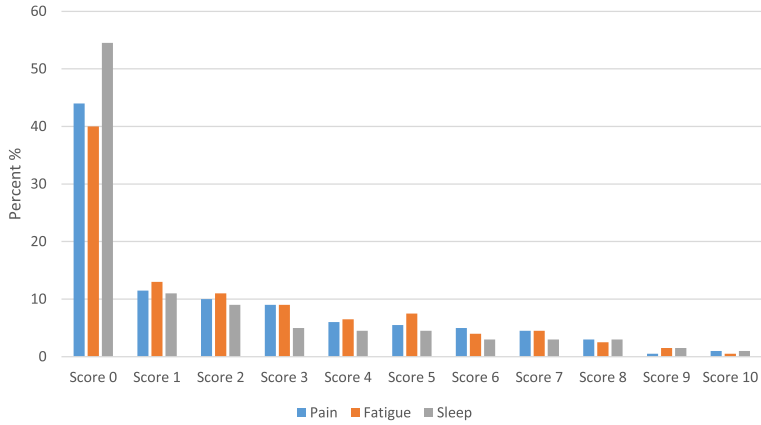


Fig. 1 Distribution of scores 0–10 on pain, fatigue, sleep

Table 5 Multiple linear regression on the MDASI sum score with age, sex, education, comorbidity and depression as explanatory variables ($N = 2021$)

	MDASI sum score[a]		
	Adjusted $R^2 = 0.45$		
	B	95% CI	p
Age groups			0.446
18–29 years	0.397	−2.761, 3.555	0.805
30–39 years	−0.449	−2.985, 2.087	0.728
40–49 years	− 1.214	−3.318, 0.890	0.258
50–59 years	0.735	−1.244, 2.715	0.466
60–69 years	0.292	−1.583, 2.167	0.760
70–80 years (ref)	–		
Sex			0.109
Women	0.99	−0.222, 2.202	0.109
Men (ref)	–		
Education			0.002
Elementary and/or primary school	2.591	0.759, 4.423	0.006
Second level (high school)	2.029	0.695, 3.363	0.003
Third level (university or university college) (ref)	–		
Comorbidities			0.000
0 (ref)	–		
1–2	3.452	2.116, 4.789	0.000
≥3	10.693	8.627, 12.760	0.000
Depression			0.000
PHQ score	4.685	4.412, 4.958	0.000

[a]Demographic and disease-related variables that were significantly correlated with MDASI sum score in the univariable analyses were entered as covariates

number of comorbidities and higher MDASI sum score were significantly associated with higher score on the interference item walking ($p < 0.001$). Further, the multivariable regression analyses showed that PHQ score and MDASI sum score were significantly associated ($p < 0.001$) with mood, relations and enjoyment of life as dependent variables.

Discussion

This study provides the first Norwegian reference values for the MDASI based on data from 2021 men and women aged 18–80 years collected in 2015. The most frequent symptoms overall were fatigue, drowsiness and pain. Fatigue, pain and disturbed sleep had the highest mean scores. The mean scores for fatigue were highest in the youngest age group (18–29 years). The presence of one or more comorbidities, increasing levels of depressive symptoms and lower level of education were significantly associated with a higher MDASI sum score. Comorbidity showed the strongest association; having three or more comorbidities increased the MDASI sum score with 10 points in average. Sex was not significantly associated with MDASI sum score when education, depression and comorbidities were controlled for in the regression model.

The Health Study of Nord-Trøndelag County (HUNT 3) found that the prevalence of chronic pain was 36% among women and 25% among men, and that the prevalence increased with age [43]. A random sample of participants were followed with annual measures over 4 years [44]. Here, pain intensity ranging from no pain to very mild, mild, moderate, severe and very severe pain was included to identify clinically important pain. A cut-off between mild and moderate pain may identify individuals with complex pain [45]. In our study, a cut off

Table 6 Multiple linear regression with the six interference items as the outcomes for all respondents included ($N = 2021$) [a]

	General activity			Mood			Working			Relations			Walking			Enjoyment of life		
	Adjusted $R^2 = 0.460$			Adjusted $R^2 = 0.584$			Adjusted $R^2 = 0.516$			Adjusted $R^{2=}$ 0.543			Adjusted $R^2 = 0.291$			Adjusted $R^2 = 0.589$		
	B	95% CI	p	B	95% CI	P	B	95% CI	p	B	95% CI	p	B	95% CI	p	B	95% CI	p
Age groups																		
18–29 years	0.100	−0.339, 0.539	0.655	0.66	0.32, 0.995	0.000	0.04	−0.39, 0.48	0.848	0.43	0.08, 0.77	0.016	− 0.98	−1.42, − 0.54	0.000	− 0.25	− 0.60, 0.11	0.176
30–39 years	0.181	0.173, 0.534	0.316	0.58	0.32, 0.85	0.000	0.27	− 0.08, 0.62	0.126	0.35	0.08, 0.63	0.012	−0.77	−1.12, − 0.42	0.000	− 0.10	−0.39, 0.18	0.473
40–49 years	0.381	0.087, 0.674	0.011	0.54	0.32, 0.76	0.000	0.37	0.08, 0.65	0.013	0.36	0.13, 0.59	0.002	−0.53	−0.83, − 0.24	0.000	− 0.02	− 0.26, 0.22	0.857
50–59 years	0.233	− 0.044, 0.509	0.099	0.13	− 0.08, 0.34	0.213	0.07	−0.20, 0.34	0.610	0.07	−0.15, 0.28	0.546	−0.38	−0.66, − 0.10	0.007	−0.06	− 0.28, 0.17	0.628
60–69 years		−0.036, 0.489	0.091	0.07	−0.30, 0.27	0.502	0.04	−0.22, 0.296	0.771	0.14	−0.07, 0.35	0.187	−0.14	−0.41, 0.12	0.287	0.06	−0.16, 0.27	0.609
70–80 years (ref)	–			–			–			–			–			–		
Sex																		
Women	0.084	−0.085, 0.253	0.331	0.02	−0.11, 0.15	0.75	0.22	0.06, 0.39	0.009	0.02	−0.12, 0.16	0.775	0.06	−0.11, 0.23	0.470	−0.13	−0.27, 0.01	0.066
Men (ref)	–			–			–			–			–			–		
Education			0.536	–														
Second level, first stage	0.145	−0.111, 0.402	0.267	–			0.11	−0.14, 0.36	0.387	–			0.36	0.10, 0.61	0.007	0.08	−0.13, 0.29	0.446
Second level, second stage	0.026	−0.160, 0.212	0.783	–			0.10	−0.09, 0.28	0.31	–			0.05	−0.14, 0.23	0.611	−0.14	−0.29, 0.02	0.077
Third level (ref)	–			–			–			–			–			–		
Comorbidities																		
0	−0.672	−0.967, −0.377	0.000	−0.07	−0.30, 0.16	0.530	−0.48	− 0.77, − 0.19	0.001	0.26	0.02, 0.50	0.033	−0.81	−1.10, − 0.51	0.000	−0.17	− 0.41, 0.07	0.160
1–2	−0.384	− 0.660, − 0.107	0.007	−0.01	− 0.22, 0.21	0.960	− 0.32	−0.60, − 0.05	0.020	0.34	0.12, 0.56	0.003	−0.78	−1.06, − 0.51	0.000	− 0.18	−0.41, 0.04	0.107
≥ 3 (ref)	–									–			–			–		
Depression																		
PHQ score	0.175	0.127, 0.224	0.000	0.34	0.30, 0.37	0.000	0.26	0.21, 0.31	0.000	0.36	0.32, 0.40	0.000	−0.03	−0.08, 0.02	0.224	0.47	0.43, 0.51	0.000
Symptom burden																		
MDASI sum	0.075	0.068, 0.081	0.000	0.06	0.06, 0.67	0.000	0.08	0.07, 0.08	0.000	0.06	0.05, 0.06	0.000	0.06	0.05, 0.07	0.000	0.05	0.05, 0.06	0.000

[a] Demographic and disease-related variables that were significantly correlated with MDASI sum score in the univariable analyses were entered as covariates

≥1 was chosen to identify the presence of a symptom. By increasing the cut off to ≥3, the prevalence was about 34% for pain, which corresponds to the finding in the HUNT 3 study.

Previous studies have shown that women generally report a higher number of symptoms than men [3, 5, 46, 47]. A Norwegian population study [47] also found that women reported a higher number of symptoms than men, although the association between somatic symptoms and anxiety and depression was equally strong in men and women indicating that the difference in prevalence of these conditions between the sexes could not explain the difference in the reported number of somatic symptoms. Elnegaard et.al [2]. found no sex differences for almost 2/3 of the reported symptoms leading to contact with a general practitioner in their population study. In our study, more women (15%) than men (9%) reported depression on the PHQ-9. This might explain why sex was not associated with symptom sum score when controlling for depression.

Across the lifespan, depression is almost twice as common in women as in men. The prevalence of major depressive episode worldwide is approximately 5% [48]. However, major depressive disorder is different from feelings of sadness which also may lead to increased symptom score. The PHQ-9 is a tool that can be used to identify and assess depression, but it is important to also assess contextual factors like alternative psychiatric diagnoses, a medical illness, or the side-effects of medication [49]. We used the PHQ-9 as a measure of depressive symptoms, and not as a measure of depressive disorder. Symptom criteria for depression overlap symptoms of cancer and other comorbidities, e.g. fatigue, poor appetite and sleep problems [50]. In patients with increased symptom burden, exclusion of somatic symptom criteria in the PHQ-9 may reduce the likelihood of being false positive categorized as depressed [42]. In this study, the four somatic depression symptoms in the PHQ-9 were excluded to avoid overlap with the MDASI. We found a significant association between higher levels of depressive symptoms and higher MDASI sum score.

Comorbidities were significantly associated with an increased MDASI sum score in our study. A cross-sectional study from the USA [51] found that symptom scores on all domains were significantly worse in people with multiple sclerosis than in the general population, also after adjusting for age and sex. Similarly, a study found that patients with systemic lupus erythematosus had symptom scores that indicated poorer average health status compared with the general population [52]. A survey among patients with type 2 diabetes in primary care found that the study population reported more problems with physical functioning and pain compared to the general population [53]. This illustrates the importance of reference values when comparing differences in daily function for populations with a specific disease and the general population. It is important to adjust for comorbidities when comparing different populations in terms of level of symptom scores. This also applies to other variables that significantly affect the symptom level, like depression and education. The independent variables included in the multiple regression model explained 45% of the variance in MDASI sum score. By controlling for relevant associated factors, potential bias is likely to be reduced.

Comorbidity, depression and MDASI sum score were significantly associated with the interference items general activity and work. Depression and MDASI sum score were negatively associated with enjoyment, mood and relations to other people. Bruusgaard et.al [4]. found a strong linear association between the number of self-reported symptoms and decreased functional status in the Norwegian Ullensaker population study. Anxiety and depression were symptoms that had substantially higher explanatory power on functional status than other symptoms [4]. This in in agreement with the findings in our study, with depressive symptoms being associated with all interference items but walking. These findings indicate that interference is influenced by other variables than just symptoms. This does not only apply to the emotional domains like enjoyment and mood, but also to the more functional ones like work and general activity.

Limitations

The randomly drawn sample was assumed to be representative of the general Norwegian population with respect to age, sex, and place of living. However, only 36% of the sample responded to the survey. Compared to collection of Norwegian reference values for the SF-36 in 1996 and 2002 this response rate was low [32]. The decline in response rates from 67% in 1996 to 36% in 2015 is in line with other postal surveys [3, 23, 54, 55]. Another Norwegian study found that health-related quality of life was relatively stable in two cross-sectional studies over an 8 year period despite the response rate being 68% in the first study and 35% in the second [56]. Surveys are used to describe large populations, and high response rates are valued to reduce the risk of bias. However, nonresponse bias is only indirectly related to nonresponse rates and there is little empirical support for the notion that low response rates are more prone to nonresponse bias than samples with higher response rates [57]. The fact that response rates in sample surveys in general have declined over the past decades is challenging for population studies [57]. Innovation in epidemiologic studies should involve development of recruitment techniques that optimize participation [58]. A large Danish population study from 2015 [2] used web-based questionnaires and had a response rate of

52%. In our study, the paper-based questionnaire was not available in an electronic version.

The fact that a large proportion of the respondents had university level education may be considered as a potential bias regarding the representativity of the sample. According to Statistics Norway [59] 32% of the Norwegian population had higher education in 2015, 41% had finished high school and 27% had finished elementary school, corresponding to 46, 37 and 17% respectively in our study. This should be considered when using the reference values in groups with low education.

When comparing the sample to the actual composition of the Norwegian population, 15% of the population was 67 years or above in 2015, while 27% of the responders were in the same age group [32]. About 21% of the Norwegian population was between 18 and 29 years, while only 5% of this age group participated in the survey. The opposite pattern was seen for the older population. Thus, it is highly likely that the high mean scores for symptoms in the youngest age group are not entirely representative for the general population of the same age. The relatively high symptom scores in the youngest age group compared to the older age groups may indicate an unhealthy bias in the youngest age group and a healthy bias among the older age groups. Taken together, these factors suggest that the reference values might be biased due to selection among the youngest participants. Regrettably, our data did not permit further analyses to illuminate this.

In accordance with other frequently used PROMs-questionnaires, the MDASI assesses the most common cancer-related symptoms. The MDASI has been translated into and validated in several languages [27, 60, 61]. However, the MDASI has not gone through a complete psychometric validation in a Norwegian cancer population. Following our study this may be a natural next step, as the symptoms of the MDASI and the fact that it specifically assesses the interference with daily living caused by these symptoms, makes it a highly relevant tool for patient-centered care and follow-up. Such a study should also include other questionnaires- such as the Quality of Life Questionnaire-Core 30 (QLQ-C30) [62] and the Brief Pain Inventory (BPI) [63], both which are validated and frequently used in Norway. However, given that the MDASI symptoms are common among cancer patients, and that the answering format is similar to other tools, we assume the Norwegian MDASI to have both high face validity and convergent validity, as is also shown in studies from other countries [60, 61, 64].

Conclusions

This study provides the first Norwegian reference values for the MDASI. The presence of one or more comorbidities, increased levels of depressive symptoms and lower level of education were significantly associated with higher MDASI sum score. These covariates must be controlled for when using the reference values.

Abbreviations

BPI: Brief Pain Inventory; EORTC: European Organisation for Research and Treatment of Cancer; FQ: Fatigue Questionnaire; HUNT: Health Study of Nord-Trøndelag; MDASI: M.D. Anderson Symptom Inventory; PHQ-9: Patient Health Questionnaire 9; PROMs: Patient Reported Outcome Measures; PROs: Patient Reported Outcomes; REC: Regional Committee for Medical and Health Research Ethics; SCQ: Self-Administered Comorbidity Questionnaire; SD: Standard Deviation; SF-36: Short-Form Health Survey-36; QLQ-C30: Quality of Life Questionnaire-Core 30; QoL: Quality of Life

Acknowledgements

Not applicable.

Authors' contributions

Conzeptualization: JHL, MJH. Methodology: MJH, KSG, JHL. Formal analysis: HK, ØS. Project administration: MJH, KSG, JHL. Writing original draft: HK. Supervision, writing review and editing: JHL, KSG, SK, ØS, CEK, MJH. All authors read and approved the final manuscript.

Funding

This study received funding from Holes legat, the Cancer Trust, St. Olavs hospital, Trondheim University Hospital (Project No. 35715), the Norwegian Hospital Foundation (Project No. 335007), Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, and St. Olavs hospital, Trondheim University Hospital.

Availability of data and materials

The dataset used and analysed during the current study is available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was performed according to the rules of the Helsinki declaration. All respondents received written information about the study. Return of the questionnaires was taken to indicate written, informed consent. The Regional Committee for Medical and Health Research Ethics (REC) South East Norway approved the survey (2014/1172).

Consent for publication

Not applicable.

Competing interests

HK, KSG, ØS, CEK and MJH have no declared conflicts of interests. Eir Solutions AS was established in 2015 with SK, JHL, and NTNU Technology Transfer AS as shareholders. No income, dividend, or financial benefits are related to the work presented here nor in relation to Eir in any way.

Author details

[1]European Palliative Care Research Centre (PRC), Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, and St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway. [2]Cancer Clinic, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway. [3]Regional Advisory Unit in Palliative Care, Department of Oncology, Oslo University Hospital, Oslo, Norway. [4]European Palliative Care Research Centre (PRC), Department of Oncology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [5]Institute of Basic Medical Sciences, Faculty of Medicine, University of Oslo, Oslo, Norway. [6]National advisory unit for late effects after cancer treatment, Oslo University Hospital, and University of Oslo, Oslo, Norway. [7]Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, Trondheim, Norway.

Received: 17 December 2019 Accepted: 12 June 2020
Published online: 23 June 2020

References

1. Food and Drug Administration USDoHaHS. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical product Development to

Support Labeling Claims; 2009. Available from: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm193282.pdf.
2. Elnegaard S, Andersen RS, Pedersen AF, Larsen PV, Sondergaard J, Rasmussen S, et al. Self-reported symptoms and healthcare seeking in the general population--exploring "the symptom iceberg". BMC Public Health. 2015;15:685.
3. McAteer A, Elliott AM, Hannaford PC. Ascertaining the size of the symptom iceberg in a UK-wide community-based survey. Br J Gen Pract. 2011;61(582):e1–11.
4. Bruusgaard D, Tschudi-Madsen H, Ihlebaek C, Kamaleri Y, Natvig B. Symptom load and functional status: results from the Ullensaker population study. BMC Public Health. 2012;12:1085.
5. Kjeldsberg M, Tschudi-Madsen H, Dalen I, Straand J, Bruusgaard D, Natvig B. Symptom reporting in a general population in Norway: results from the Ullensaker study. Scand J Prim Health Care. 2013;31(1):36–42.
6. Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. Br J Cancer. 2009;101(Suppl 2):S80–6.
7. Creed FH, Davies I, Jackson J, Littlewood A, Chew-Graham C, Tomenson B, et al. The epidemiology of multiple somatic symptoms. J Psychosom Res. 2012;72(4):311–7.
8. Food and Drug Administration. Guidelines for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Rockville MD: U.S. Department of Health and Human Services; 2006.
9. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res. 2012;21(8):1305–14.
10. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol. 2004;22(4):714–24.
11. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. BMC Health Serv Res. 2013;13:211.
12. Porter ME. What is value in health care? N Engl J Med. 2010;363(26):2477–81.
13. Kaasa S, Loge JH, Aapro M, Albreht T, Anderson R, Bruera E, et al. Integration of oncology and palliative care: a lancet oncology commission. Lancet Oncol. 2018;19(11):e588–653.
14. Jordan K, Aapro M, Kaasa S, Ripamonti CI, Scotté F, Strasser F, et al. European Society for Medical Oncology (ESMO) position paper on supportive and palliative care. Ann Oncol. 2018;29(1):36–43.
15. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. Jama. 2017;318(2):197–8.
16. Denis F, Basch E, Septans A-L, Bennouna J, Urban T, Dueck AC, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. Jama. 2019;321(3):306–7.
17. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst. 2017;109(9). https://doi.org/10.1093/jnci/djx029.
18. van de Poll-Franse LV, Mols F, Gundy CM, Creutzberg CL, Nout RA, Verdonck-de Leeuw IM, et al. Normative data for the EORTC QLQ-C30 and EORTC-sexuality items in the general Dutch population. Eur J Cancer. 2011;47(5):667–75.
19. Jensen RE, Potosky AL, Moinpour CM, Lobo T, Cella D, Hahn EA, et al. United States population-based estimates of patient-reported outcomes measurement information system symptom and functional status reference values for individuals with Cancer. J Clin Oncol. 2017;35(17):1913–20.
20. Hinz A, Singer S, Brahler E. European reference values for the quality of life questionnaire EORTC QLQ-C30: Results of a German investigation and a summarizing analysis of six European general population normative studies. Acta Oncologica (Stockholm, Sweden). 2014;53(7):958–65.
21. Nolte S, Liegl G, Petersen MA, Aaronson NK, Costantini A, Fayers PM, et al. General population normative data for the EORTC QLQ-C30 health-related quality of life questionnaire based on 15,386 persons across 13 European countries, Canada and the unites states. Eur J Cancer. 2019;107:153–63.
22. Brucker PS, Yost K, Cashy J, Webster K, Cella D. General population and cancer patient norms for the functional assessment of Cancer therapy-general (FACT-G). Eval Health Prof. 2005;28(2):192–211.
23. Hjermstad MJ, Fayers PM, Bjordal K, Kaasa S. Using reference data on quality of life--the importance of adjusting for age and gender, exemplified by the EORTC QLQ-C30 (+3). Eur J Cancer. 1998;34(9):1381–9.
24. Scott N, Fayers P, Aaronson N, Bottomley A, de Graeff A, Groenvold M, et al. EORTC QLQ-C30 reference values manual. 2nd ed. Brussels: EORTC Quality of Life Group; 2008. p. 427.
25. Champion VL, Wagner LI, Monahan PO, Daggy J, Smith L, Cohee A, et al. Comparison of younger and older breast cancer survivors and age-matched controls on specific and overall quality of life domains. Cancer. 2014;120(15):2237–46.
26. Bøhn S-KH, Thorsen L, Kiserud CE, Fosså SD, Lie HC, Loge JH, et al. Chronic fatigue and associated factors among long-term survivors of cancers in young adulthood. Acta Oncol. 2019;58(5):753–62.
27. Cleeland CS, Mendoza TR, Wang XS, Chou C, Harle MT, Morrissey M, et al. Assessing symptom distress in cancer patients: the M.D. Anderson symptom inventory. Cancer. 2000;89(7):1634–46.
28. Kirkova J, Davis MP, Walsh D, Tiernan E, O'Leary N, LeGrand SB, et al. Cancer symptom assessment instruments: a systematic review. J Clin Oncol. 2006;24(9):1459–73.
29. Reilly CM, Bruner DW, Mitchell SA, Minasian LM, Basch E, Dueck AC, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer. 2013;21(6):1525–50.
30. Bring. (cited 2019 October 31st); Available from: https://www.bring.no/.
31. Ware J. The SF-36 health survey. 2nd ed. Philadelphia: Lippincott Raven; 1996.
32. Jacobsen EL, Bye A, Aass N, Fossa SD, Grotmol KS, Kaasa S, et al. Norwegian reference values for the short-form health survey 36: development over time. Qual Life Res. 2018;27(5):1201–12.
33. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. J Psychosom Res. 1993;37(2):147–53.
34. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13.
35. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. patient health questionnaire. Jama. 1999;282(18):1737–44.
36. Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The self-administered comorbidity questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum. 2003;49(2):156–63.
37. Mendoza TR, Wang XS, Lu C, Palos GR, Liao Z, Mobley GM, et al. Measuring the symptom burden of lung cancer: the validity and utility of the lung cancer module of the M. D Anderson symptom inventory. Oncologist. 2011;16(2):217–27.
38. Wang XS, Williams LA, Eng C, Mendoza TR, Shah NA, Kirkendoll KJ, et al. Validation and application of a module of the M. D. Anderson symptom inventory for measuring multiple symptoms in patients with gastrointestinal cancer (the MDASI-GI). Cancer. 2010;116(8):2053–63.
39. Rokstad K, Straand J, Sandvik H. Patient encounters in general practice. An epidemiological survey in more and Romsdal. Tidsskrift for den Norske laegeforening: tidsskrift for praktisk medicin, ny raekke. 1997;117(5):659–64.
40. Dewolf L, Koller M, Velikova G, Johnson C, Scott N, Bottomley A. EORTC Quality of Life Group Translation Procedure. 3rd ed; 2009.
41. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). Arlington: American Psychiatric Publishing; 2013.
42. Lie HC, Hjermstad MJ, Fayers P, Finset A, Kaasa S, Loge JH. Depression in advanced cancer--assessment challenges and associations with disease load. J Affect Disord. 2015;173:176–84.
43. Landmark T, Romundstad P, Dale O, Borchgrevink PC, Vatten L, Kaasa S. Chronic pain: one year prevalence and associated characteristics (the HUNT pain study). Scand J Pain. 2013;4(4):182–7.
44. Landmark T, Dale O, Romundstad P, Woodhouse A, Kaasa S, Borchgrevink PC. Development and course of chronic pain over 4 years in the general population: The HUNT pain study. Eur J Pain (London, England). 2018;22(9):1606–16.
45. Jensen MK, Sjogren P, Ekholm O, Rasmussen NK, Eriksen J. Identifying a long-term/chronic, non-cancer pain population using a one-dimensional

verbal pain rating scale: an epidemiological study. Eur J Pain (London, England). 2004;8(2):145–52.
46. Bardel A, Wallander MA, Wallman T, Rosengren A, Johansson S, Eriksson H, et al. Age and sex related self-reported symptoms in a general population across 30 years: patterns of reporting and secular trend. PLoS One. 2019; 14(2):e0211532.
47. Haug TT, Mykletun A, Dahl AA. The association between anxiety, depression, and somatic symptoms in a large population: the HUNT-II study. Psychosom Med. 2004;66(6):845–51.
48. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health. 2013;34:119–38.
49. Malhi GS, Mann JJ. Depression. Lancet. 2018;392(10161):2299–312.
50. Rayner L, Lee W, Price A, Monroe B, Sykes N, Hansford P, et al. The clinical epidemiology of depression in palliative care and the predictive value of somatic symptoms: cross-sectional survey with four-week follow-up. Palliat Med. 2011;25(3):229–41.
51. Amtmann D, Bamer AM, Kim J, Chung H, Salem R. People with multiple sclerosis report significantly worse symptoms and health related quality of life than the US general population as measured by PROMIS and NeuroQoL outcome measures. Disabil Health J. 2018;11(1):99–107.
52. Mahieu MA, Ahn GE, Chmiel JS, Dunlop DD, Helenowski IB, Semanik P, et al. Fatigue, patient reported outcomes, and objective measurement of physical activity in systemic lupus erythematosus. Lupus. 2016;25(11):1190–9.
53. Homco J, Rodriguez K, Bardach DR, Hahn EA, Morton S, Anderson D, et al. Variation and change over time in PROMIS-29 survey results among primary care patients with type 2 diabetes. J Patient-Centered Res Rev. 2019;6(2): 135–47.
54. Christensen AI, Ekholm O, Kristensen PL, Larsen FB, Vinding AL, Glumer C, et al. The effect of multiple reminders on response patterns in a Danish health survey. Eur J Pub Health. 2015;25(1):156–61.
55. Mannetje A, Eng A, Douwes J, Ellison-Loschmann L, McLean D, Pearce N. Determinants of non-response in an occupational exposure and health survey in New Zealand. Aust N Z J Public Health. 2011;35(3):256–63.
56. Fossa SD, Hess SL, Dahl AA, Hjermstad MJ, Veenstra M. Stability of health-related quality of life in the Norwegian general population and impact of chronic morbidity in individuals with and without a cancer diagnosis. Acta Oncologica (Stockholm, Sweden). 2007;46(4):452–61.
57. Groves RM. Nonresponse rates and nonresponse Bias in household surveys. Public Opin Q. 2006;70(5):646–75.
58. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol. 2007;17(9):643–53.
59. Statistics Norway; 2020. Available from: https://www.ssb.no/befolkning.
60. Guirimand F, Buyck JF, Lauwers-Allot E, Revnik J, Kerguen T, Aegerter P, et al. Cancer-related symptom assessment in France: validation of the French M. D. Anderson symptom inventory. J Pain Symptom Manag. 2010; 39(4):721–33.
61. Wang XS, Wang Y, Guo H, Mendoza TR, Hao XS, Cleeland CS. Chinese version of the M. D. Anderson symptom inventory: validation and application of symptom measurement in cancer patients. Cancer. 2004; 101(8):1890–901.
62. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst. 1993;85(5):365–76.
63. Cleeland CS, Ryan KM. Pain assessment: global use of the brief pain inventory. Ann Acad Med Singap. 1994;23(2):129–38.
64. Ivanova MO, Ionova TI, Kalyadina SA, Uspenskaya OS, Kishtovich AV, Guo H, et al. Cancer-related symptom assessment in Russia: validation and utility of the Russian M. D. Anderson symptom inventory. J Pain Symptom Manag. 2005;30(5):443–53.

Publisher's Note

Paper II

original report

Development of EirV3: A Computer-Based Tool for Patient-Reported Outcome Measures in Cancer

abstract

Purpose Immediate transfer of patient-reported outcome measures (PROMs) for use in medical consultations is facilitated by electronic assessments. We aimed to describe the rationale and development of Eir version 3 (EirV3), a computer-based symptom assessment tool for cancer, with emphasis on content and user-friendliness.

Methods EirV3's specifications and content were developed through multiprofessional, stepwise, and iterative processes (from 2013 to 2016), with literature reviews on traditional and electronic assessment and classification methods, formative iterative usability tests with end-users, and assessment of patient preferences for paper versus electronic assessments.

Results EirV3 has the following two modules: Eir-Patient for PROMs registration on tablets and Eir-Doctor for presentation of PROMs in a user-friendly interface on computers. Eir-Patient starts with 19 common cancer symptoms followed by specific, in-depth questions for endorsed symptoms. The pain section includes a body map for pain location and intensity, whereas physical functioning, nutritional intake, and well-being are standard questions for all. Data are wirelessly transferred to Eir-Doctor. Symptoms with intensity scores ≥ 3 (on a 0 to 10 scale) are marked in red, with brighter colors corresponding to higher intensity, and supplemented with graphs displaying symptom development over time. Usability results showed that patients and health care providers found EirV3 to be intuitive, easy to use, and relevant. When comparing PROM assessments on paper versus tablets (n = 114), 19% of patients preferred paper, 41% preferred tablets, and 40% had no preference. Median intraclass correlation coefficient between paper and tablets (0.815) was excellent.

Conclusion Iterative test rounds followed by continuous improvements led to a user-friendly, applicable symptom assessment tool, EirV3, developed for and by end-users. EirV3 is undergoing international testing of clinical and cross-cultural adaptability.

Clin Cancer Inform. © 2017 by American Society of Clinical Oncology

Hilde Krogstad
Cinzia Brunelli
Kari Sand
Eivind Andersen
Herish Garresori
Tarje Halvorsen
Ellinor C. Haukland
Frode Jordal
Stein Kaasa
Jon Håvard Loge
Erik Torbjørn Løhre
Sunil X. Raj
Marianne Jensen Hjermstad

Author affiliations appear at the end of this article.

H.K. and C.B. contributed equally to this work.

Corresponding author: Hilde Krogstad, MD, European Palliative Care Research Centre, Department of Clinical and Molecular Medicine, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology and St.Olavs Hospital, Trondheim University Hospital, Trondheim 7491, Norway; e-mail: hilde.krogstad@ntnu.no.

INTRODUCTION

Systematic use of patient-reported outcome measures (PROMs) in clinical practice is essential for optimal patient care.[1-3] The recognition of PROMs as independent outcomes in cancer[4,5] is consolidated by the CONSORT Patient-Reported Outcomes Extension Statement developed to improve the reporting of PROMs on patients' evaluation of symptoms, functioning, and quality of life.[5]

Benefits of routine PROM registrations have been reported, such as improved patient-physician communication[6-8] and better patient well-being.[8,9] Regular PROM assessment during treatment with immediate feedback to clinicians has proven to be efficient in informing clinicians about symptoms and problems[7] and guiding treatment decisions.[2,10,11] A recent review reported improved symptom management and higher patient satisfaction when using PROMs in the clinical consultation, because this made physicians aware of symptoms that had not been discussed before.[12,13]

Despite these findings, systematic collection and use of PROMs in clinical oncology remain uncommon.[2,11,14,15] The most common barriers are logistical problems, cumbersome administration, and time constraints.[11,14-17] These barriers may be overcome by health information technology and Web-based communication now widely available. Indeed, electronic data collection permits dynamic symptom assessment (ie, tailored questions for individual patients based on the

patients' previous responses). This results in fewer repetitive questions and reduces patient burden by avoiding long and cumbersome questionnaires. In addition, Web-based technology permits immediate transfer of patients' responses to the attending physician's desktop. When used alongside clinical data, the follow-up of patients may be more comprehensive, especially for patients who are not hospitalized.

Our research group in the European Association of Palliative Care (EAPC) Research Network[18] and the European Palliative Care Research Centre (PRC)[19] has developed several electronic symptom assessment tools over the past decade[20-26] (Appendix). Our experiences led to the Eir Project in 2013. The long-term aim is to integrate PROMs and clinical data in a user-friendly software available on all platforms for use in treatment of adult patients with cancer across disease stages and settings.

This article describes the stepwise development process toward the current Eir version, version 3 (EirV3), which has the following two modules: Eir-Patient and Eir-Doctor. More specifically, qualitative and quantitative results from iterative test rounds are presented, focusing on the rationale behind the requirements, contents, adaptation of technical specifications, usability, patient preferences, and preference for using paper or electronic versions.

METHODS

Eir has been designed following expert-driven and user-driven approaches. The first step, selection of content, is based on literature searches, expert opinions, clinical experience, and evidence-based guidelines for symptom management,[27-33] guided by iterative formative tests of preferences, needs, and skills of the end-users—patients and health care providers (HCPs). Throughout development, regular meetings and discussions were carried out in the following two main working groups: an international palliative care (PC) expert panel, consisting of 26 PC experts from Italy, Norway, the United Kingdom, Denmark, Spain, and Germany experienced in clinical oncology/PC, symptom assessment/classification, questionnaire development, and PC research and recruited from the European Palliative Care Research Centre and EAPC Research Network, and a Norwegian core working group (n = 9 to 15) consisting of experienced oncologists, PC physicians, researchers, interaction designers, graphic designers, and software developers. Altogether, results from the international meetings and local workshops (Appendix) led to the recommendations guiding the subsequent Eir development (Table 1) and to the final decisions on the content, based on reviews, guidelines, and evidence at the time (Table 2).[33-42]

Formative Usability Testing

The second step in Eir development was formative usability testing, an iterative design process conducted to detect weaknesses in the structure and content and software bugs and to problem solve issues based on end-users' input.[43] The aim of the usability tests was to obtain the opinion of patients and HCPs regarding ease of navigation, clarity of instructions, and content relevance in Eir-Patient and Eir-Doctor (Table 3; Appendix).

Equivalence Between Electronic and Paper PROM Assessments

In 2016, a comparative study was carried out among 114 patients with cancer at six Norwegian hospitals to examine agreement between PROM assessments on tablets and paper and to assess patients' preference for either method. Patients rated the intensity of 19 symptoms in EirV3. The order of assessment, either paper or tablet first, was randomly assigned, with 30 minutes between assessments. Intraclass correlation coefficients (ICCs) based on a two-way mixed effect analysis of variance, single measure and absolute agreement,[44] were used to examine agreement of tablet and paper scores. According to interpretation guidelines,[45] an ICC > 0.75 indicates excellent agreement.

Technical Specifications and Data Safety

EirV3 is a Web site using standard HTML5, CSS3, and Javascript and designed for ease of use and touch-based navigation. This allows the system to run on any hardware with a modern Web browser, including tablets, cell phones, laptops, workstations, and public terminals. It is designed for Windows Server using IIS, but also Windows Azure. The default database for storage is Microsoft SQL Server, but Azure Blob storage and document databases also work well.

Ethical Considerations

Confidentiality issues and adherence to all regulations regarding the registration, transfer, handling, and storage of data were major issues during the development process. Data communication between the device used for data entry and the storage server is secured using HTTPS over SSL. Verification of the patient's identity is ensured

Table 1. Requirements and Methods That Guided the Eir Development Process

Requirements*	Methods
Mimic a clinical consultation regarding content	Use a hierarchical, logical structure for questions Use photos of humans for body pain markings
Cover the most common cancer-related symptoms	Select symptoms based on literature reviews, clinical experience
Minimize ad hoc formulations and questions	Select items from well-validated tools If not available, reach consensus in international expert panel
Dynamic, flexible, and tailored to the individual patient	Define screening questions that guide subsequent questions if endorsed
Applicable in multiple settings (hospital, ambulatory, home care)	Ensure software compatibility with multiple platforms
User-friendly	Perform iterative usability testing in different patient samples (eg, diagnoses, settings, fit and frail)
Feasible	Ensure easy handling, self-explanatory layout, and immediate back-up
Immediate transfer of all PROMs from Eir-Patient to Eir-Doctor	Ensure a design in Eir-Doctor that immediately presents all PROMs on the same screen in Eir-Doctor, adapt for Wi-Fi use
Longitudinal presentation of patient data	Programmed with reader-friendly diagrams, charts, and output in Eir-Doctor
Safe transfer and storage of data	Collaborate with IT specialists and data protection supervisors to comply with all safety and confidentiality regulations
Applicable across cultures	Use scales and items from well-validated tools and questionnaires, available in multiple languages Perform international testing
Output reports on patient and group level	Enable data extraction as separate files, prints, and so on
Compatibility with existing databases	Incorporate Eir into electronic patient records

Abbreviation: IT, information technology; PROM, patient-related outcome measure.

*Consensus on these requirements was reached based on literature searches, expert opinions, clinical experience, and evidence-based guidelines for symptom management, as well as workshops, international expert meetings, and usability testing.

using token-based authentication, with support for authentication protocols (eg, OAuth, OpenId, and SAML2.0). Patient data are stored on secure servers hosted by each clinic. Data are encrypted using Advanced Encryption Standard requiring an encryption key to access the database. Access to patients' PROMs in Eir-Doctor is password protected.

The Regional Committee for Medical and Health Research Ethics Central Norway approved the comparative study, confirming that formal approval was not required for the usability tests (REK-2014/212, REK-2015/185).

RESULTS

Eir-PatientV3

Eir-Patient addresses all 12 symptoms in the EAPC Basic Dataset[31] (ie, pain, tiredness, drowsiness, nausea, reduced appetite, breathlessness, depression, anxiety, well-being, sleep, constipation, vomiting), supplemented by four items particularly related to chemotherapy (ie, numbness in hands or feet, diarrhea, mouth sores, dry mouth) and another four items adjusted from the Patient-Generated Subjective Global Assessment[42] for assessment of nutritional status (ie, altered sense of taste, altered sense of smell, problems swallowing, early satiety) and physical activity.

Dynamic Symptom Assessment

For Eir to be dynamic and patient tailored (Table 1), a symptom assessment hierarchy was developed as per requirements in the working groups. The opening question mimics a common start of a clinical consultation with a general question about the patient's well-being today (Fig 1). Then there is a symptom screening section (Level 0) followed by intensity ratings of all endorsed symptoms (Level 1) and specific questions on symptom characteristics (Level 2; Table 2).[33-41,46] To keep the number of questions to a minimum, it was decided to add follow-up questions only if the international expert panel considered this to be of clinical relevance. In the last section, questions on height, current weight, food intake, and current level of physical functioning are for all patients.

Eir-DoctorV3

PROMs reported on the tablet by the patient are immediately available in Eir-Doctor to focus the patient-physician communication on symptoms that need attention and treatment. The Eir-Doctor opening screen displays symptom scores in descending order of intensity from high to low, with scores ≥ 3 in red, indicating clinical

Table 2. The Dynamic Structure for Symptom Assessment in Eir

Symptom	Level 0 Screening	Level 1 Intensity	Level 2 Characterization	Source
Well-being	—	0-10 NRS	None	
Pain	Yes/no	0-10 NRS	Pain location: body map Neuropathic pain: verbal descriptors Breakthrough pain: intensity of pain flares, triggering factors	Kaasa et al[33]; Brunelli et al[34]; Zeppetella and Davies[35]; Portenoy and Hagen[36]; Hagen et al 2008[37]
Tiredness	Yes/no	0-10 NRS	None	
Drowsiness	Yes/no	0-10 NRS	None	
Nausea	Yes/no	0-10 NRS	None	
Reduced appetite	Yes/no	0-10 NRS	None	
Breathlessness	Yes/no	0-10 NRS	Shortness of breath at rest	
Depression	Yes/no	0-10 NRS	Patient Health Questionnaire-9	Kroenke et al[38]
Anxiety	Yes/no	0-10 NRS	General anxiety disorder-2	Kroenke et al[39]
Insomnia	Yes/no	0-10 NRS	Problems falling asleep Problems sleeping all night Whether insomnia interferes with daily activities	Based on Insomnia Severity Index[40]
Constipation	Yes/no	0-10 NRS	Last bowel movement	CTCAE[41]
Vomiting	Yes/no	0-10 NRS	Frequency	CTCAE
Numbness in fingers or toes	Yes/no	0-10 NRS	None	
Diarrhea	Yes/no	0-10 NRS	Frequency Blood in stools	CTCAE
Mouth sores	Yes/no	0-10 NRS	None	
Dry mouth	Yes/no	0-10 NRS	None	
Altered sense of taste	Yes/no	0-10 NRS	None	
Altered sense of smell	Yes/no	0-10 NRS	None	
Problems swallowing	Yes/no	0-10 NRS	None	
Early satiety	Yes/no	0-10 NRS	None	

Abbreviations: CTCAE, Common Terminology Criteria for Adverse Events; NRS, numerical rating scale.

significance (Fig 2). A graph on the right shows symptom intensity over time, if available. Well-being, physical activity, nutritional intake, and weight are shown on top, because these are considered key factors in patient-centered treatment.

Formative Usability Tests of Eir-Patient and Eir-Doctor

Patients were recruited from the cancer outpatient clinic (Table 3) and were heterogeneous with respect to age, sex, cancer diagnosis, and treatment intent (curative, adjuvant, or palliative). Overall, they had few problems using Eir-Patient and appreciated that the physician received updated information about their clinical condition.

The questions per se posed few difficulties for patients, although some patients with related symptoms (eg, tiredness, lack of appetite, and depression) found some of the follow-up questions to be overlapping. Most of the outpatients had a limited number of symptoms and thus relatively few questions to which to respond. As expected, using EirV3 was perceived as more demanding for PC patients with a high symptom burden compared with patients who were in a better physical condition.

Observations of patients using Eir revealed that they did not notice all elements on the screen at a time; they focused mainly on the middle and inadvertently skipped items on the left and right sides. Even when they skipped the instructions on the screen, patients found it easy to navigate in Eir (eg, moving forward or backward, finding the right answer, and having the answer registered). However, the latter posed some difficulties for patients who either did not position the tablet in the right angle or who had fingers that were too dry or too cold to obtain sufficient pressure on their

Table 3. Iterative Usability Testing of Eir

Time	Type of Test	Participants	Procedures	Main Findings and Subsequent Changes
September-October 2013	Test of a computerized pain body map	Outpatients with cancer (n = 10)	Observations of patients using different way of marking pain on a tablet Subsequent debriefing interviews Field notes	Finding: Shading on the pain area of the body map did not seem intuitive to patients. They preferred to tap or press the relevant area Change: Tapping or pressing the area of the pain location was sufficient for the area to be marked in red.
November 2013	Test of Eir-PatientV1	Outpatients with cancer (n = 7)	Observations of patient completing Eir-PatientV1 Think-aloud strategy Subsequent debriefing interviews Field notes	Finding: When patients did not find a relevant response alternative, they tended to choose another. Change: The alternative "None of these" was added. Finding: Some patients did not manage to get their taps registered. Change: Short and long clicks or taps, as well as swipes, are registered. Finding: The zooming function of the body map and too many navigation buttons on the same screen image were confusing. Change: The layout was improved, and the number of navigation buttons was reduced. Finding: Most patients did not read the instructions regarding completion. Change: Instructions were available by clicking on a Help button.
January-May 2014	Changes made in content, functionality, and layout; development of EirV2			
May 2014	Pilot test of Eir-PatientV2	Outpatients with cancer (n = 7)	Observations of patient completing Eir-PatientV2 Think-aloud method Subsequent debriefing interviews Field notes	Finding: Patients had trouble understanding that they could not mark more than 1 painful area on the body map at the time. Change: An information page was added before the pain section. Finding: If patients had trembling hands, they accidentally double-clicked on the Next button and skipped a page. Change: Rapid double-clicks are registered as 1 tap (1 registration).
June-December 2014	Clinical test of Eir-PatientV2 and Eir-DoctorV2*	Outpatients with cancer (n = 42); physicians in cancer department (n = 8)	Observations of patient completing Eir-PatientV2 Think-aloud strategy Subsequent debriefing interviews Observations of physicians using Eir-DoctorV2 in consultation Regular group discussions with physicians during the test period Field notes	Findings for Eir-Patient Finding: Some elements on the screen went unnoticed; some elements were misunderstood; the elements in the middle of the screen were read first. Change: The number of elements on each screen was reduced. Question and response alternatives were placed in the middle. Finding: Taps were not registered as a result of cold/dry fingers or long nails. Change: Optional use of stylus. Finding: Patients accidentally quit Eir and had to start all over. Change: The tablets were locked to Eir.

(Continued on following page)

Table 3. Iterative Usability Testing of Eir (Continued)

Time	Type of Test	Participants	Procedures	Main Findings and Subsequent Changes
				Findings for Eir-Doctor Finding: Physicians misunderstood the summarized information of well-being, nutrition, and physical functioning on the opening screen and did not intuitively understand (or remember) what questions the patient had answered. Change: Extra information was added to clarify what information had been given by the patient. Finding: Detailed information on well-being, nutrition, and physical functioning was left out of Eir-Doctor. Change: All these variables were presented in 1 click. Finding: The list of symptoms could be difficult to follow if the patient has registered several symptoms. Change: More sorting functions for symptoms added (eg, high to low on intensity and development of intensity since last registration).
January-May 2015	Changes made in content, functionality, and layout; development of EirV3			
May-June 2015	Pilot test of Eir-PatientV3	Outpatients with cancer (n = 9)	Observations of patients using Eir-PatientV3 Think-aloud method Subsequent debriefing interviews Field notes	Findings: If the patient had more than 1 painful area, the second pain section was initiated with a confusing question. Change: New question added. Finding: Patients found the question about physical function confusing, because this item had too many response alternatives that were not mutually exclusive. Changes: Question was changed to a validated question on physical function with fewer response options

*Changes in Eir-Patient-versions led to immediate changes in the corresponding Eir-Doctor-versions, thus numbering of versions is identical.

touch for registration. The pain body map with zoom functions and related follow-up questions turned out to be the most challenging part of Eir-Patient. Difficulties were related to marking of the painful area, primarily because patients tried to mark multiple areas at a time, even if instructions told them not to. They also found some of the follow-up pain questions confusing, particularly those related to pain descriptors (eg, "burning" and "pins and needles"), whereas some patients missed an option for marking radiating pain. The technologic features and explanations were revised accordingly in EirV3. All follow-up questions applied to each pain site, and the number of elements on each screen was reduced (eg, by skipping some of the instructions for navigation or answers, dropping a progress bar, and consistently centering the relevant items on the screen). Increasing the user-friendliness was also pursued by adding a "Help" function; adding the response alternative "None of these," as appropriate; and accepting different types of taps, swipes, and drags for registration.

When testing Eir-Doctor, physicians defined the graphical presentation of symptom trajectories as a key factor to monitor effect of treatments. They also mentioned that the current display in EirV3, which resulted from iterative rounds of feedback from clinical testing, made them aware of symptoms they had not known troubled the patient. Physicians found it useful to start the consultations with the list of symptoms and intensity scores. Because the patient's symptoms are ordered by intensity, the list and order of

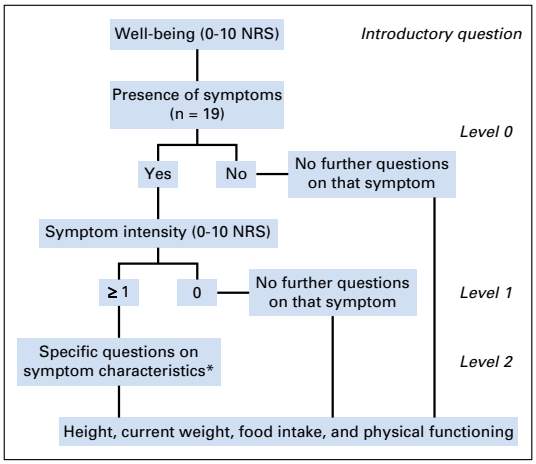


Fig 1. Symptom assessment algorithm. NRS, numerical rating scale. (*) Details in Table 2.

symptoms vary from one patient to the other. Some physicians preferred a fixed order, whereas others preferred high intensity as the default. All physicians regretted the fact that EirV3 is not yet integrated into the electronic patient records, because this would enhance the clinical decision making by combining individual patient data from different sources.[34]

Equivalence Between Electronic- and Paper-Based Assessment

Of the 114 patients included in the paper and pencil versus electronic assessments comparative study, 110 patients (97%) completed both versions, 59 patients (54%) on tablets first and 51 patients (46%) on paper first. Mean age was 64.5 years (range, 27 to 86 years), and median Karnofsky performance score was 90 (range, 50 to 100). GI cancer was most common (47%), followed by prostate cancer (10%), breast cancer (9%), and malignant melanoma (9%). Eighty-nine percent of patients had metastatic disease. Overall, the median ICC was high (0.81; Table 4), with excellent values (> 0.75) for 15 of the 19 items (range, 0.64 [vomiting] to 0.92 [tiredness]). Overall, 41% of the patients preferred assessment on tablets, 19% preferred paper, and 40% had no preference. Preference for electronic assessment was more frequent among patients with higher education and patients with previous digital experience.

DISCUSSION

This study presents the requirements behind, the methods used, and the results achieved during the stepwise iterative development process of EirV3, an electronic symptom assessment system for cancer care. The main objective was to improve clinical consultations by focusing on the patient's perspective, through immediate transfer of PROMs to the HCP's computer. Thus, EirV3 represents something beyond a direct electronic version of paper PROMs, as is frequently done.[47-49] The real-time visual presentation of individually tailored PROMs supplemented with graphs for symptom development cannot be achieved by the paper-and-pencil format.

Fig 2. Eir-Doctor opening screen.

Table 4. Results From the Study Examining Equivalence Between Electronic and Paper Patient-Reported Outcome Measures

Symptoms	ICC	95% CI	Mean
Well-being	0.73	0.63 to 0.81	3.12
Pain	0.89	0.84 to 0.92	2.43
Numbness	0.87	0.82 to 0.91	2.06
Shortness of breath	0.83	0.76 to 0.88	2.40
Drowsiness	0.89	0.85 to 0.93	3.20
Tiredness	0.92	0.88 to 0.94	3.92
Insomnia	0.75	0.65 to 0.82	2.39
Anxiety	0.81	0.73 to 0.87	2.68
Depression	0.80	0.72 to 0.86	2.04
Nausea	0.76	0.67 to 0.83	1.06
Vomiting	0.65	0.53 to 0.75	0.37
Diarrhea	0.88	0.83 to 0.91	1.02
Constipation	0.90	0.86 to 0.93	1.81
Lack of appetite	0.91	0.87 to 0.93	2.05
Mouth sores	0.88	0.83 to 0.92	0.49
Dry mouth	0.82	0.75 to 0.88	2.50
Altered sense of taste	0.74	0.64 to 0.81	1.92
Altered sense of smell	0.77	0.69 to 0.84	1.29
Problems swallowing	0.72	0.61 to 0.80	0.69

NOTE. Median ICC for all items was 0.81 (25th-75th quartile, 0.75 to 0.89).

Abbreviation: ICC, intraclass correlation coefficient.

The content in EirV3 covers a wide range of common cancer-related symptoms. Some argue that electronic PROM tools should be diagnosis or treatment specific to capture relevant clinical information,[14,50] whereas Eir was developed for use in adult patients with cancer, independent of cancer diagnosis, treatment, age, and stage of disease. Thus, one may question the specificity of the included symptoms. However, relevance for an individual patient is documented by well-validated tools and guidelines and enhanced by the presentation in a dynamic, electronic format. This way, patients receive tailored questions based on their current symptom status. Because the primary aim is to improve symptom management, the cutoff levels that decide the subsequent in-depth questions deserve attention. The cutoff in EirV3 is > 1 and is purposefully low not to risk overlooking symptoms. Cutoffs for high values are important, and based on common clinical practice, systematic reviews,[51] and clinical studies,[52] values $\geqslant 3$ were flagged to alert physicians.

We regard the continuous involvement of end-users—patients with cancer and HCPs—as extremely valuable, leading to close collaboration and immediate improvements. Physicians' feedback on Eir-Doctor was paramount for improvement of several functional issues. The list of symptoms in Eir-DoctorV3 (Fig 2) was perceived as beneficial for a quick overview of the current situation, even if some preferred a fixed order. Physicians frequently commented that they liked the graphical presentation of symptom trajectories and that they occasionally became aware of symptoms they did not know troubled the patient. The integration of Eir into the hospitals' records is a priority that implies security issues related to patient confidentiality and data storage.

So far, results from this thorough, systematic, and iterative development process indicate that EirV3 is user-friendly and self-explanatory for most patients. Usability issues of the first versions (eg, shortcomings regarding layout and the pain body map) led to immediate system modifications. Many of these changes, such as reducing the number of elements on each screen and centering the text, were done to reduce the likelihood of errors, thereby optimizing reliability. In our opinion, this emphasizes the importance of including end-users to improve the usability of any tool, be it digital or on paper. This was also the benefit of developing and testing Eir-Patient and Eir-Doctor in parallel, as feedback from physicians could be used for amendments of Eir-Patient, and vice versa.

Most patients regarded Eir as intuitively easy to use and appreciated its relevance and that results reached the physicians immediately. However, this was true on the group level. It may be that the perceptions varied among subgroups of patients (eg, fit *v* frail patients, patients with few symptoms *v* those with many). As a result of a generally higher symptom burden, completion was more demanding for patients in the palliative outpatient unit than in the oncology unit, potentially supporting the issue about subgroup differences, corresponding with results from other studies using computerized assessment.[21,53-56] The most negative comments were that Eir is not yet automatically incorporated into the electronic medical records and that it should be opened in an Internet browser, not connected with the regular hospital network.

Patients judged to be cognitively impaired were not included in the studies. However, it could be that some patients with mild cognitive impairment may find it easier to use an electronic tool, but this

needs to be thoroughly examined using a cognitive screening tool and a simpler electronic tool, which was beyond the scope of this work.

The comparative study examining equivalence between electronic and paper-based PROMs showed excellent agreement between the two methods. However, it should be noticed that the mean symptom intensity scores were low (Table 4). This may indicate that more patients were fit than frail and calls for purposive, maybe even stratified, sampling in forthcoming Eir studies to examine use in frailer patients. The issue regarding subgroups relates to generalizability and validity and cannot be examined by formative testing. However, this is not related to electronic PROMs tools per se, but applies to most formative process developments. In Norway, 97% of all households (with at least one person age < 75 years) had access to the Internet in 2015.[57] Lack of access to the Internet is probably not a limiting factor. Electronic health records are implemented in most Norwegian hospitals. Considering this, preference for electronic assessment was not overwhelming. This was a short questionnaire, however, so the response method might be of less importance in this context.

Eir is still in development, which implies an evaluation of the pros and cons of the development methods. The obvious next steps on our agenda consist of summative methods to systematically assess and quantify validation and usability issues.[58] Topics to investigate are the feasibility of using EirV3 in different settings, including home care, the frequency of use, and how it is being used by patients and HCPs in inpatient and outpatient units. Moreover, we need to assess the perceived usefulness of electronic PROMs in improving patient outcomes such as better symptom management, satisfaction with care and communication with HCPs, time of completion for distinct and vulnerable patients, and the degree of errors and system flaws (eg, down time). Automatic alerts when a patient has completed Eir will be developed. However, in the presented studies, study nurses were responsible for notifying the clinicians. The summative phase of Eir development has started with four small studies and one international validation study.

Some studies have documented an improvement in symptoms with systematic collections of PROMs, either electronically or on paper.[8,13,59,60] A recent randomized controlled trial concluded that this was attributed to the systematic monitoring that led to immediate symptom management in patients with a high symptom burden.[13] However, it is interesting that better satisfaction with patient-HCP communication is still the most prominent effect of systematic PROM registrations,[61] 15 years after the first publication by Velikova et al.[62]

Two take-home messages apply. First, even if newer studies do not show statistically significant effects of PROMs, the work toward patient-centered, electronic tools should continue to promote clinical uptake. Second, no tools are intended to replace the face-to-face interaction between patients and HCPs; instead, they should be regarded an asset for putting the patient's perspective in the center of the communication.

In conclusion, overall, technologic advances have led to an abundance of electronic PROM tools. In contrast to many others, EirV3 is not a direct electronic version of a paper-based questionnaire, but a dynamic tool adapted to the individual patient. EirV3 resembles a clinical consultation, and patients and HCPs endorsed the immediate transfer of PROMS to the physician's computer. Integration with electronic medical records is likely to improve symptom management and patient care by combining individual patient data from many sources simultaneously.

DOI: https://doi.org/10.1200/CCI.17.00051
Published online on ascopubs.org/journal/cci on September 5, 2017.

AUTHOR CONTRIBUTIONS

Conception and design: Hilde Krogstad, Kari Sand, Eivind Andersen, Tarje Halvorsen, Stein Kaasa, Jon Håvard Loge, Erik Torbjørn Løhre, Sunil X. Raj, Marianne Jensen Hjermstad
Administrative support: Kari Sand, Stein Kaasa, Marianne Jensen Hjermstad
Provision of study material or patients: Herish Garresori, Ellinor C. Haukland, Stein Kaasa, Marianne Jensen Hjermstad
Collection and assembly of data: Hilde Krogstad, Cinzia Brunelli, Kari Sand, Herish Garresori, Ellinor C. Haukland, Frode Jordal, Marianne Jensen Hjermstad
Data analysis and interpretation: Hilde Krogstad, Cinzia Brunelli, Kari Sand, Stein Kaasa, Erik Torbjørn Løhre, Marianne Jensen Hjermstad
Manuscript writing: All authors
Final approval of manuscript: All authors
Agree to be accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Hilde Krogstad
No relationship to disclose

Cinzia Brunelli
No relationship to disclose

Kari Sand
No relationship to disclose

Eivind Andersen
No relationship to disclose

Herish Garresori
No relationship to disclose

Tarje Halvorsen
Honoraria: AstraZeneca

Ellinor C. Haukland
No relationship to disclose

Frode Jordal
No relationship to disclose

Stein Kaasa
Stock and Other Ownership Interests: Eir Solutions AS

Jon Håvard Loge
Stock and Other Ownership Interests: Eir Solutions AS

Erik Torbjørn Løhre
No relationship to disclose

Sunil X. Raj
No relationship to disclose

Marianne Jensen Hjermstad
No relationship to disclose

ACKNOWLEDGMENT

We thank Eivind Sorteberg, BEKK Consulting AS, for providing information about the technical solutions in the development of Eir. We also express our gratitude to the following physicians and nurses who participated in the study examining equivalence between electronic and paper patient-reported outcome measure assessments: Åse Merete Thuen, Stavanger University Hospital; Kirsten Vera Hildegard Engljæhringer, MD, and Trude Merete Kristiansen, Nordland Hospital Trust; Ole Kristian Andersen, MD, Arve Nordbø, and Agnetha Lund, Vestfold Hospital Trust; Ingunn Johansen, Marianne Bøe, and Malin Hammerstad, Østfold Hospital Trust; Kristian Wennevold, MD, Ålesund Hospital; and Torbjørn Øvreness and Cinzia Marini, St Olavs Hospital, Trondheim University Hospital.

Affiliations

Hilde Krogstad, **Cinzia Brunelli**, **Kari Sand**, **Tarje Halvorsen**, **Stein Kaasa**, **Jon Håvard Loge**, **Erik Torbjørn Løhre**, **Sunil X. Raj**, and **Marianne Jensen Hjermstad**, European Palliative Care Research Centre, Norwegian University of Science and Technology (NTNU) and St Olavs Hospital, Trondheim University Hospital; **Hilde Krogstad**, **Tarje Halvorsen**, **Erik Torbjørn Løhre**, and **Sunil X. Raj**, Cancer Clinic, St Olavs Hospital, Trondheim University Hospital; **Eivind Andersen**, NTNU Technology Transfer AS, Trondheim; **Stein Kaasa**, **Jon Håvard Loge**, and **Marianne Jensen Hjermstad**, Oslo University Hospital, Oslo; **Herish Garresori**, Stavanger University Hospital, Stavanger; **Ellinor C. Haukland**, Nordland Hospital Trust, Bodø; **Frode Jordal**, Østfold Hospital Trust, Grålum, Norway; and **Cinzia Brunelli**, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico, Istituto Nazionale dei Tumori, Milano, Italy.

REFERENCES

1. Grol R, Grimshaw J: From best evidence to best practice: Effective implementation of change in patients' care. Lancet 362:1225-1230, 2003
2. Basch E, Abernethy AP: Supporting clinical practice decisions with real-time patient-reported outcomes. J Clin Oncol 29:954-956, 2011
3. Reeve BB, Mitchell SA, Dueck AC, et al: Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. J Natl Cancer Inst 106:dju129, 2014
4. NIH State-of-the-Science Statement on symptom management in cancer: pain, depression, and fatigue. NIH Consens State Sci Statements 19:1-29, 2002
5. Calvert M, Blazeby J, Altman DG, et al: Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. JAMA 309:814-822, 2013
6. Boyes A, Newell S, Girgis A, et al: Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl) 15:163-171, 2006
7. Detmar SB, Muller MJ, Schornagel JH, et al: Health-related quality-of-life assessments and patient-physician communication: A randomized controlled trial. JAMA 288:3027-3034, 2002
8. Velikova G, Booth L, Smith AB, et al: Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomized controlled trial. J Clin Oncol 22:714-724, 2004
9. Basch E, Deal AM, Kris MG, et al: Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. J Clin Oncol 34:557-565, 2016

10. Kaasa S, Loge JH, Fayers P, et al: Symptom assessment in palliative care: A need for international collaboration. J Clin Oncol 26:3867-3873, 2008
11. Greenhalgh J: The applications of PROs in clinical practice: What are they, do they work, and why? Qual Life Res 18: 115-123, 2009
12. Kotronoulas G, Kearney N, Maguire R, et al: What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 32:1480-1501, 2014
13. Strasser F, Blum D, von Moos R, et al: The effect of real-time electronic monitoring of patient-reported symptoms and clinical syndromes in outpatient workflow of medical oncologists: E-MOSAIC, a multicenter cluster-randomized phase III study (SAKK 95/06). Ann Oncol 27:324-332, 2016
14. Velikova G, Awad N, Coles-Gale R, et al: The clinical value of quality of life assessment in oncology practice: A qualitative study of patient and physician views. Psychooncology 17:690-698, 2008
15. Snyder CF, Jensen RE, Geller G, et al: Relevant content for a patient-reported outcomes questionnaire for use in oncology clinical practice: Putting doctors and patients on the same page. Qual Life Res 19:1045-1055, 2010
16. Santana MJ, Feeny D, Weinkauf J, et al: The use of patient-reported outcomes becomes standard practice in the routine clinical care of lung-heart transplant patients. Patient Relat Outcome Meas 1:93-105, 2010
17. Rose M, Bezjak A: Logistics of collecting patient-reported outcomes (PROs) in clinical practice: An overview and practical examples. Qual Life Res 18:125-136, 2009
18. European Association for Palliative Care Research Network: Do you want to take part in palliative care research? http://www.eapcnet.eu/Themes/Research.aspx
19. Norwegian University of Science and Technology: European Palliative Care Research Centre. www.ntnu.edu/prc
20. Hølen JC, Hjermstad MJ, Loge JH, et al: Pain assessment tools: Is the content appropriate for use in palliative care? J Pain Symptom Manage 32:567-580, 2006
21. Fyllingen EH, Oldervoll LM, Loge JH, et al: Computer-based assessment of symptoms and mobility in palliative care: Feasibility and challenges. J Pain Symptom Manage 38:827-836, 2009
22. Hjermstad MJ, Lie HC, Caraceni A, et al: Computer-based symptom assessment is feasible in patients with advanced cancer: Results from an international multicenter study, the EPCRC-CSA. J Pain Symptom Manage 44:639-654, 2012
23. Di Maio M, Perrone F: Quality of Life in elderly patients with cancer. Health Qual Life Outcomes 1:44, 2003
24. Raj SX, Klepstad P, Brunelli C, et al: COMBAT study: Computer Based Assessment and Treatment—A computerized clinical decision support system for evaluation and treatment of pain and other cancer related symptoms. Palliat Med 28:549-550, 2014 (abstr)
25. Jaatun EA, Hjermstad MJ, Gundersen OE, et al: Development and testing of a computerized pain body map in patients with advanced cancer. J Pain Symptom Manage 47:45-56, 2014
26. Andreassen Jaatun EA, Haugen D, Dahl Y, et al: Designing a reliable pain drawing tool: Avoiding interaction flaws by better tailoring to patients' impairments. Pers Ubiquitous Comput 19:635-648, 2015
27. Knudsen AK, Brunelli C, Klepstad P, et al: Which domains should be included in a cancer pain classification system? Analyses of longitudinal data. Pain 153:696-703, 2012
28. Knudsen AK, Aass N, Fainsinger R, et al: Classification of pain in cancer patients: A systematic literature review. Palliat Med 23:295-308, 2009
29. Caraceni A, Hanks G, Kaasa S, et al: Use of opioid analgesics in the treatment of cancer pain: Evidence-based recommendations from the EAPC. Lancet Oncol 13:e58-e68, 2012
30. Rayner L, Price A, Hotopf M, et al: The development of evidence-based European guidelines on the management of depression in palliative cancer care. Eur J Cancer 47:702-712, 2011
31. Sigurdardottir KR, Kaasa S, Rosland JH, et al: The European Association for Palliative Care basic dataset to describe a palliative care cancer population: Results from an international Delphi process. Palliat Med 28:463-473, 2014
32. Fearon K, Strasser F, Anker SD, et al: Definition and classification of cancer cachexia: An international consensus. Lancet Oncol 12:489-495, 2011
33. Kaasa S, Apolone G, Klepstad P, et al: Expert conference on cancer pain assessment and classification: The need for international consensus—Working proposals on international standards. BMJ Support Palliat Care 1:281-287, 2011
34. Brunelli C, Bennett MI, Kaasa S, et al: Classification of neuropathic pain in cancer patients: A Delphi expert survey report and EAPC/IASP proposal of an algorithm for diagnostic criteria. Pain 155:2707-2713, 2014
35. Zeppetella G, Davies AN: Opioids for the management of breakthrough pain in cancer patients. Cochrane Database Syst Rev 10:CD004311, 2013
36. Portenoy RK, Hagen NA: Breakthrough pain: Definition and management. Oncology (Williston Park) 3:25-29, 1989 (suppl 8)

37. Hagen NA, Stiles C, Nekolaichuk C, et al: The Alberta Breakthrough Pain Assessment Tool for cancer patients: A validation study using a Delphi process and patient think-aloud interviews. J Pain Symptom Manage 35:136-152, 2008
38. Kroenke K, Spitzer RL, Williams JB: The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 16: 606-613, 2001
39. Kroenke K, Spitzer RL, Williams JB, et al: Anxiety disorders in primary care: Prevalence, impairment, comorbidity, and detection. Ann Intern Med 146:317-325, 2007
40. Morin C: Insomnia: Psychological Assessment and Management. New York, NY, Guilford Press, 1993
41. National Cancer Institute: Common Terminology Criteria for Adverse Events (CTCAE) Version 4.0. https://ctep.cancer.gov/protocoldevelopment/electronic_applications/ctc.htm
42. Bauer J, Capra S, Ferguson M: Use of the scored Patient-Generated Subjective Global Assessment (PG-SGA) as a nutrition assessment tool in patients with cancer. Eur J Clin Nutr 56:779-785, 2002
43. Stinson J, McGrath P, Hodnett E, et al: Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 12:e30, 2010
44. McGraw K, Wong S: Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30-46, 1996
45. Cicchetti DV: Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 6:284-290, 1994
46. Løhre ET, Klepstad P, Bennett MI, et al: From "breakthrough" to "episodic" cancer pain? A European Association for Palliative Care Research Network expert Delphi survey toward a common terminology and classification of transient cancer pain exacerbations. J Pain Symptom Manage 51:1013-1019, 2016
47. Holzner B, Giesinger JM, Pinggera J, et al: The Computer-based Health Evaluation Software (CHES): A software for electronic patient-reported outcome monitoring. BMC Med Inform Decis Mak 12:126, 2012
48. Basch E, Iasonos A, Barz A, et al: Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. J Clin Oncol 25:5374-5380, 2007
49. Jensen RE, Snyder CF, Abernethy AP, et al: Review of electronic patient-reported outcomes systems used in cancer clinical care. J Oncol Pract 10:e215-e222, 2014
50. Snyder CF, Wu AW, Miller RS, et al: The role of informatics in promoting patient-centered care. Cancer J 17:211-218, 2011
51. Oldenmenger WH, de Raaf PJ, de Klerk C, et al: Cut points on 0-10 numeric rating scales for symptoms included in the Edmonton Symptom Assessment Scale in cancer patients: A systematic review. J Pain Symptom Manage 45:1083-1093, 2013
52. Seow H, Sussman J, Martelli-Reid L, et al: Do high symptom scores trigger clinical actions? An audit after implementing electronic symptom screening. J Oncol Pract 8:e142-e148, 2012
53. Erharter A, Giesinger J, Kemmler G, et al: Implementation of computer-based quality-of-life monitoring in brain tumor outpatients in routine clinical practice. J Pain Symptom Manage 39:219-229, 2010
54. Abernethy AP, Herndon JE II, Wheeler JL, et al: Improving health care efficiency and quality using tablet personal computers to collect research-quality, patient-reported data. Health Serv Res 43:1975-1991, 2008
55. Velikova G, Wright EP, Smith AB, et al: Automated collection of quality-of-life data: A comparison of paper and computer touch-screen questionnaires. J Clin Oncol 17:998-1007, 1999
56. Kearney N, Kidd L, Miller M, et al: Utilising handheld computers to monitor and support patients receiving chemotherapy: Results of a UK-based feasibility study. Support Care Cancer 14:742-752, 2006
57. Statistics Norway: ICT usage in households, 2015, 2nd quarter. https://www.ssb.no/en/teknologi-og-innovasjon/statistikker/ikthus/aar/2015-10-01
58. Theofanos MF, Quesenbery W: Towards the design of effective formative test reports. J Usability Stud 1:27-45, 2005
59. Johansen MA, Berntsen GK, Schuster T, et al: Electronic symptom reporting between patient and provider for improved health care service quality: A systematic review of randomized controlled trials. Part 2: Methodological quality and effects. J Med Internet Res 14:e126, 2012
60. Yount SE, Rothrock N, Bass M, et al: A randomized trial of weekly symptom telemonitoring in advanced lung cancer. J Pain Symptom Manage 47:973-989, 2014
61. Berry DL, Blumenstein BA, Halpenny B, et al: Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. J Clin Oncol 29:1029-1035, 2011
62. Velikova G, Brown JM, Smith AB, et al: Computer-based quality of life questionnaires may contribute to doctor-patient interactions in oncology. Br J Cancer 86:51-59, 2002

63. Helbostad JL, Holen JC, Jordhoy MS, et al: A first step in the development of an international self-report instrument for physical functioning in palliative cancer care: A systematic literature review and an expert opinion evaluation study. J Pain Symptom Manage 37:196–205, 2009

64. Rayner L, Price A, Hotopf M, et al: The development of evidence-based European guidelines on the management of depression in palliative cancer care. Eur J Cancer 47:702–712, 2011

65. Hjermstad MJ, Fayers PM, Haugen DF, et al: Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: A systematic literature review. J Pain Symptom Manage 41:1073–1093, 2011

66. Haugen DF, Hjermstad MJ, Hagen N, et al: Assessment and classification of cancer breakthrough pain: A systematic literature review. Pain 149:476–482, 2010

67. Martin-Moreno J, Albreht T, Krnl SR (eds): Boosting Innovation and Cooperation in European Cancer Control. Key Findings From the European Partnership for Action Against Cancer. Ljubjana, Slovenia, National Institute of Public Health of the Republic of Slovenia, 2013

APPENDIX

Former Development of Electronic Symptom Assessment Tools From Our Group

The Patient Assessment Tool-Computerized study (2007). In this descriptive study, patients with advanced cancer responded to 59 questions and a pain body map on touchscreen computers.[21] The selection of items was based on systematic reviews[20,63] and surveys among patients with cancer and expert groups. Most patients (93%) were able to report symptoms directly on the computer

The European Palliative Care Research Collaborative Computerized Symptom Assessment study (2008 to 2009). This international, multicenter study used a more sophisticated tablet version in 1,017 patients with advanced cancer. Patients were recruited from 17 centers and eight countries (Norway, the United Kingdom, Austria, Germany, Switzerland, Italy, Canada, and Australia). They responded to questions on symptoms, nutritional intake, and physical and emotional functions. The software was programmed in four languages (English, German, Italian, and Norwegian) and contained several skip sessions to reduce patient burden; if the patient had no pain, the rest of the pain section was omitted.[22] In agreement with results from other computerized assessment studies,[21,53-56] the completion rate was high (95%), with more missing information and need for assistance associated with higher age and lower performance status, similar to results when using paper-and-pencil assessments.[23]

Continuous software improvements and small-scale tests were performed based on feedback from patients and health care providers in the European Palliative Care Research Collaborative Computerized Symptom Assessment study. In 2012, a tablet version that included treatment recommendations for pain and depression was used in a Norwegian clinical trial of 143 outpatients with cancer.[24] Two studies comparing different versions of a computerized pain body map in randomized order and testing different ways of marking pain were also conducted and demonstrated the need to optimize the user-friendliness by simplifying the design for the frailest patients.[25,26]

Expert Meetings and Workshops to Decide the Content and Development of Eir

Between 2013 and 2015, regular meetings were held by the international expert panel and the core working group, in addition to two workshops.

International expert panel. The international expert group participated in workshops and roundtable discussions addressing symptom assessment, classification, and management in 2013 and 2014. Relevant symptom dimensions and validated symptom assessment tools for the choice of specific items were identified, aggregated, and presented to researchers and clinicians from different specialties to reach consensus regarding relevance and importance.[10,31,33,64-66]

The first international Eir expert group meeting in 2013 was organized as part of the European Partnership for Action Against Cancer.[67] Here, 26 participants discussed computerized symptom assessment and development of Eir. The meeting was organized with short introductions about the objectives of Eir and symptom assessment followed by plenary discussions and two workshops in which the participants worked in groups, addressing symptom assessment and treatment guidelines. The following decisions were made at the meeting:

- Eir's content should be based on evidence-based or consensus-based assessment methods.
- Eir should have a hierarchical structure, with an introductory question prior to a screening section on symptoms, followed by a section on symptom intensity and yet another for characterization for endorsed symptoms (Fig 1).
- Patients' registrations in Eir should be immediately transferred and visually presented.
- Eir should be user-friendly and relevant for heterogeneous cancer populations.
- Eir should be easy to adapt to cultural and clinical preferences.

The second international expert meeting (2013; n = 9) was arranged after the participants had tested the first tablet version of Eir-Patient (EirV1). Here, feedback regarding content and layout was collected and summarized for each of the screen images. Subsequent discussions resulted in consensus on which symptoms to include in Eir and the structure and presentation of the included items. The third international meeting (2014) consisted with experts in neuropathic and breakthrough pain (n = 5) and focused on achieving consensus on how to screen these pain types in Eir.

Norwegian core working group. The core working group (n = 15) consisted of oncologists, palliative care physicians, researchers, interaction designers, graphic designers, and software developers. Members of the core working group were the first to test each new feature of Eir, as part of the iterative development. Regular multiprofessional meetings were organized with group members, the software development team, and designers to discuss functionality and features and decide refinements.

Workshops. Prior to the development EirV1 in 2013 (Table 1), two national workshops were conducted to assess the needs and preferences of end-users. The first workshop (2013) presented the overall idea and intention of Eir to physicians, nurses, designers, and patients as participants (n = 20). Furthermore, the intended features of Eir-Patient, such as content, layout, and functionalities, were presented, and feedback suggestions from the participants were collected. The participants were positive about using an electronic tool and could foresee several advantages related to easy collection and more focus on symptom assessment, including immediate access to patients' patient-reported outcome measure scores and perhaps also improved communication.

The second workshop in 2013 was conducted with five physicians who suggested different ways of presenting patient-reported outcome measures in Eir-Doctor, either with as much information as possible on the opening screen in Eir-Doctor or to highlight only the most relevant information (eg, symptoms with the highest intensity, those with the most pronounced increase, or a combination of these).

Formative Usability Testing Methods

Formative usability tests[43] on separate sections (eg, general pain and breakthrough pain), as well as on more complete versions of Eir, were repeatedly performed by patients at the Cancer Clinic, St Olavs Hospital, Trondheim University Hospital, during the entire development process. Reports from these tests were presented to the core working group in weekly meetings, together with ideas for changes as drawn sketches or on a monitor. Consensus was reached regarding how to eliminate identified usability problems and to meet user preferences.

In November 2013, EirV1 was tested by outpatients with cancer for the first time (Table 3). On the basis of results from tests in end-users and feedback and discussion in all groups, major improvements from EirV1 to EirV2 were made in the first two quarters of 2014. The most important changes aimed to improve the user interface. For example, the display was radically changed to make the distinction between response options clearer, the buttons were slightly moved, and the layout on all questions in the symptom screening and follow-up sections were standardized. EirV2 included the Eir-Patient and Eir-Doctor modules, which were tested in 42 inpatients and outpatients and eight physicians at the Cancer Clinic in 2014 (Table 3). Patients completed Eir-Patient on tablets in the waiting room, and physicians used Eir-Doctor in the consultations. Usability data were collected through interviews and observations.

Paper III

This article is not included due to copyright restritions.

ISBN 978-82-326-5050-7 (printed ver.)
ISBN 978-82-326-5051-4 (electronic ver.)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology