Quality of life assessment and screening tool development for endometriosis

Nina Julie Verket

Institute of Clinical Medicine

Faculty of Medicine

University of Oslo

and

Research Center for Obstetrics and Gynecology

Department of Obstetrics and Gynecology

Oslo University Hospital Ullevål

UiO : Institute of Clinical Medicine
University of Oslo

Oslo University Hospital

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-718-5

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Table of contents

Acknowledgements

I wish to thank two individuals who were influential for my research and choice of field: Mr. Niels Christian Danbolt (professor of neurobiology), my supervisor from the medical education research specialization program (Forskerlinjen), for teaching me the fundamentals of a research laboratory and control groups, whose research group I left due to intense dislike of rats (the brains of which we used in research). Mr. Kristen Olav Lind (senior consultant of obstetrics and gynecology at Nordland Hospital Vesterålen), for once upon a time inspiring me to want to become a gynecologist, although I chose not to become one.

I wish to thank my supervisors:
Mr. Erik Qvigstad (senior consultant of gynecology), my now retired previous main supervisor, for giving me full academic freedom and responsibility, the former which I longed for, the latter which I occasionally cursed him for. Mr. Tom Gunnar Tanbo (professor and senior consultant of reproductive medicine), my previous co-supervisor who kindly took on the role of main supervisor in 2018, for academic supervision, timely support, and fair judgment. Mr. Leiv Sandvik (professor and senior statistician), my de facto third supervisor, for weekly academic brainstorming and for continued friendship.

I wish to thank my co-authors Ms. Marit Helen Andersen (professor of interdisciplinary health sciences), Mr. Till Uhlig (professor and senior consultant of rheumatology), and Ms. Ragnhild Sørum Falk (associate professor and senior statistician). I hope I will have the pleasure of collaborating with them in the future.

I wish to thank Mr. Anton Langebrekke (senior consultant of gynecology and leading clinical expert on endometriosis), for the grace and kindness of letting me in on clinical experience and thoughts on the field of endometriosis. I wish to thank Ms. Karen Bertelsen and the Norwegian Endometriosis Association (Endometrioseforeningen – Norge) for their support and contribution to the research which forms the basis of the present thesis.

I wish to thank Mr. Kjartan Moe (research fellow and junior consultant of obstetrics) for comradeship and letting-out-steam lunches.

For inspiration, I wish to thank the Amazons of the Department of Obstetrics and Gynecology, Oslo University Hospital Ullevål, in alphabetical order, Ms. Anne Flem Jacobsen (professor and senior consultant of obstetrics), Ms. Katariina Laine (associate professor and senior consultant of obstetrics), Ms. Marit Lieng (professor and senior consultant of gynecology), and Ms. Annetine Staff (professor and senior consultant of gynecology).

Lastly, I would like to thank my father, who has been, and still is, my safe haven.

To my daughter Sophie (6 years), my son Sebastian (9 years), and my husband Anders. I hope you are proud.

Abbreviations

AUC	Area Under the receiver operating characteristic Curve
BMI	Body Mass Index
BP	Bodily Pain
COC	Combined Oral Contraceptive
EHP-30	Endometriosis Health Profile-30
EndoCost	Endometriosis Cost assessment
ERI-1	Endometriosis Risk Index variant 1
ERI-2	Endometriosis Risk Index variant 2
FN	False Negative
FP	False Positive
GH	General Health
GnRH	Gonadotropin-Releasing Hormone
MCS	Mental Component Summary
MH	Mental Health
MRI	Magnetic Resonance Imaging
NPV	Negative Predictive Value
NSAID	NonSteroidal Anti-Inflammatory Drug
PCS	Physical Component Summary
PF	Physical Functioning
PPV	Positive Predictive Value
RA	Rheumatoid Arthritis
RAQoL	Rheumatoid Arthritis Quality of Life scale
RE	Role-Emotional
RP	Role-Physical
SF	Social Functioning
SF-36	Short Form-36
SF-36v1	Short Form-36 version 1
SF-36v2	Short Form-36 version 2
TN	True Negative
TP	True Positive

TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
VT	ViTality
WHO	World Health Organization

List of papers

1. **Verket NJ, Andersen MH, Sandvik L, Tanbo TG, Qvigstad E.** Lack of cross-cultural validity of the Endometriosis Health Profile-30. J Endometr Pelvic Pain Disord. 2018;10(2):107-115.

2. **Verket NJ, Uhlig T, Sandvik L, Andersen MH, Tanbo TG, Qvigstad E.** Health-related quality of life in women with endometriosis, compared with the general population and women with rheumatoid arthritis. Acta Obstet Gynecol Scand. 2018;97(11):1339-1348.

3. **Verket NJ, Falk RS, Qvigstad E, Tanbo TG, Sandvik L.** Development of a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis: cross-sectional study. BMJ Open. 2019;9(12):e030346.

Thesis summary

Endometriosis is a relatively common chronic inflammatory gynecological disease affecting women of reproductive age. It can cause significant pain and infertility. There is no cure, and symptomatic treatment can vary from occasional use of over-the-counter pain-killers to multiple extensive surgeries. Thus, the potential consequences of endometriosis can be substantial and last several decades, significantly impacting quality of life. The thesis is based on three papers (listed on page 8).

1. The aim of the first paper was to evaluate the measurement properties of the Norwegian version of the disease-specific quality of life questionnaire Endometriosis Health Profile-30, and thereby its suitability for future use in endometriosis research in Norway.

   Novel finding: Of the five scales of the Endometriosis Health Profile-30 (*pain*, *control & powerlessness*, *emotional well-being*, *social support*, and *self-image*), the construct self-image does not seem to be measured appropriately by the Norwegian version, suggesting a lack of cross-cultural validity of the Endometriosis Health Profile-30.

2. The aim of the second paper was to compare quality of life in women with endometriosis, women from the general population, and women with rheumatoid arthritis using the generic quality of life questionnaire Short form-36.

   Novel finding: Women with endometriosis seemed to have poorer mental quality of life compared with women with rheumatoid arthritis, despite similar pain scores. Women with endometriosis had significantly reduced mean scores for the three SF-36 mental domain subscales *vitality*, *social functioning*, and *mental health*.

As endometriosis is estrogen dependent, hormonal contraceptive treatment, by inducing a hyper-progestogenic state, is the mainstay of long-term symptom suppression. However, cases of women with extensive surgical findings despite past history of oral contraceptive treatment, have questioned the appropriateness of liberal use of hormonal treatments without follow-up. Therefore, development of strategies for early diagnosis of women that would benefit from closer follow-up and care is a research priority. However, the lack of specific symptoms and diagnostic tools, combined with definite diagnosis requiring surgery, is making this task challenging, causing substantial diagnostic delay for many patients. It follows from the lack of specific symptoms and diagnostic tools that the longest delay takes place in primary care.

3. The aim of the third paper was to identify predictors of endometriosis among factors commonly associated with endometriosis and available to physicians through medical interview. Further, if successful, to combine these to develop and internally validate a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis.

> Novel finding: The predictors *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* demonstrated the strongest association with disease. A prediction model based on these two predictors appears to be a relatively efficient screening tool for endometriosis.

All three papers are based on cross-sectional data from postal surveys conducted among women with and without endometriosis and women with rheumatoid arthritis. Women with endometriosis were recruited from the Norwegian Endometriosis Association. The response rate was 41.9%. We had no access to medical records. Women without endometriosis were recruited from a randomly selected sample of women residing in Oslo. The response rate was 14.9%. Women with rheumatoid arthritis were recruited from the Oslo Rheumatoid Arthritis Registry. The response rate was 59.7%. The main weakness of all three papers is likely selection bias. Thus, results from the second and third papers should be considered indicative and in need of further investigation in future studies.

Introduction

Endometriosis is a common chronic inflammatory gynecological disease mainly affecting women of reproductive age. It is characterized by endometrium-like tissue in aberrant locations, such as on the peritoneum or pelvic organs (bladder, uterus, ovaries, rectum). This endometrium-like tissue may cause inflammatory reactions, with or without accompanying adhesions, fibrosis, scarring, and neuronal infiltration (1). The main symptoms of endometriosis are pain, most frequently dysmenorrhea (painful periods) and dyspareunia (painful intercourse), and infertility (1-3). Both disease expression and disease progression vary markedly (4). Disease onset can be as early as adolescence, with disease persistence throughout child bearing age until a presumed burn out at menopause. There is no cure, and symptomatic treatment can vary from occasional use of over-the-counter pain-killers to multiple extensive surgeries with adhesiolysis and organ resection or removal. Thus, the potential consequences of early onset progressive endometriosis can be substantial and last multiple decades.

Quality of life

It is not surprising that endometriosis, associated with both pain and infertility, can affect quality of life (5). Although many have an intuitive understanding of what quality of life comprises, definitions of quality of life are controversial (6, 7).

The World Health Organization (WHO) defines health as a state of complete physical, mental, and social well-being and not merely the absence of disease or infirmity (8). Health in this broad sense is interchangeably called health status or quality of life. However, quality of life means different things to different people, and takes on different meanings according to the area of application (6). This wider perception of quality of life is evident in WHOs definition of quality of life as an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards, and concerns (9). It is a broad ranging concept affected in a complex way by the

person's physical health, psychological state, personal beliefs, social relationships, and their relationship to salient features of their environment (9).

The term health-related quality of life is frequently used to indicate a narrower definition of quality of life comprising only aspects relevant to clinical trials. These can include general health, physical functioning, physical symptoms and toxicity, emotional functioning, cognitive functioning, role functioning, social well-being and functioning, sexual functioning, and existential issues (6).

In our studies, the term health-related quality of life is defined as a multidimensional concept that refers to the patient's general perception of the effect of disease and treatment on physical, psychological, and social aspects of daily life (10-12). Throughout this thesis, the term quality of life is used instead of health-related quality of life.

Measuring quality of life

It has been argued that clinical outcomes measured by investigators, in many cases ought to be supplemented by patient-reported outcomes in clinical trials. A patient-reported outcome is "any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else" (11, 13). Such patient-reported outcomes are often regarded as indicators of quality of life.

The importance of including patient perspectives is perhaps easiest to convey using examples from cancer research: In evaluation of a new cancer therapy, a decrease in quality of life may outweigh an increase in survival by 3 months, thereby favoring the "old" therapy (14). Evaluation of endometriosis therapies, associated with varying symptom relief and side-effects, should similarly include patient-reported outcomes. Quality of life instruments may also reveal other issues that are equally or more important to patients.

Quality of life instruments, or questionnaires, can be generic, applicable to patients with a variety of conditions and often also healthy subjects, or disease-specific (6). Disease-specific instruments may detect change in important aspects of certain conditions not accessible by

generic instruments (15). Few disease-specific quality of life instruments have been developed in the field of endometriosis. Of these, the Endometriosis Health Profile-30 (EHP-30) is the most thoroughly evaluated and extensively administered (16, 17). The original English version was developed in the UK and first presented in 2001 (16). The items, or questions, were generated from in-depth interviews of 25 patients with endometriosis visiting a gynecology clinic at a large tertiary referral hospital in Oxford. The EHP-30 has been translated into multiple languages, and the translated versions have demonstrated good psychometric properties (18-23). The EHP-30 has not yet been validated in any of the Scandinavian languages.

The aim of the first study was to evaluate the psychometric properties of the Norwegian version EHP-30 and thereby its suitability for future use in endometriosis research in Norway.

Quality of life comparison with rheumatoid arthritis

A study assessing availability of quality of life instruments across 30 medical specialties, showed greatest development and evaluation of quality of life instruments for rheumatology and musculoskeletal disorders and cancer, and relatively less for burns and trauma, intensive care, and gynecology (24). Given that many gynecological diseases, including endometriosis, are chronic and associated with substantial psychological distress, the authors express their surprise concerning the lack of advancement in assessment of outcomes from the patient's perspective in gynecology (24).

Few studies have compared quality of life between endometriosis and other diseases (25). This may partly explain why many women with endometriosis report not being taken seriously by the general public and health care professionals. Comparisons with common references easy to relate to, such as more acknowledged patient groups, would aid public communication of disease burden associated with endometriosis (5).

The aim of the second study was to compare quality of life in women with endometriosis, women from the general population, and women with rheumatoid arthritis.

We chose to compare endometriosis with rheumatoid arthritis for several reasons. First, rheumatology is a specialty with a long tradition of quality of life research, also in Norway. Second, rheumatoid arthritis is a disease well acknowledged in the general public and among healthcare professionals. Further, there are some striking similarities between rheumatoid arthritis and endometriosis. Both diseases are chronic inflammatory diseases affecting mostly women (26). Inflammation localizes partly on membranes lining sterile, closed compartments (synovial membranes and peritoneum), and pain is a main symptom in both diseases. The two diseases may even have some inflammatory mediation pathways in common. A gonadotropin-releasing hormone (GnRH) antagonist (cetrorelix) has recently been shown to produce rapid anti-inflammatory effects for a subset of rheumatoid arthritis patients with higher gonadotropin levels resistant to traditional treatment (27). Another GnRH antagonist (elagolix) has been shown to reduce endometriosis-associated pain (28).

Endometriosis – diagnosis and management

Many aspects associated with endometriosis can be thought to impact quality of life.

Diagnosis is difficult and substantial diagnostic delay is common. The two most common pain symptoms, dysmenorrhea and dyspareunia, are not only common among women without endometriosis, but also considered by many as physiological. Non-invasive tests with sufficient predictive properties, including blood biomarkers, are lacking (29). Ultrasonography and magnetic resonance imaging may aid detection of ovarian endometriotic cysts and deep endometriosis, but not peritoneal endometriosis, and are highly dependent on the experience of the investigator (30). To date, definite diagnosis still requires laparoscopic visualization of abnormal patches of tissue, accurate assessment of which is also highly investigator dependent (31). Thus, it is not surprising that for some patients it takes years – several studies report a mean delay of ~7 years – before diagnosis is established (32-34). It follows from the lack of diagnostic tools that the longest delay takes place in primary care (35, 36). Because infertility as a symptom is swiftly followed up by standard algorithms, the diagnostic delay associated with endometriosis-related infertility is often much shorter than for endometriosis-related pain.

To date there is no cure for endometriosis. Treatment is symptomatic, and may be either medical or surgical. As endometriosis is estrogen dependent, hormonal therapy, by inducing a hypo-estrogenic (GnRH agonists) or hyper-progestogenic state (oral contraceptives, progestins), is the mainstay of medical symptom suppression, often combined with analgesics such as NSAIDs. Hormonal contraceptives, including combined oral contraceptive (COC) pills and progestins (e.g. progestin-only pills and levonorgestrel-releasing intrauterine systems), are recommended as first-line treatment (4, 37-39). COCs are among the most commonly prescribed first-line treatments and thought to suppress cell proliferation and enhance programmed cell death in eutopic endometrial tissue, two factors possibly involved in growth and/or recurrence of endometriosis (40). GnRH agonists are given to patients who fail to respond to first-line therapy (of which several options have often been tried). These, however, are associated with menopausal side-effects. For extended treatment, it is supplemented with add-back therapy (with low levels of estrogen or progestin) to avoid bone mineral density loss. Symptom recurrence is common following therapy cessation.

Further, surgical therapy is not straight-forward. The extent of surgical therapy depends on the desire to conceive. Conservative surgery with preservation of reproductive organs is preferred in women who desire to conceive. Definite surgery with aggressive removal of visible disease and attempted restoration of anatomy, may in worst cases involve pelvic organ resection and/or removal requiring multidisciplinary therapy with urological and colorectal surgery (41). Even then, post-surgical medical therapy may be crucial to limit symptom recurrence (42).

Screening tool for endometriosis

The development of a screening tool to aid in the diagnosis of endometriosis is considered to be one of the top ten research questions for endometriosis (4, 43, 44). Cases of women with extensive surgical findings despite past history of oral contraceptive treatment for dysmenorrhea, has questioned the appropriateness of liberal use of hormonal treatments without follow-up (45). Therefore, development of strategies for early diagnosis of women who would benefit from closer follow-up and care, is a priority. However, the lack of specific

symptoms and diagnostic tools, combined with definite diagnosis requiring surgery, is making this task challenging.

In recent years, several other pain-related behaviors have gained interest as possible facilitators of early diagnosis, such as the use of oral contraceptives due to dysmenorrhea and absenteeism from school due to dysmenorrhea (45, 46). Endometriosis-related symptoms have been separately examined in women with endometriosis and women from the general population, but not compared between these groups, which is a prerequisite for symptom-based screening tool development.

The aims of the third study were to identify predictors of disease among a few factors commonly associated with endometriosis and available to physicians through medical interview, and if successful, to combine these to develop and internally validate a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis.

The performance of a screening tool is measured by sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV):

	With disease	**Without disease**	
Test positive	True positive (TP)	False positive (FP)	**PPV** = $\frac{TP}{TP+FP}$
Test negative	False negative (FN)	True negative (TN)	**NPV** = $\frac{TN}{FN+TN}$
	Sensitivity = $\frac{TP}{TP+FN}$	**Specificity** = $\frac{TN}{FP+TN}$	

Sensitivity is the proportion of individuals with the disease which the test identifies correctly as having the disease.

Specificity is the proportion of individuals without the disease which the test identifies correctly as not having the disease.

Positive predictive value is the proportion of individuals with a positive test who actually have the disease.

Negative predictive value is the proportion of individuals with a negative test who actually are without the disease.

Prevalence data are needed to estimate screening tool performance. However, obtaining reliable prevalence estimates for endometriosis is difficult. Endometriosis prevalence is commonly referred to be ~5-10% among women of reproductive age (4). Prevalence of 10% is likely overstated as these higher estimates have been based on high-risk populations, such as patients attending gynecological surgical departments or infertility clinics (47, 48). In an American cross-sectional survey of women aged 18-49 years from 2012, 6.1% reported having endometriosis, of which about half were diagnosed surgically (49). General population-based estimates from European and Israeli databases suggest more modest prevalences of 0.8-1.8% (50-54). The lower prevalence estimates are likely understated as these lower estimates have been solely based on physician-registered diagnosis of endometriosis. In our third study, we chose to use relatively modest prevalence estimates to avoid overestimation of the performance of our prediction models.

Aims of the thesis

The aims of the present thesis were to expand on and put into context the aims of the three studies that the thesis is based on:

1. To evaluate the measurement properties of the Norwegian version of the disease-specific quality of life questionnaire Endometriosis Health Profile-30, and thereby its suitability for future use in endometriosis research in Norway.

2. To compare quality of life in women with endometriosis, women from the general population, and women with rheumatoid arthritis, using the generic quality of life questionnaire Short form-36.

3. To identify predictors of endometriosis among factors commonly associated with endometriosis and available to physicians through medical interview, and if successful, to combine these to develop and internally validate a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis.

Participants, materials, and methods

Study populations

Our studies were based on data from three study populations: 1) women with endometriosis, 2) women from the general population, and 3) women with rheumatoid arthritis. An overview is presented below:

Women with endometriosis	Women from the general population	Women with rheumatoid arthritis
• Recruitment source: Norwegian Endometriosis Association • Age 18-45 years • Data collected in 2012/2013 • Response rate: 41.9% • *n* = 157	• Recruitment source: Random sample of women residing in Oslo • Age 18-45 years • Data collected in 2012/2013 • Response rate: 14.9% • *n* = 156	• Recruitment source: Oslo Rheumatoid Arhtritis Registry • Age 23-96 years • Data collected in 2009 • Response rate: 59.7% • *n* = 837

Women with endometriosis were recruited from the Norwegian Endometriosis Association. Inclusion criteria were 18-45 years of age and surgically confirmed diagnosis.

Women from the general population were recruited from a random sample of women 18-45 years of age living in Oslo, Norway. Inclusion criteria were 18-45 years of age and no known diagnosis of endometriosis.

Women with rheumatoid arthritis were included from the Oslo Rheumatoid Arthritis Registry, which is assumed to be 85% complete and representative of adult patients with rheumatoid arthritis residing in Oslo, Norway (55, 56).

Oslo Rheumatoid Arthritis Registry

A cross-sectional postal survey including Short form-36 version 1 (SF-36v1), was conducted among members of the Oslo Rheumatoid Arthritis Register in 2009 (57). The survey had a response rate of 59.7% (57). Respondents and non-respondents were similar for age, gender distribution, and disease duration (57). A total of 837 questionnaires from female respondents (age range 23-96 years) were available for analysis, 126 of which were from women below 46 years of age.

Study design and data collection

Cross-sectional data collection from women with endometriosis and women from the general population was performed from August 2012 to August 2013. A set of two anonymous postal questionnaires including the Endometriosis Health Profile-30 (EHP-30), Short form-36 version 2 (SF-36v2), and candidate predictors of endometriosis, was sent to potential participants (the first questionnaire of the set of two anonymous questionnaires sent to potential participants, "Spørreskjema 1", can be found in the appendix). Participants were asked to fill in the second questionnaire one month after completing the first for test-retest reliability analysis of the EHP-30 (for further information, please see pages 25-26).

Initially, questionnaires were sent to a random sample of 150 women with endometriosis. Of these, 60 successfully completed and returned the first questionnaire. Based on this preliminary response rate, questionnaires were sent to a second random sample of 225 women with endometriosis not contacted in the first round. In total, 162 of 375 women with endometriosis successfully completed and returned the first questionnaire. Five of these reported that their diagnosis had not been confirmed surgically and were excluded. Thus, 157 women with endometriosis were included, giving a response rate of 41.9%.

Correspondingly, the same set of questionnaires were first sent to a random sample of 300 women from the general population. Of these, 43 successfully completed and returned the first questionnaire. Based on this preliminary response rate, questionnaires were sent to a second random sample of 750 women from the general population not contacted in the first

round. In total, 159 of 1050 women from the general population successfully completed and returned the first questionnaire. Although the questionnaire included a letter asking only women without endometriosis to participate, three women reported having endometriosis and were excluded. Thus, 156 women from the general population were included, giving a response rate of 14.9%.

Among the 157 women with endometriosis, 94 completed and returned the second questionnaire at a later date. Of these, 10 reported change in treatment or starting new treatment since completing the first questionnaire. Excluding these, test-retest reliability of EHP-30 could be assessed in 84 of the respondents.

Quality of life

Endometriosis Health Profile-30

The EHP-30 was used to measure quality of life. The EHP-30 is a disease-specific quality of life questionnaire which includes a core and a modular questionnaire (17). The responses are based on patient experiences during the 4 weeks prior to answering the questionnaire. The core questionnaire is composed of 30 items grouped into 5 scales: *pain* (11 items), *control & powerlessness* (6 items), *emotional well-being* (6 items), *social support* (4 items), and *self-image* (3 items). The modular questionnaire is composed of 23 items grouped into 6 scales: *work life* (5 items), *relationship with children* (2 items), *sexual intercourse* (5 items), *medical profession* (4 items), *treatment* (3 items), and *infertility* (4 items). The modular questionnaire is characterized by the possibility of responding only to scales which the patient deems relevant to her. All EHP-30 scales can achieve a minimum score of 0, indicating low disability, and a maximum score of 100, indicating high disability. All items of a scale must be answered to be able to calculate a scale score. The only exception is the scale *sexual intercourse*, where each item may be relevant independently of the other items of the same scale. Thus, the scale score for the scale *sexual intercourse* is calculated by omitting items which are not relevant.

Norwegian version Endometriosis Health Profile-30

At the time of study initiation, a Norwegian version of the EHP-30 core questionnaire was available from Isis Innovation (later renamed Oxford University Innovation) by purchase of a licence. However, a Norwegian version of the EHP-30 modular questionnaire was not available. The Norwegian language has two distinct written varieties, “bokmål” and “nynorsk” (58). “Bokmål” is the most commonly used variety. The EHP-30 core questionnaire had therefore been translated to “bokmål” Norwegian. The translation and cultural adaptation had been conducted by Oxford Outcomes. We therefore requested Oxford Outcomes to conduct the translation and cultural adaptation of the EHP-30 modular questionnaire to “bokmål” Norwegian according to the same guidelines used for the core questionnaire (59):

1. *Forward translations*: The EHP-30 was translated into Norwegian by two independent translators who are native Norwegian speakers. The two forward translations were reconciled into a third translation by an in-country investigator. Any issues that arose from this stage were discussed with the Oxford Outcomes project manager.
2. *Back translations*: The reconciled translation was back translated into English by two independent translators who are native English speakers and fluent in the Norwegian language and who had no prior knowledge of the EHP-30. The back translations were reviewed against the original EHP-30 by the Oxford Outcomes project manager. Any issues arising from this review were passed to the in-country investigator for comment.
3. *Clinician review*: The translation was passed to a clinician specializing in the appropriate area in Norway. The clinician reviewed the translation to ensure that it was linguistically and culturally appropriate for use in Norway, and that it was acceptable for use with patients. Any suggestions or issues were passed to the in-country investigator who, in conjunction with the Oxford Outcomes project manager, worked to resolve any problems and further refine the translation.
4. *Developer review*: The translation was reviewed by the instrument developer.
5. *Cognitive debriefing*: The translation was given to five women with endometriosis in Norway who are all native Norwegian speakers. They were asked to read through and complete the Norwegian version EHP-30. Following completion, the women were asked a series of questions aimed at gauging their comprehension of the wording of the translation. The answers, along with any other relevant comments and suggestions were summarized in a report, followed by review of results by the Oxford Outcomes project

manager. Any issues arising were sent to the in-country investigator for further review or revision.

After completion of the translation and cultural adaptation by Oxford Outcomes, a licence to use the Norwegian version EHP-30 core and modular questionnaire was obtained from Isis Innovation.

Short form-36

The Short-form 36 (SF-36) was used to measure quality of life (60). The SF-36 is a generic quality of life questionnaire composed of 36 items, one item assessing health change and 35 items grouped into 8 scales: *physical functioning* (PF; 10 items), *role limitations due to physical problems* (RP; 4 items), *bodily pain* (BP; 2 items), *general health perceptions* (GH; 5 items), *vitality* (VT; 4 items), *social functioning* (SF; 2 items), *role limitations due to emotional problems* (RE; 3 items), and *mental health* (MH; 5 items) (61, 62). All scales can achieve a minimum score of 0 (worst health), and a maximum score of 100 (best health). From the eight scales, two linear combinations are commonly computed: physical component summary (PCS) and mental component summary (MCS) (63). PCS is calculated by weighting physical domain subscales PF, RP, BP, and GH positively and mental domain subscales VT, SF, RE, and MH negatively. MCS is calculated by weighting mental domain subscales positively and physical domain subscales negatively. PCS and MCS are standardized to have a mean of 50 and a standard deviation of 10 based on a general population sample. SF-36v2 contains small changes in wording and layout compared to SF-36v1 (64). For items of the scales RP and RE, the number of response categories was increased from two (yes/no) to five (all of the time/most of the time/some of the time/a little of the time/none of the time), reducing floor and ceiling effects (65). PCS and MCS of SF-36v1 used in our study are based on a Norwegian general population sample collected in 1996. PCS and MCS of SF-36v2 used in our studies are based on a U.S. general population sample collected in 2009 (65). QualityMetric Health Outcomes™ Scoring Software 4.5 from OptumInsight Life Sciences, Inc. was used to score SF-36v2.

Candidate predictors of endometriosis

The candidate predictors were chosen based on three criteria: 1) They had to be applicable to most, if not all female adolescents. By this criterion, variables such as dyspareunia (according to surveys from 99700 Norwegian high school students from 2016 to 2018, about half have had intercourse by the age of 18), ultrasound/MRI findings, surgical findings, infertility, and previous pregnancies were excluded as candidate predictors (66). 2) They had to be simple and comprehensible to young adolescents, without the need for supplementary explanation. By this criterion, variables such as pelvic pain (we were for example not confident in adolescents' ability to readily localize symptoms as from the pelvis) and the concept of cyclic vs non-cyclic symptoms, were excluded. 3) They had to be available from early stages of the disease and reasonably frequent. By this criterion, variables such as dysuria and dyschezia were excluded. The following candidate predictors (with the questions, Q, and answer alternatives, A, given in parenthesis) were included in the final questionnaire sent to women with endometriosis and women from the general population:

1. Age at menarche
   (Q: How old were you when you had your first period?)
2. Severe dysmenorrhea in adolescence
   (Q: Did you have very painful periods as a teenager?)
   (A: Never/Rarely/Sometimes/Often/Always)
3. Absenteeism from school due to dysmenorrhea
   (Q: Did you have to be absent from school – junior high school/high school – because of painful periods?)
   (A: Never/Rarely/Sometimes/Often/Always)
4. Use of painkillers due to dysmenorrhea in adolescence
   (Q: Did you use painkillers for painful periods as a teenager?)
   (A: Never/Rarely/Sometimes/Often/Always)
5. Use of oral contraceptives due to dysmenorrhea in adolescence
   (Q: Did you use oral contraceptives because of painful periods as a teenager?)
   (A: Yes/No)
6. Family history of endometriosis
   (Q: Does anyone in your family have endometriosis?)

(A: Yes/No/Irrelevant)

Sample size calculation

The sample size calculation was based on the evaluation of the measurement properties of the Norwegian version Endometriosis Health Profile-30. Correlation coefficients played a central role in this study. We used Fisher's z transformation to estimate 95% confidence interval for a correlation coefficient *r* (67). The confidence interval for a correlation coefficient *r* is widest when $r = 0.50$. We consider it sufficient with a precision of ± 0.10, i.e. that the length of the confidence interval for *r* is at most 0.20 (18). For a correlation coefficient of 0.50 with a sample of 150 patients, this confidence interval will be 0.40-0.60. We therefore decided to include 150 women with endometriosis and 150 women without endometriosis in our study.

Statistical analysis

PAPER 1: Lack of cross-cultural validity of the Endometriosis Health Profile-30

Norwegian version Endometriosis Health Profile-30 (EHP-30) data and SF-36 data from women with endometriosis were used to assess construct validity, reliability, and interpretability of the Norwegian version EHP-30.

Validation of instruments is the process of determining whether there are grounds for believing that the instrument measures what it is intended to measure, and that it is useful for its intended purpose (6). Assessment of reliability consists of determining that a scale or

measurement yields reproducible and consistent results (6). Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument (68).

Construct validity: Exploratory factor analysis was used to assess structural validity (69). Principal components analysis with varimax rotation was used to identify the different potential components with eigenvalues greater than 1 (70). Items with factor loadings ≥ 0.40 in a factor were included in the factor. The SF-36 was used for hypotheses-testing to assess convergent validity (71-73). We hypothesized the strongest correlations between EHP-30 *pain* and SF-36 *bodily pain*, and EHP-30 *emotional well-being* and SF-36 *mental health*. We further expected a strong correlation between EHP-30 *social support* and SF-36 *social functioning*, and EHP-30 *work life* and SF-36 *role-physical*. After obtaining the results of the factor analyses, we hypothesized a strong correlation between EHP-30 *control & powerlessness* and SF-36 *bodily pain*, and EHP-30 *relationship with children* and SF-36v2 *role-physical*. Associations between scales of the EHP-30 and the SF-36 were calculated by Spearman's rho correlation coefficient. There are no widely accepted criteria for defining a strong versus moderate versus weak correlation (74). In our study, values 0.20-0.39 were considered to indicate weak correlations, values 0.40-0.59 moderate, values 0.60-0.79 strong, and values 0.80-1.00 very strong correlations (74).

Reliability: Cronbach's alpha and corrected item-total correlations were used to measure internal consistency. Cronbach's alpha above 0.70 were considered to indicate acceptable internal consistency reliability for group comparisons, and values above 0.90 for individual comparisons (71). Item-total correlations were corrected for overlap by omitting the item from the parent scale total. Item-total correlations above 0.40 were considered to indicate acceptable internal consistency (75). Intraclass correlation coefficients for agreement and paired t-tests were used to measure test-retest reliability. Intraclass correlation coefficients above 0.70 were considered to indicate acceptable reliability for group comparisons, and values above 0.90 for individual measurements over time (71, 76). Significant differences in mean scores ($p < 0.05$) were considered to indicate poor reliability. No significant differences in mean scores were considered to indicate acceptable reliability.

Interpretability: Data completeness, mean scores and standard deviations, floor and ceiling effects, and skewness of score distribution were used to describe the distribution of item

responses (72). Floor or ceiling effects were considered present if more than 15% of respondents scored the minimum value of 0 or the maximum value of 100, respectively (76).

PAPER 2: Health-related quality of life in women with endometriosis, compared with the general population and women with rheumatoid arthritis

Short form-36 (SF-36) data from women with endometriosis, women from the general population, and women with rheumatoid arthritis were included in the analyses. Mean SF-36 scale scores for the groups were compared using the independent samples t-test. The assumption of distribution normality was checked and found to be adequately met. Linear regression analysis was used to adjust for age, BMI, diagnostic delay and/or disease duration (available for the endometriosis group and the rheumatoid arthritis group), infertility (available for the endometriosis group and the control group), and pain (represented by the score for the SF-36 scale *bodily pain*).

PAPER 3: Development of a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis: cross-sectional study

Development of risk indices: Women with endometriosis (case) and women from the general population (control) with complete data for the candidate predictors presented on page 24-25 (154 cases and 145 controls) were included in the analyses.

Two different approaches were used to develop two risk indices: Endometriosis Risk Index variant 1 (ERI-1) based on logistic regression analysis, and Endometriosis Risk Index variant 2 (ERI-2) based on lasso regression analysis. Logistic regression analysis is one of the most frequently used methods to develop prediction models by selecting relevant predictors and combining them statistically into a multivariable model (77). However, logistic regression may overestimate performance. We therefore applied lasso regression analysis, a penalization procedure that performs both variable selection and regularization, during model development, as recommended in the TRIPOD checklist for developing and validating prediction models (77).

In the regression analyses, *age at menarche* was included as a continuous variable. To increase test power, the ordered categorical variables *severe dysmenorrhea in adolescence* and *absenteeism from school due to dysmenorrhea* were included as continuous variables based on linearity of the beta coefficients, supporting the assumption of the categories (never/rarely/sometimes/often/always) being equally spaced. The ordered categorical variable *use of painkillers due to dysmenorrhea in adolescence* was recoded into three categories (never/rarely, sometimes, and often/always) based on deviations from linearity of the beta coefficients. *Use of oral contraceptives due to dysmenorrhea in adolescence* was included as a dichotomous (yes/no) variable. The categorical variable *family history of endometriosis* was recoded into two categories (yes and no/irrelevant/missing) to be able to handle the real-life response category "Irrelevant" (for example if adopted). Missing responses were also included in this dichotomous categorization due to the likelihood of blank responses being comparable to participants simply not knowing. Further, a sensitivity analysis was performed, i.e. a re-analysis with an alternative dichotomous categorization (yes/no) for the candidate predictor *family history of endometriosis*, excluding the responses "Irrelevant" and missing (142 cases and 130 controls).

First, univariable and multivariable logistic regression analysis were performed to assess the relationship between the six candidate predictors and endometriosis. Backward stepwise variable selection was performed using $p \leq 0.157$ as criterion (corresponding to Akaike

information criteria). The results were presented as beta coefficients and odds ratios with 95% confidence intervals based on 1000 bootstrap samples. ERI-1 was based on the relative ratio between the beta coefficients. Second, lasso regression analysis was performed with 10-fold cross-validation and 1000 bootstrap samples, as implemented in the R package *mami*. The results were presented as means of the lasso regression coefficients with 95% confidence intervals. ERI-2 was based on the relative ratios between the lasso regression coefficients.

Internal validation: Women with endometriosis (case) and women from the general population (control) with complete data for the predictors included in ERI-1 and ERI-2 (155 cases and 148 controls) were included in the analyses.

The predictive abilities of the two risk indices, ERI-1 and ERI-2, were described by area under the receiver operating characteristic curve (AUC). Sensitivity and specificity for different cut-off values of the risk indices were calculated, as well as positive and negative predictive values (PPV and NPV). To take into account the likelihood of skewed representativeness of the patient sample towards high symptom burden, we considered the following hypothetical prevalences of endometriosis in the general population: 0.1%, 0.5%, 1%, and 2% (78).

PAPERS 1-3

A significance level of 5% was used if not otherwise stated. All analyses were performed with IBM SPSS Statistics version 22, except lasso regression analysis (paper 3) which was performed with Stata/SE version 15 and R version 3.5.

Summary of results

1. **PAPER 1: Lack of cross-cultural validity of the Endometriosis Health Profile-30**

   Factor analysis could not support construct validity of the scale *self-image* of the core questionnaire and the scale *treatment* of the modular questionnaire. The Norwegian-version EHP-30 demonstrated acceptable internal consistency and test-retest reliability except for the scale *relationship with children* of the modular questionnaire. Floor effects were observed for the scales *self-image* (20.1%), *work life* (33.9%), *relationship with children* (34.2%), and *medical profession* (20.5%).

   Of the five scales of the Endometriosis Health Profile-30 (*pain*, *control & powerlessness*, *emotional well-being*, *social support*, and *self-image*), the construct self-image does not seem to be measured appropriately by the Norwegian version, suggesting a lack of cross-cultural validity of the Endometriosis Health Profile-30.

2. **PAPER 2: Health-related quality of life in women with endometriosis, compared with the general population and women with rheumatoid arthritis**

   Compared with women from the general population, women with endometriosis had significantly poorer overall quality of life. Pain seemed to be strongly associated with reduced quality of life, infertility only weakly. Compared with women with rheumatoid arthritis <46 years, women with endometriosis had equal overall physical quality of life, yet significantly poorer mean *bodily pain* scores. Compared with all women with rheumatoid arthritis (age 23-96 years), women with endometriosis had better overall physical quality of life yet similar mean *bodily pain* scores. These findings are consistent with the progressive decrease in physical function and increase in pain associated with rheumatoid arthritis. Compared with both younger (<46 years) and all (age 23-96 years) women with rheumatoid arthritis, women with endometriosis had significantly poorer overall mental quality of life.

Among women with endometriosis, there was no significant difference in quality of life between infertile and non-infertile women. Among infertile women with endometriosis, women without children had significantly poorer mental quality of life compared with women with children.

Women with endometriosis seemed to have poorer mental quality of life compared with women with rheumatoid arthritis despite similar pain scores.

3. **PAPER 3: Development of a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis: cross-sectional study**

Regression analysis was used to develop two endometriosis risk indices. The predictors *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* demonstrated the strongest association with disease. Endometriosis Risk Index variant 1 (ERI-1) included these two predictors only. Endometriosis Risk Index variant 2 (ERI-2) included two more: *severe dysmenorrhea in adolescence* and *use of painkillers due to dysmenorrhea in adolescence*. However, these two predictors had the lowest weight among the predictors included in ERI-2. For the hypothetical prevalences of endometriosis in the general population 0.1%, 0.5%, 1%, and 2%, both prediction models "ERI-1 $\geq$ 5" (score range 0-6) and "ERI-2 $\geq$ 33" (score range 0-44) ascertained endometriosis with PPV equal to 2.0%, 9.4%, 17.2%, and 29.6%, respectively, and NPV was at least 98% for all values considered. Thus, no apparent additional value was observed for ERI-2 relative to ERI-1. However, this issue should be investigated in an external validation study.

The predictors *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* demonstrated the strongest association with disease. A prediction model based on these two predictors appears to be a relatively efficient screening tool for endometriosis.

Discussion of findings

The findings in all three studies were somewhat unexpected. Considering that cross-cultural validation of the Endometriosis Health Profile-30 (EHP-30) has been performed for several translated versions, it was a surprise to find that the construct *self-image*, representing one of five scales of the EHP-30, does not seem to be measured appropriately by the Norwegian version, suggesting a lack of cross-cultural validity of the EHP-30. Considering the level of pain and physical dysfunction associated with rheumatoid arthritis, it was a surprise to find women with endometriosis suffering poorer mental quality of life compared with women with rheumatoid arthritis despite similar pain scores. Considering that screening tool development has been a long-term research priority, it was a surprise to find that a prediction model based on frequent *absenteeism from school due to dysmenorrhea* and positive *family history of endometriosis* appears to be a relatively efficient screening tool for endometriosis.

Interpretation of findings – paper 1

Does the EHP-30 scale *self-image* measure self-image?

As mentioned in the section of "Participants, materials, and methods", page 25, validation of an instrument is the process of determining whether there are grounds for believing that the instrument measures what it is intended to measure, that it is useful for its intended purpose (6). Factor analysis of the Norwegian version EHP-30 core questionnaire suggested that the construct self-image does not seem to be measured appropriately (79). In other words, the scale *self-image* of the Norwegian version EHP-30 does not seem to measure self-image. The scale *self-image* consists of three items (listed on the next page). The first two items concern the effect of endometriosis on the choice of clothing and appearance, and the third concerns the effect of endometriosis on confidence. In factor analysis, the first two items loaded on the scale *social support*, while the third loaded on the scale *emotional well-being*. The lack of consistent association between appearance and confidence is convincing, but may vary with age. Subtle differences in exploratory factor analysis technique, i.e. performed with or without predefinition of five factors for the core questionnaire, may have masked a similar finding in other translated versions (22, 69).

Items of the scale self-image of the EHP-30 core questionnaire	
28	… have you … felt frustrated as you cannot always wear the clothes you would choose?
29	… have you … felt your appearance has been affected?
30	… have you … lacked confidence?

Double negative

Double negative is the use of two negatives (i.e. words that mean "no") in the same phrase or sentence. In questionnaires, a double-negative usually creates unnecessary confusion for the respondent. Item 15 of the scale *control & powerlessness*, "have you felt unable to forget your symptoms?", and the response option "never", compose a double negative formulation, and may thus confuse some respondents. Thus, not infrequently, we could observe a tendency of item 15 scoring the polar opposite of the rest of the items of the scale *control & powerlessness*. In questionnaires, double negative formulations should be avoided, so this item may benefit from rephrasing.

Item 15 of the scale control & powerlessness of the EHP-30 core questionnaire	
15	… have you … felt unable to forget your symptoms? – NEVER/ALWAYS? *(Norwegian: … følt deg ute av stand til å glemme symptomene dine? – ALDRI/ALLTID?)*

The Norwegian version EHP-30 – language issues

In factor analysis of the Norwegian version EHP-30 core questionnaire, item 20 of the scale *emotional well-being* loaded on the scale *pain*. In item 20, the English word "miserable" is used. This word has an emotional loading in the English language, and item 20 is therefore part of the scale *emotional well-being*. However, the Norwegian translated word of "miserable" ("elendig") implicates description of a more physical state (as in "I feel really unwell"). This may explain why item 20 loaded on the scale *pain* in our factor analysis. Thus, an alternative Norwegian translated word with a clearer emotional loading (such as "ulykkelig", meaning "unhappy"), may be more suitable in item 20 of the Norwegian version EHP-30.

Item 23 of the scale *emotional well-being* contains the word "violent". However, the Norwegian translated word of "violent" ("voldelig") is not used to describe an emotional state. Thus, this item never achieved maximum score. This may explain the lower item-total correlation demonstrated by item 23 compared with the other items of the same scale. It may be better to remove the word "violent" from item 23 of the Norwegian version EHP-30.

Thorough translation and cultural adaptation procedures were performed according to recommended guidelines (described in more detail in the section "Norwegian version Endometriosis Health Profile-30", page 22) (59). However, the issues raised above suggest that a cultural validation step is important and necessary to ensure adequate measurement properties of a translated version of a quality of life questionnaire.

Items 20 and 23 of the scale emotional well-being of the EHP-30 core questionnaire

20	… have you … felt miserable? *(Norwegian: … følt deg elendig?)*
23	… have you … felt violent or aggressive? *(Norwegian: … følt deg voldelig eller aggressiv?)*

Does the EHP-30 measure quality of life?

The scale *pain* of the EHP-30, comprising 11 items of 30, has a dominating role in the EHP-30. By comparison, the scale *bodily pain* of the SF-36 comprises 2 items of 36. None of the 30 items of the Rheumatoid Arthritis Quality of Life Scale (RAQoL), a disease-specific quality of life questionnaire for rheumatoid arthritis, include the word "pain", even though pain is a dominating symptom in rheumatoid arthritis (80). Thus, pain may represent a disproportionately large part of quality of life in the EHP-30. This raises the question of whether the EHP-30 adequately measures quality of life. In clinical studies of endometriosis, pain is most often the primary endpoint (further detailed in section "Core outcome measures for chronic pain studies", page 36). If one third of the EHP-30 also relates to pain, this may lead to superfluous double counting.

Content validity is the degree to which the content of a quality of life instrument is an adequate reflection of the construct to be measured (i.e. quality of life) (68). Face validity is the degree to which (the items of) a quality of life instrument indeed looks as though they are

an adequate reflection of the construct to be measured (68). Face validity is closely related to content validity. The main distinction is that face validity concerns the critical review of an instrument *after* it has been constructed, while the greater part of content validation consists of ensuring that comprehensive and thorough development procedures have been followed and documented. The EHP-30 claims to measure quality of life, one third of which it relates to pain.

Aside from the question of cross-cultural validity raised in our study, it can also be asked whether the EHP-30 has a face validity problem and perhaps a content validity problem, too. Content validity is optimized by including a wide range of individuals in the development process. Thus, the lack of representativeness of the patient sample used to develop the items of the original EHP-30, may have reduced its content validity. The EHP-30 was developed from in-depth interviews of 25 patients with endometriosis visiting a gynecology clinic at a large tertiary referral hospital in Oxford (16). Patients with endometriosis referred to a large tertiary hospital are likely to be in considerable pain, as (often treatment resistant) pain is one of the most important reasons for referring a patient with endometriosis. Thus, it is not surprising that interviews of these patients would result in a quality of life questionnaire of which a third is about pain. The EHP-30 is argued to be applicable to all patients with endometriosis. However, the phrasing of items of the pain scale (listed on the next page) referring to <u>the</u> pain and not simply pain, suggests that pain has to be present, not taking into account the possibility of absence of pain. This would pose a difficulty in using EHP-30 for evaluating treatments as patients may become pain free (albeit for shorter periods of time) following treatment. The phrasing of items 13-17 (listed on the next page) referring to <u>your</u> symptoms and not simply symptoms, may represent a similar problem.

There may also be an imbalance of scales with reference to quality of life in the EHP-30 modular questionnaire. The modular questionnaire includes the following scales: *work life*, *relationship with children, sexual intercourse, health professionals*, *treatment*, and *infertility*. Regarding face validity, at least two scales appear to be missing: “relationship with partner” and “school life”. For example, in the EndoCost study, a multicenter study of 931 women visiting tertiary care centers because of endometriosis-associated symptoms, one third of the women reported significant problems with their partner due to endometriosis (5).

Items of the scale *pain* of the EHP-30 core questionnaire	
1	… have you … been unable to go to social events **because of the pain**?
2	… have you … been unable to do jobs around the home **because of the pain**?
3	… have you … found it difficult to stand **because of the pain**?
4	… have you … found it difficult to sit **because of the pain**?
5	… have you … found it difficult to walk **because of the pain**?
6	… have you … found it difficult to exercise or do the leisure activities you would like to do **because of the pain**?
7	… have you … lost your appetite and/or been unable eat **because of the pain**?
8	… have you … been unable to sleep properly **because of the pain**?
9	… have you … had to go to bed/lie down **because of the pain**?
10	… have you … been unable to do the things you want to do **because of the pain**?
11	… have you … felt unable to cope **with the pain**?

Items 13-17 of the scale *control & powerlessness* of the EHP-30 core questionnaire	
13	… have you … felt frustrated because **your symptoms** are not getting better?
14	… have you … felt frustrated because you are not able to control **your symptoms**?
15	… have you … felt unable to forget **your symptoms**?
16	… have you … felt as though **your symptoms** are ruling your life?
17	… have you … felt **your symptoms** are taking away your life?

Core outcome measures for chronic pain studies

Developing new quality of life instruments is difficult (6). Thus, the developers of the EHP-30 deserve all credit for their efforts and achievements. However, the challenge should not be underestimated. Fayers and Machin, the authors of the book "Quality of life: The assessment, analysis and reporting of patient-reported outcomes", even advice readers not to develop new instruments unless absolutely necessary (6). They also recommend considering using or building upon existing instruments wherever possible. The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recommended core outcome measures that should be considered in the design of all chronic pain clinical trials (81):

1. Pain
    - 11-point (0-10) numerical rating scale of pain intensity
    - Usage of rescue analgesics
2. Physical functioning (either one of two measures)
    - Multidimensional Pain Inventory Interference Scale
    - Brief Pain Inventory interference items
3. Emotional functioning (at least one of two measures)
    - Beck Depression Inventory
    - Profile of Mood States
4. Participant ratings of global improvement and satisfaction with treatment
    - Patient Global Impression of Change
5. Symptoms and adverse events
    - Passive capture of spontaneously reported adverse events and symptoms and use of open-ended prompts
6. Participant disposition
    - Detailed information regarding participant recruitment and progress through the trial, including all information specified in the CONSORT guidelines

The recommendations above are based primarily on the perspectives of clinicians and researchers. An IMMPACT survey conducted to identify relevant domains of patient-reported outcomes from the perspective of people who experience chronic pain, identified 19 aspects of daily life (82). In addition to pain reduction, the most important aspects were enjoyment of life, emotional well-being, fatigue, weakness, and sleep-related problems (81). Some of these aspects not included in the EHP-30, may be relevant for the quality of life of women with endometriosis.

Interpretation of findings – paper 2

Factors that may explain differences in mental quality of life

Apart from a smaller study with a sample size of 20-30 patients in each patient group, our second study was the first to compare quality of life in women with endometriosis and women with rheumatoid arthritis using a larger sample size (>100).

Considering the level of pain and physical dysfunction associated with rheumatoid arthritis, it seemed reasonable to assume poorer quality of life in women with rheumatoid arthritis compared with women with endometriosis. Rheumatoid arthritis is a systemic, inflammatory, autoimmune disease mostly of older, post-reproductive age (26). It is associated with pain, swelling and stiffness of joints, fatigue, diverse comorbidities as well as a substantially reduced overall quality of life in mental and, even more so, in physical domains compared to general populations (83, 84). However, despite similar pain scores, in our study women with endometriosis were found to have poorer mental quality of life compared with women with rheumatoid arthritis.

According to Norwegian normative SF-36 data from 2015, women have overall poorer quality of life compared to men (85). The gender difference is most pronounced and clinically relevant in the youngest age group (18-29 years), in which women score 10 points lower than men for the scales *bodily pain* and *vitality* (85). For both sexes, high age is positively associated with *vitality* and *mental health* scale scores, and negatively associated with all other scores (85). High education is positively associated will all scores (85). Similarly, according to Swiss normative SF-36 data from 2015, high age, male gender, and absence of chronic conditions, are positively associated with better mental quality of life (86). Young age, high education, and absence of chronic conditions, are positively associated with better physical quality of life (86). Such factors could explain some of the differences in quality of life found in our study between women with endometriosis and women with rheumatoid arthritis. Thus, attributing the differences in mental quality of life associated with endometriosis and rheumatoid arthritis solely to effects of pain, is too simple. However, our findings lend support to the conclusion that endometriosis can have equally negative effects on quality of life as rheumatoid arthritis, albeit different in shape and form. The issue needs further investigation.

Lack of patient's perspective in gynecological research

As mentioned in the introduction, page 13, the lack of public acknowledgement of marked pain and low quality of life experienced by many women with endometriosis may be partly due to the lack of studies comparing pain and quality of life associated with endometriosis and other non-gynecological diseases. However, the general lack of advancement in assessment of outcomes from the patient's perspective in the field of gynecology, may perhaps reflect a general attitude, both in the general public and among health care professionals, of neglecting patient perspectives regarding non-malignant diseases affecting only women (24). It may also reflect a medical profession previously dominated by male physicians with limited personal experience and therefore less understanding of patient perspectives associated with female symptoms and diseases. Further, few other organs can be removed with less impact on physical function or survival than the female reproductive organs. Gynecological diseases may therefore be perceived as less threatening and erroneously assumed to have less impact on quality of life.

In addition to the field of gynecology, less development and evaluation of quality of life instruments was also demonstrated for burns and trauma and intensive care (24). The care offered by these specialties (burns and trauma, intensive care, and gynecology) have in common the tradition of intensive, short-term treatment with no follow-up. Immediate care of burns and trauma and intensive care can all be considered part of emergency medicine. In emergency medicine, lack of follow-up is a natural characteristic of the specialty. This does not apply to gynecology.

Lack of follow-up may be more common in surgical specialties (including gynecology and for example orthopedics) than in medical specialties (such as rheumatology). For example, at the gynecological department of Oslo University Hospital, patients with endometriosis receiving surgical therapy are mostly discharged without follow-up. The tradition of no follow-up for patients with endometriosis is in stark contrast to the traditions of clinical practice in rheumatology, where follow-up is considered an inherent part of the care of chronic diseases such as rheumatoid arthritis. However, in the study assessing availability of quality of life instruments across 30 specialties (mentioned in the introduction), 29 of 3921 reports on development and evaluation of quality of life instruments were from gynecology versus 65 from orthopedics, indicating that gynecology being a surgical specialty cannot be

the only explanation for the general lack of advancement in assessment of outcomes from the patient's perspective.

The recognition of endometriosis as a chronic disease is relatively recent. It seems reasonable to assume that routine follow-up will be included in the treatment of endometriosis at the department of gynecology in the future.

Endometriosis registries

We know relatively little about quality of life associated with endometriosis, even less about long-term consequences of having the disease. Compared with rheumatoid arthritis, endometriosis tends to occur at an earlier age, hitting women from the onset of fertility and during a formative stage when it comes to identity, career, relationships, and family. The indirect consequences associated with endometriosis are therefore probably different and may be substantial. Establishment of endometriosis registries would likely provide such knowledge.

Interpretation of findings – paper 3

Screening for different purposes

Screening is commonly associated with population-based screening programs, such as the Norwegian Breast Cancer Screening Program. The breast cancer screening is performed by inviting women 50 to 69 years of age to biennial mammography screening examinations.

Previous studies on screening tool development for endometriosis have been based on patients attending gynecological surgical departments and infertility clinics (87-91). One of the main objectives has been to improve selection on whom to perform diagnostic laparoscopic surgery, thereby reducing the personal and institutional costs associated with unnecessary procedures. Our prediction model development study differs from previous studies in having as its main focus early identification of endometriosis in the general population.

Balancing the yields and costs of screening

Yields of a screening tool have to be balanced against the costs of screening and the costs of making the confirmatory diagnosis. For breast cancer screening it has been estimated that identifying one case requires examining 170 women by palpation and mammography and taking nine biopsy specimens (92). By comparison, the costs associated with our screening method would be quite low, requiring only information available from medical interview, while the costs of making confirmatory diagnosis by laparoscopy are relatively high.

With mammography screening, all procedures are performed on subjects mainly without symptoms. With endometriosis screening, according to our prediction models, all subjects would be in considerable pain. Even if a diagnostic laparoscopy turns out negative, it would likely be of benefit in diagnostic follow-up of dysmenorrhea causing frequent absenteeism from school or work. Breast cancer takes lives. Endometriosis can have significant impact on quality of life, but is not a malignant disease. Thus, comparing as well as balancing of yields and costs is not at all uncomplicated.

Suitable samples for prediction models aimed at early identification?

It is important for a screening tool for endometriosis to identify a woman with endometriosis that would benefit from early identification of disease. Not all cases of endometriosis may benefit from early detection. A single patch of peritoneal endometriosis identified by surgery for other reasons may not necessarily require treatment or early identification. Thus, for the purposes of prediction model development, it may be important to look for patient samples with a type of endometriosis that would benefit from early identification. These may consist of women with high symptom burdens and/or advanced disease.

The patient sample included in our prediction model development study consists of patients with high symptom burden (78). The findings in a French study suggest that the predictors identified in our study, frequent *absenteeism from school due to dysmenorrhea* and positive *family history of endometriosis*, are more in line with advanced endometriosis than endometriosis in general (46). In a cross-sectional study comparing adolescent markers among women with endometriosis, women with deeply infiltrating endometriosis were found to have higher absenteeism from school during menstruation and a more positive family history of endometriosis than women with superficial peritoneal endometriosis and/or ovarian endometriomas (46). In a genome-wide association study regarding heredity of

endometriosis, moderate and severe endometriosis showed greater genetic burden than minimal or mild endometriosis (93). Thus, our models may be more predictive of advanced endometriosis than of endometriosis in general.

Choice of candidate predictors to include in screening tool development

For a screening tool for endometriosis to enable early diagnosis, the predictors included must be available to most women from a young age. Candidate predictors that exclude too many subjects or are mainly available at a later stage of the disease, may not be suitable.
For example, painful intercourse may exclude too many subjects to be potentially useful as a predictor. According to surveys from 99700 Norwegian high school students from 2016 to 2018, about half have had intercourse by the age of 18, and not all students who have had intercourse, continue to have it regularly (66).

Other examples are information on ultrasound findings, surgical findings, infertility, and previous pregnancies, which are often not available until relatively late in the course of the disease. Most of the diagnostic delay would likely have taken place by the time this information is available. Furthermore, women are delaying parenthood, thereby delaying the potential diagnosis of infertility. In Norway, the mean age at first delivery was 29.5 years in 2018 (94). The total fertility rate, the average number of children born alive per woman in the course of her reproductive period, has declined from 1.98 in 2009 to 1.56 in 2018 (94). Thus, previous pregnancies are becoming less suitable as a candidate predictor.

The limitation of pain as a candidate predictor

In general, reporting of pain, such as intensity or frequency of dysmenorrhea, is subject to substantial individual variation and shown to have limited predictive value. Thus, frequency of dysmenorrhea was not predictive of endometriosis in our analyses when adjusting for other predictors. A study on pain symptoms among women with chronic pelvic pain undergoing diagnostic laparoscopic surgery, illustrates how difficult pain can be as a predictor. Forty pain descriptions derived from in-depth interviews with women with chronic pelvic pain were compared between women with and without endometriosis and between women with superficial and deep endometriosis. Women with deep endometriosis were shown to be more likely to report “shooting rectal pain” and a sense of their “insides pulled down”, compared with women with superficial endometriosis. However, although significant difference was found for the pain description “shooting rectal pain”, no difference was found for the

seemingly similar pain description “sharp rectal pain”. Correspondingly, significant difference was found for “insides pulled down”, but not for “pulling inside” (89).

In our study, the predictors *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* demonstrated the strongest association with disease. Interference of pain with daily life, such as absenteeism from school due to dysmenorrhea, is less common and may be less subject to individual variation. In future prediction model development studies, candidate predictors measuring interference of pain with daily function, may be more suitable than those measuring pain intensity or frequency.

Sensitivity or PPV?

The performance of a screening tool is measured by sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). It is common to only report sensitivity and specificity. However, high sensitivity and specificity do not necessarily entail an acceptable PPV.

PPV measures the proportion of subjects with a positive test who actually have the disease. As diagnosis of endometriosis can only be made by surgery, there is a high likelihood for subjects identified with a positive screening test to undergo a diagnostic laparoscopic procedure. It is therefore particularly important that PPV is high for a screening tool for endometriosis.

	With disease	Without disease	
Test positive	True positive (TP)	False positive (FP)	**PPV** = $\frac{TP}{TP+FP}$
Test negative	False negative (FN)	True negative (TN)	**NPV** = $\frac{TN}{FN+TN}$
	Sensitivity = $\frac{TP}{TP+FN}$	**Specificity** = $\frac{TN}{FP+TN}$	

The table above and the tables below will be used to demonstrate that PPV is highly dependent on prevalence. For a disease with a prevalence range comparable to that of endometriosis, the number of subjects with the disease in any given sample from the general population would be low, giving low numbers of both true positives and false negatives. The

number of individuals without the disease, and thereby the number of false positives and true negatives, would be correspondingly large.

For example, assuming a prevalence of 5% and a population sample of n = 1000, for a screening tool for endometriosis with sensitivity and specificity of 80%, PPV would be 17%:

	With disease (n = 50)	Without disease (n = 950)	
Test positive	True positive = 40	False positive = 190	**PPV** = $\frac{40}{40+190} \approx$ **0.17**
Test negative	False negative = 10	True negative = 760	**NPV** = $\frac{760}{10+760} \approx$ **0.99**
	Sensitivity = 0.80	**Specificity = 0.80**	

Assuming the same prevalence of 5%, increasing sensitivity and specificity to 90% would give a PPV value of 32%:

	With disease (n = 50)	Without disease (n = 950)	
Test positive	True positive = 45	False positive = 95	**PPV** = $\frac{45}{45+95} \approx$ **0.32**
Test negative	False negative = 5	True negative = 855	**NPV** = $\frac{855}{5+855} \approx$ **0.99**
	Sensitivity = 0.90	**Specificity = 0.90**	

Assuming a prevalence of 2% and a population sample of n = 1000, for a screening tool with sensitivity and specificity of 90%, PPV would be 16%:

	With disease (n = 20)	Without disease (n = 980)	
Test positive	True positive = 18	False positive = 98	**PPV** = $\frac{18}{18+98} \approx$ **0.16**
Test negative	False negative = 2	True negative = 882	**NPV** = $\frac{882}{2+882} \approx$ **1.00**
	Sensitivity = 0.90	**Specificity = 0.90**	

By comparison, our prediction models with sensitivity 10.3% and specificity 100% (calculations performed with assumed true specificity of 99.5%), provide a PPV value of 30%, almost two times higher for the same assumed prevalence of 2%.

$$\textbf{PPV} = \frac{\boldsymbol{True\ positive}}{\boldsymbol{True\ positive + False\ positive}} \qquad \textbf{Specificity} = \frac{\boldsymbol{True\ negative}}{\boldsymbol{False\ positive + True\ negative}}$$

From the fraction of PPV above, the number of true positives will be low for endometriosis due to its prevalence. The lower the number of false positives, the higher the PPV. From the fraction of specificity, the lower the number of false positives, the higher the specificity. Thus, the higher the specificity, the lower the number of false positives.

From this and the sample calculations above, we can conclude that for a screening tool for endometriosis it is important that specificity is very high, preferably as close to 100% as possible. High sensitivity does not significantly impact PPV.

Methodological considerations

A happy mistake

All three studies were based on data collected for the study on the evaluation of the measurement properties of the Norwegian version EHP-30. Data from women with and without endometriosis were collected. Attentive readers will have noticed that the publication "Lack of cross-cultural validity of the Endometriosis Health Profile-30" (paper 1) does not include data from women without endometriosis. As mentioned in the introduction, generic quality of life instruments, such as the SF-36, can be applied to a general population. However, disease-specific instruments, such as the EHP-30, cannot. Thus, collection of EHP-30 data from women without endometriosis is a methodological error. This error occurred because we used as a template a study that made the same mistake, the study on the evaluation of the Dutch version EHP-30 (18).

The Dutch study was one of the first evaluation studies of a translated version EHP-30 (18). The Dutch version EHP-30 was applied to women without endometriosis (friends of patients, nurses, and doctors) likely for known-groups validation, a form of construct validation. Simply put, in known-groups validation, a questionnaire is demonstrated to show an expected difference between groups, in this case, between women with endometriosis and healthy control subjects. However, as previously mentioned, the EHP-30 cannot be applied to women without endometriosis. We became aware of this methodological error during data analysis by which time several publications on evaluations of other translated EHP-30s were available and in which data from women without endometriosis were not included.

The importance of luck in research has been described many times before. The error of collecting data from women without endometriosis resulted in two new publications: "Health-related quality of life in women with endometriosis, compared with the general population and women with rheumatoid arthritis" (paper 2) and "Development of a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis: cross-sectional study" (paper 3). The latter in particular offers exciting prospects. Through these studies, we were able to utilize the contribution of all our

participants. Thus, we hope we have been able to make up for the methodological error we made.

Representativeness of the study populations

Possible selection bias is a main weakness of all three studies, and applies to all three study populations: women with endometriosis, women from the general population, and women with rheumatoid arthritis (RA).

First, it is important to note that there is a general issue regarding representativeness of patient samples in endometriosis research. As mentioned in the introduction, definite diagnosis of endometriosis is only possible through visual confirmation of disease during surgery. Endometriosis severity varies and is classified upon surgery commonly as minimal/mild/moderate/severe, or as superficial peritoneal endometriosis/ovarian endometrioma/deeply infiltrating endometriosis (4, 95, 96). The exact ratio of women with minimal/mild/moderate/severe disease or superficial peritoneal endometriosis/ovarian endometrioma/deeply infiltrating endometriosis in a representative sample of women with endometriosis is unknown.

Second, our patient sample was recruited from a patient organization (the Norwegian Endometriosis Association). At the time of patient recruitment, the Norwegian Endometriosis Association was just short of 500 members. The representativeness of the patient sample is likely skewed towards high symptom burden (97). High prevalence of reported dyschezia and irregular bowel movements combined with high prevalence of reported bowel affection in our patient sample, may suggest overrepresentation of women with deeply infiltrating

endometriosis, thereby more moderate to severe disease (88, 98). However, we do believe that participants with mild forms of endometriosis are also included.

Third, by recruiting participants from the Norwegian Endometriosis Association, we minimized the possibility of patients being in active treatment settings when symptoms are often at their extreme states. The alternative recruitment sources to patient associations are gynecological surgical departments and infertility clinics. Apart from the fact that women with milder forms of endometriosis are likely underrepresented in the former and overrepresented in the latter, participants recruited from hospitals are likely in active treatment settings (i.e. undergoing diagnosis/treatment for marked pain or infertility). Both settings imply high levels of stress and would not have been ideal for any of our three studies. Endometriosis registries would have been a preferable recruitment source, as they would provide more representative samples and allow longitudinal observations. However, endometriosis registries do not currently exist in Norway.

Fourth, participants with RA were not excluded in the endometriosis group. RA is a disease of older age. Due to the low prevalence of RA among women of reproductive age, this weakness would have minor effect on the results. A study from UK reported an RA prevalence of 0.12% among women of reproductive age (99). Assuming a similar RA prevalence, it is likely that at most one of 157 participants may have had RA.

First, although the control group was recruited from a randomly selected sample from the general population, the response rate at 14.9% was lower than we preferred. However, the low response rate follows an overall international trend of declining response rates to postal surveys (100). Reminders may have increased the response rate and could have been performed (101).

The quality of life data for women from the general population in our studies are similar to the Norwegian normative SF-36 data collected in 2015 (85). The response rate for the 2015 survey was also low, especially among younger subjects (~20% for those 18-39 years). Comparison of Norwegian normative SF-36 data from 1996, 2002, and 2015, indicates that quality of life has been relatively stable over a 19-year span despite a decline in the overall response rate from 67% in 1996 to 36% in 2015 (85). The 2015 respondents had higher *role-emotional* scale scores (the change mostly represented by the highest age group), and lower *bodily pain* scale scores (the change mostly represented by respondents 30-49 years), than the 1996 and 2002 respondents. The mean *bodily pain* scale scores of our control group are higher than those of female respondents 18-49 years in the 2015 survey. Thus, our selection bias may partly consist of an overrepresentation of individuals without pain in the control group.

The prevalence of *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* in the control group in the present study, were comparable to those found in a Finnish survey involving 1103 adolescent girls from the general population, in which 2.7% reported having a first degree relative with endometriosis, and 5% reported regular absenteeism from school or voluntary activities because of menstrual pain (102).

Second, participants with RA were not excluded in the control group. However, as mentioned above, this weakness would have minor effect on the results.

Participants with endometriosis were not excluded in the RA group. However, due to the low prevalence of endometriosis, and because endometriosis is thought to burn out at menopause, this weakness would have minor effect on the results. In our RA group, 126 of 837 participants (15.1%) were younger than 46 years of age. Assuming an endometriosis

prevalence of ~5% among women of reproductive age, around six (126 × 0.05) of 837 participants (<1%) may have had endometriosis.

No access to medical records

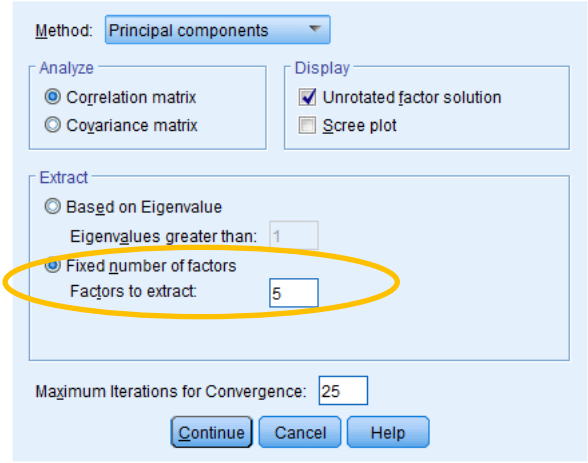
Another limitation of our study is that we were unable to corroborate the reported diagnosis of endometriosis. The questionnaires were anonymous and we did not have access to medical records. Even if we did have access to medical records, we would not have had access to information on classification of disease severity, as such classification is seldom standardly performed in clinical practice at gynecological departments, and if recorded, most often only for research purposes.

Methodological considerations – paper 1

Factor analysis – comparison with other studies

Factor analysis resulted in a three-factor model for the Norwegian version EHP-30 core questionnaire. In other cross-cultural validation studies, factor analysis either confirmed the original five-factor structure or suggested a four-factor model with merger of the scales *pain* and *control & powerlessness*. Thus, our finding deviates from previous ones.

So far, all cross-cultural validation studies of the EHP-30 applying factor analysis techniques have performed exploratory factor analysis, rather than confirmatory factor analysis which may have been more appropriate (discussed in more detail below) (72). Exploratory factor analysis has been performed with forced five-factor extraction in some studies, and without forced five-factor extraction in others. In exploratory factor analysis with forced five-factor extraction, as the term suggests, five factors are forcibly extracted from the data. Using SPSS, this is technically accomplished by choosing the alternative: Extract: Fixed number of factors (please see the figure on the next page).



In a study evaluating the choices of factor analysis techniques used in the validation of SF-36, exploratory factor analysis was considered appropriate if the aim of the study was to explore (hence the term exploratory factor analysis) the factor structure of the SF-36 without a prior hypothesis (69). Confirmatory factor analysis was considered more appropriate if the aim of the study was to confirm the existing first-order 8-factor structure or the second-order 2-factor structure (physical health subscale and mental health subscale) (69). Thus, if exploratory factor analysis is used to validate the EHP-30, it should be performed without a prior hypothesis, i.e. without forced five-factor extraction.

Exploratory factor analysis with forced five-factor extraction seems to have been performed for the American and Portuguese versions of the EHP-30 suggesting a five- and four-factor model, respectively (22, 103). Our study yielding a three-factor model, may simply be due to the fact that we did not employ forced extraction of a fixed number of factors. During the course of our study, we did carry out exploratory factor analysis with forced five-factor extraction on our data. This resulted in a four-factor model very similar to the Portuguese one (please see the table on the next page) (22).

Factor analysis of the 30 items of the EHP-30 core questionnaire with forced five factor extraction, suggesting a four-factor model.

	Items of the EHP-30 core questionnaire	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	Been unable to go to social events because of the pain?	**0.82**				
2	Been unable to do jobs around the home because of the pain?	**0.79**				
3	Found it difficult to stand because of the pain?	**0.80**				
4	Found it difficult to sit because of the pain?	**0.81**				
5	Found it difficult to walk because of the pain?	**0.82**				
6	Found it difficult to exercise or do the leisure activities you would like to do because of the pain?	**0.77**				
7	Lost your appetite and/or been unable to eat because of the pain?	**0.72**				
8	Been unable to sleep properly because of the pain?	**0.71**				
9	Had to go to bed/lie down because of the pain?	**0.77**				
10	Been unable to do the things you want to do because of the pain?	**0.80**				
11	Felt unable to cope with the pain?	**0.74**				
12	Generally felt unwell?	**0.62**				
13	Felt frustrated because your symptoms are not getting better?	**0.57**	0.42	0.44		
14	Felt frustrated because you are not able to control your symptoms?	**0.58**		0.48		
15	Felt unable to forget your symptoms?					**0.74**
16	Felt as though your symptoms are ruling your life?	**0.53**		0.43		0.46
17	Felt your symptoms are taking away your life?	**0.55**				0.47
18	Felt depressed?		**0.69**			
19	Felt weepy/tearful?		**0.66**			
20	Felt miserable?	**0.57**	0.53			
21	Had mood swings?		**0.81**			
22	Felt bad tempered or short tempered?		**0.81**			
23	Felt violent or aggressive?		**0.59**			
24	Felt unable to tell people how you feel?			**0.76**		
25	Felt others do not understand what you are going through?			**0.75**		
26	Felt as though others think you are moaning?	0.43		**0.47**		
27	Felt alone?			**0.65**		
28	Felt frustrated as you cannot always wear the clothes you would choose?				**0.68**	
29	Felt your appearance has been affected?				**0.70**	
30	Lacked confidence?		0.55		**0.63**	

Principal components analysis with varimax rotation. Only factor loadings ≥ 0.40 are shown.
In the original EHP-30, items 1-11 belong to the scale "*pain*", items 12-17 to the scale "*control & powerlessness*", items 18-23 to the scale "*emotional well-being*", items 24-27 to the scale "*social support*" and items 28-30 to the scale "*self-image*".

Factor analysis was not performed for the Brazilian Portuguese (article is not available in English), Australian, or Italian versions (20, 104, 105). For the Dutch, Persian, and Chinese versions, it is not specified whether exploratory factor analysis was performed with or without forced five-factor extraction (18, 19, 21). The only studies clearly stating that exploratory factor analysis was performed without forced five-factor extraction are those of the original English, French, and Norwegian versions, for which four-, five-, and three-factor structures were suggested, respectively (17, 23). Thus, the cross-cultural validation of the five-factor structure of the EHP-30 core questionnaire is far from established.

Exploratory factor analysis vs. confirmatory factor analysis

As mentioned in the previous section, according to a study evaluating the choices of factor analysis techniques used in the validation of SF-36, confirmatory factor analysis was considered more appropriate if the aim of the study was to confirm the existing factor structure (69). Thus, in hindsight, use of confirmatory factor analysis, rather than exploratory factor analysis, may be more appropriate in cross-cultural validation studies such as ours. Furthermore, according to the COSMIN guidelines, confirmatory factor analysis (and not exploratory factor analysis) is recommended for cross-cultural validation of QoL questionnaires (72).

During the course of our study, we did carry out confirmatory factor analysis on our data. The results of confirmatory factor analysis are expressed as several scores of goodness-of-fit (to the hypothesized structure). None of our preliminary scores of goodness-of-fit achieved recommended threshold values. Thus, both exploratory factor analysis without forced five-factor extraction and confirmatory factor analysis suggested lack of cross-cultural validity of the Norwegian version EHP-30. Ultimately, we chose to proceed with exploratory factor analysis without forced five-factor extraction to facilitate comparison with other studies. By using the same technique as other studies, we also believe we have contributed in drawing attention to important methodological issues concerning cross-cultural validation of the EHP-30.

Test-retest reliability: assessment of agreement vs. assessment of reliability

In evaluating test-retest reliability, a clarification of the concepts agreement and reliability is necessary. Agreement points to the question, whether scores are identical or similar or the degree to which they differ (106). In this situation, the absolute degree of measurement error

is of interest. Consequently, any variability between subjects or the distribution of the rated trait in the population does not matter (106). However, for ordered categorical variables (e.g. never, rarely, sometimes, often, always), which constitute the EHP-30, there are really no parameters of measurement error since only the ordering is important (107). Additionally, since there are no units, there are no clear parameters of measurement error (108). Thus, for the EHP-30, assessment of reliability, rather than assessment of agreement, seems appropriate.

Reliability points to the question of how well patients can be distinguished from each other despite measurement errors (109). It is typically defined as the ratio of variability between scores of the same subjects to the total variability of all scores in the sample (106). Thus, a reliability parameter has the typical basic formula (109):

$$\textbf{reliability} = \frac{\textit{variability between subjects}}{\textit{variability between subjects} + \textit{measurement error}}$$

Therefore, reliability parameters (e.g. intraclass correlation coefficients) provide information about the ability of the scores to distinguish between subjects (106). If the measurement error is small compared to the variability between subjects, the reliability parameter approaches 1. This means that the discrimination of the subjects is hardly affected by measurement error, and thus the reliability is high (109).

The US Food and Drug Administration 2009 guidance for industry on patient-reported outcome measures describes how the Agency evaluates the psychometric properties of measures intended to support medical product labelling claims (11). The guidance lists intraclass correlation coefficients (ICCs) and the assessment time period as key considerations for test–retest reliability evaluations. Others also advocate the use of ICC, the reasons for which are discussed in detail elsewhere (108, 110, 111). Thus, we chose to use ICC and paired t-tests to evaluate test-retest reliability of the Norwegian version EHP-30.

Methodological considerations – paper 2

Data collection in 2012/2013 vs 2009

Data collection was performed in 2012/2013 for the endometriosis group and the control group, and in 2009 for the RA group. Oslo Rheumatoid Arthritis Register SF-36 data from 1994 to 2009 indicate a (~16%) significant improvement in physical quality of life among rheumatoid arthritis patients over a 15-year period, mainly attributed to improved rheumatoid arthritis treatment strategies, but not in mental quality of life (57). Thus, the difference in physical quality of life between women with rheumatoid arthritis and the other two groups may be slightly smaller than what is shown in the present study.

SF-36 version 1 and 2

Another weakness of the second study is the application of different versions of SF-36 to the different study populations (SF-36v2 to the endometriosis group and the control group, and SF-36v1 to the RA group). SF-36v2 contains small changes in wording and layout compared to SF-36v1 (64). For items of the scales *role-physical* (RP) and *role-emotional* (RE), the number of response categories was increased from two (yes/no) to five (all of the time/most of the time/some of the time/a little of the time/none of the time), reducing floor and ceiling effects (65). Thus, use of different instruments may mainly have affected the results for the scales RP and RE (64). PCS and MCS are standardized to have a mean of 50 and a standard deviation of 10 based on a general population sample. PCS and MCS of SF-36v2 used in our studies are based on a U.S. general population sample collected in 2009 (65). PCS and MCS of SF-36v1 used in our study are based on a Norwegian general population sample collected in 1996. Thus, PCS and MCS of SF-36v2 and SF-36v1 are not comparable.

Classification of childless and infertile

The EHP-30 modular questionnaire includes six scales, two of which are *relationship with children* and *infertility*. In our study, women who responded to the scale *relationship with children* were classified as having a child/children, and women who reported that the scale was irrelevant were classified as childless. Women who responded to the scale *infertility* were classified as infertile, and women who reported that the scale was irrelevant, were classified as non-infertile. The term non-infertile was used instead of fertile because women who have

not had regular unprotected intercourse for 12 months or longer, and not conceived, may not necessarily be fertile.

The scale *relationship with children* is introduced with: "These questions concern the effect endometriosis has had on your relationship with your child/children during the last 4 weeks. If you do not have any children please tick here and move on to section C". Although a few respondents have added comments about their children being stepchildren, most respondents have likely been correctly classified as having a child/children or being childless. On the other hand, the scale *infertility*, relating to concerns about possible infertility, is introduced with: "These questions concern your problems conceiving during the last 4 weeks. If this section is not relevant to you please tick here." As "problems conceiving" the last 4 weeks do not indicate whether the respondent has tried to conceive for 12 months or longer, a likely misclassification of infertile and non-infertile respondents has occurred. Although the classification of infertile/non-infertile used in our study needs to be replaced with more appropriate terms, the impact on mental quality of life among women with endometriosis caused by concerns regarding possible infertility, is still a question of interest. The impact may also vary depending on clinical and psychological circumstances at the time of responding to the survey. An infertile woman in the middle of IVF treatment, may not give the same response before, after, or between treatments.

Methodological considerations – paper 3

Recall bias

We cannot exclude the possibility of recall bias. Women with endometriosis may be more liable to recall symptoms suggestive of endometriosis experienced in adolescence compared with women without endometriosis.

What could have been done differently?

With the advantage of hindsight, we would have done some things differently. For example, we would have liked to record more variables. The level of education is known to affect SF-

36 scores, and should have been recorded (85). As mentioned previously, reminders could have increased the response rate (101). However, as mentioned in the section of "Methodological Considerations" – "A happy mistake" (page 46), if we had done everything by the book, we would have been without the second and the third study.

Ethical considerations

The study was approved by the Regional Committee for Medical and Health Research Ethics, division southeastern Norway (trial registration number: 2011/2213/Regional Committee for Medical and Health Research Ethics, division south-eastern Norway B).

Questionnaire data was collected from women with endometriosis and women from the general population by anonymous postal prepaid reply. The studies did not involve any person identifying data. Participation did not involve any financial expenditure for the participant.

The Norwegian Endometriosis Association was contacted by Nina Julie Verket with a request to consider participating in the study. A representative of the endometriosis association read through the questionnaire and assessed patient burden. The Norwegian Endometriosis Association subsequently agreed to participate in the study, and kindly provided address labels of their members. The address labels of the members that were not contacted were destroyed after completion of data collection.

The author of the thesis is not in any way involved in the treatment of endometriosis patients at Oslo University Hospital. However, at hindsight, it probably would have been better if a third party had initiated contact with the Norwegian Endometriosis Association to remove any possibility of pressure the Norwegian Endometriosis Association may have felt to participate in the study.

We decided to include 150 women without endometriosis (please see “Sample size calculation”, page 25). We assumed a lowest expected response rate of 10% and applied for a recruiting sample of 1500 women. After approval from the Norwegian Tax Administration, the Norwegian Civil Registry provided names and addresses of a random sample of 1500 women 18-45 years of age living in Oslo, Norway. The list of names and addresses were destroyed after completion of data collection.

Anonymized data (including 14 variables: age, BMI, disease duration, seropositivity, SF-36 scale scores for the eight SF-36 scales, MCS and PCS) from the 837 female respondents of the SF-36 survey conducted among members of the Oslo Rheumatoid Arthritis Registry in 2009 was kindly provided by co-author professor Till Uhlig at the University of Oslo.

Conclusions and clinical implications

The scale *self-image* of the Norwegian version EHP-30 does not seem to measure the construct self-image appropriately. Thus, for Norwegian quality of life studies on women with endometriosis, the EHP-30 must be used with caution.

Women with endometriosis seem to have overall impaired quality of life compared to women from the general population. Pain seems to be a main contributing factor. Women with endometriosis with similar pain to women with rheumatoid arthritis, seem to have poorer mental quality of life than women with rheumatoid arthritis. Comparisons among women with endometriosis seem to suggest that not infertility per se, but a combination of infertility and childlessness markedly reduces mental quality of life. More studies are needed to map the extent of disease burden associated with endometriosis. Both the physical and mental disease burden associated with endometriosis may be larger than previously considered.

The predictors *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* demonstrated the strongest association with disease. A prediction model based on these two predictors appears to be a relatively efficient screening tool to aid primary care physicians in early identification of women at high risk of developing endometriosis. However, the model needs to be validated in future studies before use. If validated, it may shorten diagnostic delay markedly and enable much earlier initiation of appropriate care and follow-up for a subgroup of endometriosis patients.

Future perspectives

Two studies in progress

Two further studies based on the dataset described in the present thesis are in progress. In the first of these two studies, Endometriosis Health Profile-30 (EHP-30) data and Short form-36 (SF-36) data from women from the general population will be used.

The aim of the first study is to evaluate the measurement properties of the EHP-30 applied to women from the general population.

Endometriosis is relatively common, has no disease-specific symptoms, and is difficult to diagnose. Given several studies reporting a mean diagnostic delay of ~7 years, many women in the general population are assumed to have undiagnosed endometriosis. Further, periodically many women with endometriosis may experience little or no pain (for example following treatment), and are difficult to differentiate from women from the general population. Thus, given the nature of endometriosis, it seems reasonable to assume that the disease-specific quality of life questionnaire EHP-30 could be applicable to women from the general population as well as women with endometriosis.

Preliminary results suggest that the EHP-30 applied to women from the general population demonstrate similar measurement properties as the EHP-30 applied to women with endometriosis. If these results are confirmed, a second study comparing EHP-30 data and SF-36 data from women with endometriosis and women from the general population will be performed.

The aim of the second study is to compare the quality of life instruments EHP-30 and SF-36.

The EHP-30 and the generic quality of life questionnaire SF-36 will be compared with regard to how well they differentiate between women with endometriosis and women from the general population. The EHP-30 and the SF-36 will also be compared with regard to their coverage of different aspects of quality of life.

Quality of life instrument development is an iterative process involving ongoing instrument improvement (112). Thus, many instruments undergo development through a number of versions (e.g. Short form-36 version 2), each version being extensively reappraised (6). 19 years have passed since the EHP-30 was first introduced in 2001. To the best of our knowledge, a review of the original instrument based on the validation studies performed has not yet been conducted. We hope that our first publication, "Lack of cross-cultural validity of the Endometriosis Health Profile-30", together with the first and second study described above, will contribute to such a process if it were initiated (79).

External validation of the prediction models

The prediction models need external validation. It is my aim to contribute to an external validation study in the future. As the diagnosis of endometriosis very much is investigator dependent, it is important that such studies be carried out in clinics experienced in handling patients with endometriosis.

A Norwegian Endometriosis Registry

In the future, the design and implementation of a Norwegian endometriosis registry would provide a powerful tool for evidence development in the field of endometriosis. Registries are an important complementary data source that may extend results of clinical trials to populations not included in these studies, demonstrate real-world effects of treatments outside of research settings, and provide longitudinal data not available in clinical trial settings (113).

A patient registry may increase understanding of the natural course of disease which is largely unknown for endometriosis. It may provide an overview of real-world variations in treatment and outcomes, and thereby allow evaluation of clinical effectiveness and cost-effectiveness for different treatments. It may enable exploration of possible factors that may influence prognosis and quality of life. It may serve as a surveillance system for the occurrence of unexpected or unwanted effects of treatments. Through all these functions, a

registry may contribute to improving quality of care provided for women with endometriosis (113).

References

1. Giudice LC. Clinical practice. Endometriosis. N Engl J Med. 2010;362(25):2389-98.
2. Bulun SE. Endometriosis. N Engl J Med. 2009;360(3):268-79.
3. Vercellini P, Vigano P, Somigliana E, Fedele L. Endometriosis: pathogenesis and treatment. Nat Rev Endocrinol. 2014;10(5):261-75.
4. Zondervan KT, Becker CM, Koga K, Missmer SA, Taylor RN, Vigano P. Endometriosis. Nat Rev Dis Primers. 2018;4(1):9.
5. De Graaff AA, D'Hooghe TM, Dunselman GA, Dirksen CD, Hummelshoj L, WERF EndoCost Consortium, et al. The significant effect of endometriosis on physical, mental and social wellbeing: results from an international cross-sectional survey. Hum Reprod. 2013;28(10):2677-85.
6. Fayers PM, Machin D. Quality of life : the assessment, analysis, and reporting of patient-reported outcomes. Third ed. Chichester, West Sussex, UK ; Hoboken, NJ: John Wiley & Sons Inc.; 2016. 626 p.
7. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? Pharmacoeconomics. 2016;34(7):645-9.
8. Preamble to the Constitution of WHO as adopted by the International Health Conference, New York, 19 June–22 July 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of WHO, no. 2. p. 100.) and entered into force on 7 April 1948, (1948).
9. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. Soc Sci Med. 1995;41(10):1403-9.
10. Hays RD, Reeve BB. Measurement and Modeling of Health-Related Quality of Life: Elsevier Inc.; 2008. 241-52 p.
11. FDA. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regist. 2009;74(235):65132-3.
12. Bottomley A, Pe M, Sloan J, Basch E, Bonnetain F, Calvert M, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. Lancet Oncol. 2016;17(11):e510-e4.
13. EMA. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. European Medicines Agency; 2005. Contract No.: Doc. Ref. EMEA/CHMP/EWP/139391/2004.
14. Moinpour CM, Feigl P, Metch B, Hayden KA, Meyskens FL, Jr., Crowley J. Quality of life end points in cancer clinical trials: review and recommendations. J Natl Cancer Inst. 1989;81(7):485-95.
15. Comans TA, Nguyen KH, Mulhern B, Corlis M, Li L, Welch A, et al. Developing a dementia-specific preference--based quality of life measure (AD-5D) in Australia: a valuation study protocol. BMJ Open. 2018;8(1):e018996.
16. Jones G, Kennedy S, Barnard A, Wong J, Jenkinson C. Development of an endometriosis quality-of-life instrument: The Endometriosis Health Profile-30. Obstet Gynecol. 2001;98(2):258-64.
17. Jones G, Jenkinson C, Taylor N, Mills A, Kennedy S. Measuring quality of life in women with endometriosis: tests of data quality, score reliability, response rate and scaling assumptions of the Endometriosis Health Profile Questionnaire. Hum Reprod. 2006;21(10):2686-93.

18. van de Burgt TJ, Hendriks JC, Kluivers KB. Quality of life in endometriosis: evaluation of the Dutch-version Endometriosis Health Profile-30 (EHP-30). Fertil Steril. 2011;95(5):1863-5.
19. Nojomi M, Bijari B, Akhbari R, Kashanian M. The Assessment of Reliability and Validity of Persian Version of the Endometriosis Health Profile (EHP-30). Iran J Med Sci. 2011;36(2):84-9.
20. Maiorana A, Scafidi Fonti GM, Audino P, Rosini R, Alio L, Oliveri AM, et al. The role of EHP-30 as specific instrument to assess the quality of life of Italian women with endometriosis. Minerva Ginecol. 2012;64(3):231-8.
21. Jia S-Z, Leng J-H, Sun P-R, Lang J-H. Translation and psychometric evaluation of the simplified Chinese-version Endometriosis Health Profile-30. Hum Reprod. 2013;28(3):691-7.
22. Nogueira-Silva C, Costa P, Martins C, Barata S, Alho C, Calhaz-Jorge C, et al. Validation of the Portuguese Version of EHP-30 (The Endometriosis Health Profile-30). Acta Med Port. 2015;28(3):347-56.
23. Chauvet P, Auclair C, Mourgues C, Canis M, Gerbaud L, Bourdel N. Psychometric properties of the French version of the Endometriosis Health Profile-30, a health-related quality of life instrument. J Gynecol Obstet Hum Reprod. 2017;46(3):235-42.
24. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. BMJ. 2002;324(7351):1417.
25. Laursen BS, Bajaj P, Olesen AS, Delmar C, Arendt-Nielsen L. Health related quality of life and quantitative pain measurement in females with chronic non-malignant pain. Eur J Pain. 2005;9(3):267-75.
26. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. Lancet. 2010;376(9746):1094-108.
27. Kass A, Hollan I, Fagerland MW, Gulseth HC, Torjesen PA, Forre OT. Rapid Anti-Inflammatory Effects of Gonadotropin-Releasing Hormone Antagonism in Rheumatoid Arthritis Patients with High Gonadotropin Levels in the AGRA Trial. PLoS One. 2015;10(10):e0139439.
28. Taylor HS, Giudice LC, Lessey BA, Abrao MS, Kotarski J, Archer DF, et al. Treatment of Endometriosis-Associated Pain with Elagolix, an Oral GnRH Antagonist. N Engl J Med. 2017;377(1):28-40.
29. Nisenblat V, Prentice L, Bossuyt PM, Farquhar C, Hull ML, Johnson N. Combination of the non-invasive tests for the diagnosis of endometriosis. Cochrane Database Syst Rev. 2016;7:CD012281.
30. Nisenblat V, Bossuyt PM, Farquhar C, Johnson N, Hull ML. Imaging modalities for the non-invasive diagnosis of endometriosis. Cochrane Database Syst Rev. 2016;2:CD009591.
31. Wykes CB, Clark TJ, Khan KS. Accuracy of laparoscopy in the diagnosis of endometriosis: a systematic quantitative review. BJOG. 2004;111(11):1204-12.
32. Husby GK, Haugen RS, Moen MH. Diagnostic delay in women with pain and endometriosis. Acta Obstet Gynecol Scand. 2003;82(7):649-53.
33. Matsuzaki S, Canis M, Pouly JL, Rabischong B, Botchorishvili R, Mage G. Relationship between delay of surgical diagnosis and severity of disease in patients with symptomatic deep infiltrating endometriosis. Fertil Steril. 2006;86(5):1314-6; discussion 7.
34. Hudelist G, Fritzer N, Thomas A, Niehues C, Oppelt P, Haas D, et al. Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. Hum Reprod. 2012;27(12):3412-6.

35. Nnoaham KE, Hummelshoj L, Webster P, d'Hooghe T, de Cicco Nardone F, de Cicco Nardone C, et al. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. Fertil Steril. 2011;96(2):366-73.
36. Staal AH, van der Zanden M, Nap AW. Diagnostic Delay of Endometriosis in the Netherlands. Gynecol Obstet Invest. 2016;81(4):321-4.
37. Practice Committee of American Society for Reproductive M. Treatment of pelvic pain associated with endometriosis. Fertil Steril. 2008;90(5 Suppl):S260-9.
38. Dunselman GA, Vermeulen N, Becker C, Calhaz-Jorge C, D'Hooghe T, De Bie B, et al. ESHRE guideline: management of women with endometriosis. Hum Reprod. 2014;29(3):400-12.
39. Hirsch M, Begum MR, Paniz E, Barker C, Davis CJ, Duffy J. Diagnosis and management of endometriosis: a systematic review of international and national guidelines. BJOG. 2018;125(5):556-64.
40. Meresman GF, Auge L, Baranao RI, Lombardi E, Tesone M, Sueldo C. Oral contraceptives suppress cell proliferation and enhance apoptosis of eutopic endometrial tissue from patients with endometriosis. Fertil Steril. 2002;77(6):1141-7.
41. Zanelotti A, Decherney AH. Surgery and Endometriosis. Clin Obstet Gynecol. 2017;60(3):477-84.
42. Yap C, Furness S, Farquhar C. Pre and post operative medical therapy for endometriosis surgery. Cochrane Database Syst Rev. 2004(3):CD003678.
43. Steenberg CK, Tanbo TG, Qvigstad E. Endometriosis in adolescence: predictive markers and management. Acta Obstet Gynecol Scand. 2013;92(5):491-5.
44. Surrey E, Carter CM, Soliman AM, Khan S, DiBenedetti DB, Snabes MC. Patient-completed or symptom-based screening tools for endometriosis: a scoping review. Arch Gynecol Obstet. 2017;296(2):153-65.
45. Chapron C, Souza C, Borghese B, Lafay-Pillet MC, Santulli P, Bijaoui G, et al. Oral contraceptives and endometriosis: the past use of oral contraceptives for treating severe primary dysmenorrhea is associated with endometriosis, especially deep infiltrating endometriosis. Hum Reprod. 2011;26(8):2028-35.
46. Chapron C, Lafay-Pillet MC, Monceau E, Borghese B, Ngo C, Souza C, et al. Questioning patients about their adolescent history can identify markers associated with deep infiltrating endometriosis. Fertil Steril. 2011;95(3):877-81.
47. Mahmood TA, Templeton A. Prevalence and genesis of endometriosis. Hum Reprod. 1991;6(4):544-9.
48. Eskenazi B, Warner ML. Epidemiology of endometriosis. Obstet Gynecol Clin North Am. 1997;24(2):235-58.
49. Fuldeore MJ, Soliman AM. Prevalence and Symptomatic Burden of Diagnosed Endometriosis in the United States: National Estimates from a Cross-Sectional Survey of 59,411 Women. Gynecol Obstet Invest. 2017;82(5):453-61.
50. Ballard KD, Seaman HE, de Vries CS, Wright JT. Can symptomatology help in the diagnosis of endometriosis? Findings from a national case-control study--Part 1. BJOG. 2008;115(11):1382-91.
51. Abbas S, Ihle P, Koster I, Schubert I. Prevalence and incidence of diagnosed endometriosis and risk of endometriosis in patients with endometriosis-related symptoms: findings from a statutory health insurance-based cohort in Germany. Eur J Obstet Gynecol Reprod Biol. 2012;160(1):79-83.
52. von Theobald P, Cottenet J, Iacobelli S, Quantin C. Epidemiology of Endometriosis in France: A Large, Nation-Wide Study Based on Hospital Discharge Data. Biomed Res Int. 2016;2016:3260952.

53. Morassutto C, Monasta L, Ricci G, Barbone F, Ronfani L. Incidence and Estimated Prevalence of Endometriosis and Adenomyosis in Northeast Italy: A Data Linkage Study. PLoS One. 2016;11(4):e0154227.
54. Eisenberg VH, Weil C, Chodick G, Shalev V. Epidemiology of endometriosis: a large population-based database study from a healthcare provider with 2 million members. BJOG. 2017.
55. Kvien TK, Glennas A, Knudsrod OG, Smedstad LM, Mowinckel P, Forre O. The prevalence and severity of rheumatoid arthritis in Oslo. Results from a county register and a population survey. Scand J Rheumatol. 1997;26(6):412-8.
56. Kvien TK, Uhlig T. The Oslo experience with arthritis registries. Clin Exp Rheumatol. 2003;21(5 Suppl 31):S118-22.
57. Austad C, Kvien TK, Olsen IC, Uhlig T. Health status has improved more in women than in men with rheumatoid arthritis from 1994 to 2009: results from the Oslo rheumatoid arthritis register. Ann Rheum Dis. 2015;74(1):148-55.
58. Norway TLCo. Norwegian: Bokmål vs. Nynorsk 2015 [Available from: http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk/.
59. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94-104.
60. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care. 1993;31(3):247-63.
61. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.
62. Loge JH, Kaasa S, Hjermstad MJ, Kvien TK. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. J Clin Epidemiol. 1998;51(11):1069-76.
63. Ware JE, Kosinski M, Keller SD. Physical and mental health summary scales-a user's manual: The Health Institute, New England Medical Center, Boston, MA; 1994.
64. Jenkinson C, Stewart-Brown S, Petersen S, Paice C. Assessment of the SF-36 version 2 in the United Kingdom. J Epidemiol Community Health. 1999;53(1):46-50.
65. Maruish ME. User's manual for the SF-36v2 Health Survey: Quality Metric Incorporated; 2011.
66. Bakken A. Ungdata. Nasjonale resultater 2018, NOVA Rapport 8/18. Oslo: Norwegian Social Research (NOVA); 2018.
67. Machin D, Campbell MJ, Tan S-B, Tan S-H. Sample size tables for clinical studies. Third ed. Chichester, West Sussex, UK: John Wiley & Sons Ltd; 2009.
68. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
69. de Vet HC, Ader HJ, Terwee CB, Pouwer F. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. Qual Life Res. 2005;14(5):1203-18; dicussion 19-21, 23-4.
70. Ware JE, Jr., Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol. 1998;51(11):1159-65.

71. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11(3):193-205.
72. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19(4):539-49.
73. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res. 2013;22(8):1889-905.
74. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third ed. Harlow, Essex, UK: Pearson Education Limited; 2014.
75. Gandek B, Ware JE, Jr., Aaronson NK, Alonso J, Apolone G, Bjorner J, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol. 1998;51(11):1149-58.
76. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42.
77. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.
78. Verket NJ, Uhlig T, Sandvik L, Andersen MH, Tanbo TG, Qvigstad E. Health-related quality of life in women with endometriosis, compared with the general population and women with rheumatoid arthritis. Acta Obstet Gynecol Scand. 2018;97(11):1339-48.
79. Verket NJ, Andersen MH, Sandvik L, Tanbo TG, Qvigstad E. Lack of cross-cultural validity of the Endometriosis Health Profile-30. J Endometr Pelvic Pain Disord. 2018;10(2):107-15.
80. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. Br J Rheumatol. 1997;36(8):878-83.
81. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain. 2005;113(1-2):9-19.
82. Turk DC, Dworkin RH, Revicki D, Harding G, Burke LB, Cella D, et al. Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain. Pain. 2008;137(2):276-85.
83. Uhlig T, Loge JH, Kristiansen IS, Kvien TK. Quantification of reduced health-related quality of life in patients with rheumatoid arthritis compared to the general population. J Rheumatol. 2007;34(6):1241-7.
84. Matcham F, Scott IC, Rayner L, Hotopf M, Kingsley GH, Norton S, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis. Semin Arthritis Rheum. 2014;44(2):123-30.
85. 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. 2017.
86. Roser K, Mader L, Baenziger J, Sommer G, Kuehni CE, Michel G. Health-related quality of life in Switzerland: normative data for the SF-36v2 questionnaire. Qual Life Res. 2019;28(7):1963-77.

87. Calhaz-Jorge C, Mol BW, Nunes J, Costa AP. Clinical predictive factors for endometriosis in a Portuguese infertile population. Hum Reprod. 2004;19(9):2126-31.
88. Chapron C, Barakat H, Fritel X, Dubuisson JB, Breart G, Fauconnier A. Presurgical diagnosis of posterior deep infiltrating endometriosis based on a standardized questionnaire. Hum Reprod. 2005;20(2):507-13.
89. Ballard K, Lane H, Hudelist G, Banerjee S, Wright J. Can specific pain symptoms help in the diagnosis of endometriosis? A cohort study of women with chronic pelvic pain. Fertil Steril. 2010;94(1):20-7.
90. Nnoaham KE, Hummelshoj L, Kennedy SH, Jenkinson C, Zondervan KT, World Endometriosis Research Foundation Women's Health Symptom Survey C. Developing symptom-based predictive models of endometriosis as a clinical screening tool: results from a multicenter study. Fertil Steril. 2012;98(3):692-701 e5.
91. Lafay Pillet MC, Huchon C, Santulli P, Borghese B, Chapron C, Fauconnier A. A clinical score can predict associated deep infiltrating endometriosis before surgery for an endometrioma. Hum Reprod. 2014;29(8):1666-76.
92. Coggon DD, Rose GA, Barker DJP. Epidemiology for the uninitiated. Fifth ed. London: BMJ; 2003. 73 p.
93. Sapkota Y, Attia J, Gordon SD, Henders AK, Holliday EG, Rahmioglu N, et al. Genetic burden associated with varying degrees of disease severity in endometriosis. Mol Hum Reprod. 2015;21(7):594-602.
94. Norway S. Decline in fertility 2019 [Available from: http://www.ssb.no/en/befolkning/artikler-og-publikasjoner/decline-in-fertility--379997.
95. ASRM. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. Fertil Steril. 1997;67(5):817-21.
96. Nisolle M, Donnez J. Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. Fertil Steril. 1997;68(4):585-96.
97. De Graaff AA, Dirksen CD, Simoens S, De Bie B, Hummelshoj L, D'Hooghe TM, et al. Quality of life outcomes in women with endometriosis are highly influenced by recruitment strategies. Hum Reprod. 2015;30(6):1331-41.
98. Fauconnier A, Chapron C, Dubuisson JB, Vieira M, Dousset B, Breart G. Relation between pain symptoms and the anatomic location of deep infiltrating endometriosis. Fertil Steril. 2002;78(4):719-26.
99. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology (Oxford). 2002;41(7):793-800.
100. Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. Am J Epidemiol. 2006;163(3):197-203.
101. 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 Public Health. 2015;25(1):156-61.
102. Suvitie PA, Hallamaa MK, Matomaki JM, Makinen JI, Perheentupa AH. Prevalence of Pain Symptoms Suggestive of Endometriosis Among Finnish Adolescent Girls (TEENMAPS Study). J Pediatr Adolesc Gynecol. 2016;29(2):97-103.
103. Jenkinson C, Kennedy S, Jones G. Evaluation of the American version of the 30-item Endometriosis Health Profile (EHP-30). Qual Life Res. 2008;17(9):1147-52.
104. Mengarda CV, Passos EP, Picon P, Costa AF, Picon PD. [Validation of Brazilian Portuguese version of quality of life questionnaire for women with endometriosis (Endometriosis Health Profile Questionnaire--EHP-30)]. Rev Bras Ginecol Obstet. 2008;30(8):384-92.

105. Khong SY, Lam A, Luscombe G. Is the 30-item Endometriosis Health Profile (EHP-30) suitable as a self-report health status instrument for clinical trials? Fertil Steril. 2010;94(5):1928-32.
106. Kottner J, Streiner DL. The difference between reliability and agreement. J Clin Epidemiol. 2011;64(6):701-2; author reply 2.
107. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Cambridge: Cambridge University Press; 2011.
108. Hernaez R. Reliability and agreement studies: a guide for clinical investigators. Gut. 2015;64(7):1018-27.
109. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033-9.
110. Kraemer HC. Measurement of reliability for categorical data in medical research. Stat Methods Med Res. 1992;1(2):183-99.
111. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide to their development and use. Fifth edition. ed. Oxford, United Kingdom: Oxford University Press; 2014.
112. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. Clin Pharmacol Ther. 2011;90(5):737-42.
113. Gliklich RE, Dreyer NA, Leavy MB, Quintiles Outcome (Firm), United States. Agency for Healthcare Research and Quality, Effective Health Care Program (U.S.). Registries for evaluating patient outcomes : a user's guide. Rockville, MD: Agency for Healthcare Research and Quality,; 2014. Available from: http://www.ncbi.nlm.nih.gov/books/NBK208616/.

Appendix

Spørreskjema

1

Hvor gammel er du?		år
Hvor høy er du?	,	cm
Hvor mye veier du?	,	kg
Når begynte din siste menstruasjon? (dato for første dag av siste menstruasjon) ***Kryss av her om det ikke er relevant*** ◯	d d m m å å	
Har du menstruasjon nå? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯

Hvor gammel var du da du fikk din første menstruasjon?		år
Hadde du sterke menstruasjonssmerter som tenåring?		Aldri ◯ Sjelden ◯ Noen ganger ◯ Ofte ◯ Alltid ◯
Måtte du være hjemme fra skolen (ungdomsskole/videregående skole) på grunn av menstruasjonssmerter?		Aldri ◯ Sjelden ◯ Noen ganger ◯ Ofte ◯ Alltid ◯
Brukte du smertestillende mot menstruasjonssmerter som tenåring?		Aldri ◯ Sjelden ◯ Noen ganger ◯ Ofte ◯ Alltid ◯
Hvis ja, hvilke:		
Hvis ja, ble smertene helt borte da du brukte smertestillende?	Ja ◯	Nei ◯
Brukte du p-piller ***<u>mot menstruasjonssmerter</u>*** som tenåring?	Ja ◯	Nei ◯
Er det noen i familien din som har endometriose? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯
Hvis ja, hvem:		

Har du endometriose? **Ja** ◯ **Nei** ◯ **Vet ikke** ◯

HVIS JA, hvilket år begynte du å få symptomer? | | | | |

HVIS JA, hvilket år fikk du diagnosen hos legen? | | | | |

HVIS JA, er diagnosen blitt bekreftet ved en operasjon? Ja ◯ Nei ◯

HVIS JA, hvor har eller har du hatt endometriose?

- Vet ikke ◯
- På bukhinnen (peritoneum) ◯
- På eggstokken(e) ◯
- I skjedeveggen ◯
- På tarmen ◯
- På blæren ◯
- Andre steder i underlivet ◯
- Andre steder utenfor underlivet ◯

HVIS JA, hvilke behandlinger har du fått?:

HVIS JA, hva slags behandling(er) får du nå?:

Har du hatt underlivssmerter de siste 4 ukene? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯
Har du vært kvalm eller kastet opp de siste 4 ukene?	Ja ◯	Nei ◯
Har du vært trøtt eller manglet energi de siste 4 ukene?	Ja ◯	Nei ◯
Har du hatt smerter ved vannlating de siste 4 ukene? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯
Har du hatt smerter ved avføring de siste 4 ukene? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯
Har du hatt forstoppelse eller diaré de siste 4 ukene? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯
Har du hatt uregelmessige blødninger de siste 4 ukene? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯
Har du hatt menstruasjonssmerter de siste 4 ukene? ***Kryss av her om det ikke er relevant*** ◯	Ja ◯	Nei ◯

Spørreskjemaet ble fylt ut: (dato) | | | | | | |
d d m m å å

Hvor var du da du svarte på spørreskjemaet?:

Kommentarer:

EHP-30 DEL 1: KJERNESPØRRESKJEMA

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

		Aldri	Sjelden	Noen ganger	Ofte	Alltid
1.	vært ute av stand til å delta i sosiale aktiviteter på grunn av smerter?	☐	☐	☐	☐	☐
2.	vært ute av stand til å utføre arbeider rundt i hjemmet på grunn av smerter?	☐	☐	☐	☐	☐
3.	hatt vanskeligheter med å stå på grunn av smerter?	☐	☐	☐	☐	☐
4.	hatt vanskeligheter med å sitte på grunn av smerter?	☐	☐	☐	☐	☐
5.	hatt vanskeligheter med å gå på grunn av smerter?	☐	☐	☐	☐	☐
6.	hatt vanskeligheter med å trene eller foreta fritidsaktiviteter du har lyst til, på grunn av smerter?	☐	☐	☐	☐	☐
7.	mistet appetitten og/eller vært ute av stand til å spise på grunn av smerter?	☐	☐	☐	☐	☐

Kontroller at du bare har krysset av ***én boks til hvert spørsmål***
før du går videre til neste side

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

	Aldri	Sjelden	Noen ganger	Ofte	Alltid
8. vært ute av stand til å sove skikkelig på grunn av smerter?	☐	☐	☐	☐	☐
9. vært tvunget til å gå til sengs eller legge deg ned på grunn av smerter?	☐	☐	☐	☐	☐
10. vært ute av stand til å gjøre ting du ønsker på grunn av smerter?	☐	☐	☐	☐	☐
11. følt deg ute av stand til å takle smertene?	☐	☐	☐	☐	☐
12. generelt følt deg uvel?	☐	☐	☐	☐	☐
13. følt deg frustrert fordi symptomene dine ikke blir bedre?	☐	☐	☐	☐	☐
14. følt deg frustrert fordi du ikke er i stand til å kontrollere symptomene dine?	☐	☐	☐	☐	☐

Kontroller at du bare har krysset av ***én boks til hvert spørsmål***
før du går videre til neste side

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

	Aldri	Sjelden	Noen ganger	Ofte	Alltid
15. følt deg ute av stand til å glemme symptomene dine?	☐	☐	☐	☐	☐
16. følt det som om symptomene styrer livet ditt?	☐	☐	☐	☐	☐
17. følt det som om symptomene tar fra deg livskvaliteten?	☐	☐	☐	☐	☐
18. følt deg deprimert?	☐	☐	☐	☐	☐
19. følt deg på gråten/gråtkvalt?	☐	☐	☐	☐	☐
20. følt deg elendig?	☐	☐	☐	☐	☐
21. hatt humørsvingninger?	☐	☐	☐	☐	☐
22. vært i dårlig humør eller irritabel?	☐	☐	☐	☐	☐

Kontroller at du bare har krysset av ***én boks til hvert spørsmål***
før du går videre til neste side

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

	Aldri	Sjelden	Noen ganger	Ofte	Alltid
23. følt deg voldelig eller aggressiv?	☐	☐	☐	☐	☐
24. følt deg ute av stand til å fortelle andre hva du føler?	☐	☐	☐	☐	☐
25. følt at andre ikke forstår hva du går gjennom?	☐	☐	☐	☐	☐
26. følt det som om andre syns du klager?	☐	☐	☐	☐	☐
27. følt deg alene?	☐	☐	☐	☐	☐
28. følt deg frustrert fordi du ikke alltid kan bruke de klærne du ønsker å velge?	☐	☐	☐	☐	☐
29. følt at utseendet ditt har blitt påvirket?	☐	☐	☐	☐	☐
30. manglet selvtillit?	☐	☐	☐	☐	☐

Kontroller at du har krysset av ***én boks for hvert spørsmål***
før du går videre til Del 2

Del 2: MODULÆRT SPØRRESKJEMA

Seksjon A: Disse spørsmålene gjelder virkningen endometriose har hatt på arbeidet ditt **de siste fire ukene.** Hvis du ikke har vært i lønnet eller frivillig arbeid de siste fire ukene, kan du krysse av her ☐ og gå videre til Seksjon B.

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

		Aldri	Sjelden	Noen ganger	Ofte	Alltid
1.	vært nødt til å ta fri fra jobben på grunn av smertene?	☐	☐	☐	☐	☐
2.	vært ute av stand til å utføre plikter på jobben på grunn av smertene?	☐	☐	☐	☐	☐
3.	følt deg forlegen på jobben på grunn av symptomene?	☐	☐	☐	☐	☐
4.	hatt skyldfølelse for å ta deg fri fra jobben?	☐	☐	☐	☐	☐
5.	følt deg bekymret over ikke å kunne gjøre jobben din?	☐	☐	☐	☐	☐

Seksjon B: Disse spørsmålene gjelder virkningen endometriose har hatt på forholdet ditt til ditt/dine barn **de siste fire ukene.** Hvis du ikke har noen barn, kan du krysse av her ☐ og gå videre til Seksjon C.

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

		Aldri	Sjelden	Noen ganger	Ofte	Alltid
1.	funnet det vanskelig å ta vare på ditt/dine barn?	☐	☐	☐	☐	☐
2.	vært ute av stand til å leke med ditt/dine barn?	☐	☐	☐	☐	☐

Kontroller at du har ***besvart hver seksjon*** før du går videre til neste side

Seksjon C: Disse spørsmålene gjelder virkningen endometriose har hatt på dine seksuelle forhold **de siste fire ukene.**

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …	**Aldri**	**Sjelden**	**Noen ganger**	**Ofte**	**Alltid**
1. opplevd smerter under eller etter samleie? ***Kryss av her om det ikke er relevant*** ☐	☐	☐	☐	☐	☐
2. følt deg bekymret over å ha samleie på grunn av smertene? ***Kryss av her om det ikke er relevant*** ☐	☐	☐	☐	☐	☐
3. unngått samleie på grunn av smertene? ***Kryss av her om det ikke er relevant*** ☐	☐	☐	☐	☐	☐
4. hatt skyldfølelse fordi du ikke ønsker å ha samleie? ***Kryss av her om det ikke er relevant*** ☐	☐	☐	☐	☐	☐
5. følt deg frustrert fordi du ikke kan nyte samleiet? ***Kryss av her om det ikke er relevant*** ☐	☐	☐	☐	☐	☐

Kontroller at du har ***besvart hver seksjon*** før du går videre til neste side

Seksjon D: Disse spørsmålene gjelder dine følelser overfor medisinsk personell **de siste fire ukene**. Hvis denne seksjonen ikke er relevant for deg, kan du krysse av her ☐ og gå videre til Seksjon E

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

		Aldri	Sjelden	Noen ganger	Ofte	Alltid
1.	følt at legen(e) du bruker ikke gjør noe for deg?	☐	☐	☐	☐	☐
2.	følt at legen(e) mener at alt bare foregår i hodet ditt?	☐	☐	☐	☐	☐
3.	følt deg frustrert over legen(e)s mangel på kunnskaper om endometriose?	☐	☐	☐	☐	☐
4.	følt at du kaster bort legen(e)s tid?	☐	☐	☐	☐	☐

Seksjon E: Disse spørsmålene gjelder dine følelser **de siste fire ukene** om behandlingen du får mot endometriose. Behandling betyr enhver operasjon eller **foreskrevet** medisinering mot endometriosen. Hvis dette ikke er relevant for deg, kan du krysse av her ☐ og gå videre til Seksjon F

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …

		Aldri	Sjelden	Noen ganger	Ofte	Alltid
1.	følt deg frustrert fordi behandlingen ikke gir resultater?	☐	☐	☐	☐	☐
2.	funnet det vanskelig å takle bivirkningene av behandlingen?	☐	☐	☐	☐	☐
3.	følt deg irritert over mengden av behandlinger du måtte få?	☐	☐	☐	☐	☐

Kontroller at du har ***besvart hver seksjon*** før du går videre til neste side

Seksjon F: Disse spørsmålene gjelder dine problemer med å bli gravid **de siste fire ukene**. Hvis denne seksjonen ikke er relevant for deg, kan du krysse av her ☐ og gå videre til Del 3

HVOR OFTE I LØPET AV DE SISTE FIRE UKENE HAR DU PÅ GRUNN AV DIN ENDOMETRIOSE …		**Aldri**	**Sjelden**	**Noen ganger**	**Ofte**	**Alltid**
1.	følt deg bekymret over muligheten til ikke å få barn / flere barn?	☐	☐	☐	☐	☐
2.	følt deg utilstrekkelig fordi du kanskje ikke har vært i stand til å få / kan få barn / flere barn?	☐	☐	☐	☐	☐
3.	følt deg deprimert over muligheten til ikke å få barn / flere barn?	☐	☐	☐	☐	☐
4.	følt at muligheten til ikke å bli gravid / ikke være i stand til å bli gravid har belastet ditt personlige forhold?	☐	☐	☐	☐	☐

Kontroller at du har ***besvart alle seksjonene*** som gjelder for deg.

Din Helse og Trivsel

Dette spørreskjemaet handler om hvordan du ser på din egen helse. Disse opplysningene vil hjelpe oss til å få vite hvordan du har det og hvordan du er i stand til å utføre dine daglige gjøremål. ***Takk for at du fyller ut dette spørreskjemaet!***

For hvert av de følgende spørsmålene vennligst sett et ☒ i den ene luken som best beskriver ditt svar.

1. Stort sett, vil du si at din helse er:

Utmerket	Meget god	God	Nokså god	Dårlig
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

2. <u>Sammenlignet med for ett år siden</u>, hvordan vil du si at din helse stort sett er <u>nå</u>?

Mye bedre nå enn for ett år siden	Litt bedre nå enn for ett år siden	Omtrent den samme som for ett år siden	Litt dårligere nå enn for ett år siden	Mye dårligere nå enn for ett år siden
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

3 De neste spørsmålene handler om aktiviteter som du kanskje utfører i løpet av en vanlig dag. Er din helse slik at den begrenser deg i utførelsen av disse aktivitetene nå? Hvis ja, hvor mye?

	Ja, begrenser meg mye	Ja, begrenser meg litt	Nei, begrenser meg ikke i det hele tatt
a Anstrengende aktiviteter som å løpe, løfte tunge gjenstander, delta i anstrengende idrett	☐ 1	☐ 2	☐ 3
b Moderate aktiviteter som å flytte et bord, støvsuge, gå en tur eller drive med hagearbeid	☐ 1	☐ 2	☐ 3
c Løfte eller bære en handlekurv	☐ 1	☐ 2	☐ 3
d Gå opp trappen flere etasjer	☐ 1	☐ 2	☐ 3
e Gå opp trappen én etasje	☐ 1	☐ 2	☐ 3
f Bøye deg eller sitte på huk	☐ 1	☐ 2	☐ 3
g Gå mer enn to kilometer	☐ 1	☐ 2	☐ 3
h Gå noen hundre meter	☐ 1	☐ 2	☐ 3
i Gå hundre meter	☐ 1	☐ 2	☐ 3
j Vaske eller kle på deg	☐ 1	☐ 2	☐ 3

4. I løpet av de siste 4 ukene, hvor ofte har du hatt noen av de følgende problemer i ditt arbeid eller i andre av dine daglige gjøremål på grunn av din fysiske helse?

	Hele tiden	Mye av tiden	En del av tiden	Litt av tiden	Ikke i det hele tatt
a Du har måttet redusere tiden du har brukt på arbeid eller på andre gjøremål	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b Du har utrettet mindre enn du hadde ønsket	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c Du har vært hindret i å utføre visse typer arbeid eller gjøremål	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d Du har hatt problemer med å gjennomføre arbeidet eller andre gjøremål (f.eks. det krevde ekstra anstrengelser)	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

5. I løpet av de siste 4 ukene, hvor ofte har du hatt noen av de følgende problemer i ditt arbeid eller i andre av dine daglige gjøremål på grunn av følelsesmessige problemer (som f.eks. å være deprimert eller engstelig)?

	Hele tiden	Mye av tiden	En del av tiden	Litt av tiden	Ikke i det hele tatt
a Du har måttet redusere tiden du har brukt på arbeid eller på andre gjøremål	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b Du har utrettet mindre enn du hadde ønsket	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c Du har utførte arbeidet eller andre gjøremål mindre grundig enn vanlig	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

6. I løpet av <u>de siste 4 ukene</u>, i hvilken grad har din fysiske helse eller følelsesmessige problemer hatt innvirkning på din vanlige sosiale omgang med familie, venner, naboer eller foreninger?

Ikke i det hele tatt	Litt	En del	Mye	Svært mye
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

7. Hvor sterke kroppslige smerter har du hatt i løpet av <u>de siste 4 ukene</u>?

Ingen	Meget svake	Svake	Moderate	Sterke	Meget sterke
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

8. I løpet av <u>de siste 4 ukene</u>, hvor mye har smerter påvirket ditt vanlige arbeid (gjelder både arbeid utenfor hjemmet og husarbeid)?

Ikke i det hele tatt	Litt	En del	Mye	Svært mye
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

9. Disse spørsmålene handler om hvordan du har følt deg og hvordan du har hatt det de siste 4 ukene. For hvert spørsmål, vennligst velg det svaralternativet som best beskriver hvordan du har hatt det. Hvor ofte i løpet av de siste 4 ukene har du…

		Hele tiden	Mye av tiden	En del av tiden	Litt av tiden	Ikke i det hele tatt
a	Følt deg full av liv?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b	Følt deg veldig nervøs?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c	Vært så langt nede at ingenting har kunnet muntre deg opp?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d	Følt deg rolig og harmonisk?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
e	Hatt mye overskudd?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
f	Følt deg nedfor og deprimert?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
g	Følt deg sliten?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
h	Følt deg glad?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
i	Følt deg trett?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

10. I løpet av de siste 4 ukene, hvor ofte har din fysiske helse eller følelsesmessige problemer påvirket din sosiale omgang (som det å besøke venner, slektninger osv.)?

Hele tiden	Mye av tiden	En del av tiden	Litt av tiden	Ikke i det hele tatt
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

(IQOLA SF-36v2 Standard, Norway (Norwegian))

11. Hvor RIKTIG eller GAL er <u>hver</u> av de følgende påstander for deg?

	Helt riktig	Delvis riktig	Vet ikke	Delvis gal	Helt gal
a Det virker som om jeg blir syk litt lettere enn andre	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
b Jeg er like frisk som de fleste jeg kjenner	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
c Jeg tror at helsen min vil forverres	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
d Jeg har utmerket helse	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

Takk for at du fylte ut dette spørreskjemaet!

(IQOLA SF-36v2 Standard, Norway (Norwegian))

Papers

I

JEPPD Journal of Endometriosis and Pelvic Pain Disorders

Original Research Article

Lack of cross-cultural validity of the Endometriosis Health Profile-30

Journal of Endometriosis and Pelvic Pain Disorders
2018, Vol. 10(2) 107–115

DOI: 10.1177/2284026518780638
journals.sagepub.com/home/pev
SAGE

Nina Julie Verket[1,2], Marit Helen Andersen[3,4], Leiv Sandvik[5], Tom Gunnar Tanbo[1,6], and Erik Qvigstad[1,7]

Abstract

Introduction: The Endometriosis Health Profile-30 is a disease-specific patient-reported outcome measure of health-related quality of life. Cross-cultural validation of the Endometriosis Health Profile-30 has been performed for several translated versions. The aim of this study was to evaluate the measurement properties of a Norwegian version Endometriosis Health Profile-30.

Methods: This study was designed as a cross-sectional anonymous postal questionnaire study. A total of 157 women with endometriosis were included during a period from 2012 to 2013. Women aged 18–45 years were recruited from the Norwegian Endometriosis Association. Principal components analysis with varimax rotation was used to assess construct validity. Short Form-36 was used to determine convergent validity. Cronbach's alpha was used to measure internal consistency. Intraclass correlation coefficients and paired t-tests were used to evaluate test–retest reliability. Floor and ceiling effects were estimated.

Results: Factor analysis resulted in a three and five-factor model for the core and modular questionnaire, respectively. Factor analysis could not support construct validity of the scales self-image and treatment. The Norwegian version Endometriosis Health Profile-30 demonstrated acceptable internal consistency and test–retest reliability, except for the scale relationship with children. Floor effects were observed for the scales self-image (20.1%), work life (33.9%), relationship with children (34.2%), and medical profession (20.5%).

Conclusion: The construct self-image does not seem to be measured appropriately by the Norwegian version Endometriosis Health Profile-30, suggesting a lack of cross-cultural validity of the Endometriosis Health Profile-30. With multinational studies increasing, adequate translation, cross-cultural adaptation, and cross-cultural validation of instruments are essential to ensure equivalence in languages and cultures other than the original.

Keywords

Endometriosis, Endometriosis Health Profile-30, health-related quality of life, reliability, validity

Date received: 2 December 2017; accepted: 9 May 2018

Introduction

Chronic diseases such as endometriosis can affect health-related quality of life (HRQoL).[1] HRQoL is a multidimensional concept that refers to the patient's general perception of the effect of her disease and treatment on physical, psychological, and social aspects of daily life.[2–4] HRQoL is commonly assessed as a patient-reported outcome, that is, a clinical outcome reported directly by the patient.[3,5] A patient-reported outcome measure (PROM) of HRQoL can be generic, applicable to patients with a variety of conditions, or disease-specific.[6] Disease-specific instruments may detect change in important aspects of certain conditions not accessible by generic

[1]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[2]Research Centre for Obstetrics and Gynecology, Oslo University Hospital, Oslo, Norway
[3]Institute of Health and Society, University of Oslo, Oslo, Norway
[4]Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway
[5]Oslo Center for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway
[6]Department of Reproductive Medicine, Oslo University Hospital, Oslo, Norway
[7]Department of Gynecology, Oslo University Hospital, Oslo, Norway

Corresponding author:
Nina Julie Verket, Research Centre for Obstetrics and Gynecology, Oslo University Hospital Ullevål, Postboks 4956 Nydalen, 0424 Oslo, Norway.
Email: ninaverket@gmail.com

instruments.[7] The Endometriosis Health Profile-30 (EHP-30) is a disease-specific PROM of HRQoL consisting of a core and modular questionnaire.[8,9] The original English version was developed in the United Kingdom and first presented in 2001.[8] The items, or questions, were generated from in-depth interviews of 25 patients with endometriosis visiting a gynecology clinic at a large tertiary referral hospital in Oxford.[8]

The EHP-30 is available in many languages. Evaluation of measurement properties, that is, reliability, validity, and responsiveness, has been performed for several of these, however primarily for the core questionnaire.[10–15] With multinational and multicultural studies increasing, adequate translation, cross-cultural adaptation, and cross-cultural validation are essential to ensure equivalence of a PROM in languages and cultures other than the original.[16] The Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) group has developed user-friendly and easily applicable checklists to evaluate the methodological quality of primary studies on measurement properties.[17] According to these checklists, few, if any, of the EHP-30 validation studies have included adequate sample sizes for test–retest reliability analysis.[18] Test–retest reliability is an important aspect of reliability, ensuring that changes detected by an instrument are not random.[3] However, analysis depends on patients being in stable condition. Although endometriosis is sometimes characterized by disease fluctuation, it is also thought to be stable for longer periods of time. Fewer may be in stable condition among patients attending secondary and tertiary referral centers compared with members of patient registries and patient associations.

The aim of this study was to evaluate the measurement properties of the Norwegian version EHP-30 (NO-EHP-30) and thereby its suitability for future use in endometriosis research in Norway or as part of multinational studies.

Methods

Participants, study design, and data collection

Women with endometriosis were recruited from the Norwegian Endometriosis Association. Inclusion criteria were 18–45 years of age and surgically confirmed diagnosis. Cross-sectional data collection was performed from 2012 to 2013. A set of two anonymous postal questionnaires was sent to potential participants. Each questionnaire included questions on background information, NO-EHP-30, and Short form-36 version 2 (SF-36v2).[19] Participants were asked to fill in the second questionnaire 1 month after completing the first questionnaire, for test–retest reliability analysis. A period of 1 month between the test and retest was chosen to minimize memory effects. A period of 1 month was also thought to increase the chances of the respondents being in the same phase of their menstrual cycle, which in turn may be relevant regarding endometriosis complaints and reporting of HRQoL.

Background information

Background information included age, height, and weight. Diagnostic delay was recorded as year receiving diagnosis minus year the participant started having symptoms. Furthermore, a multiple choice question on organs/anatomic locations affected by endometriosis and two open questions inviting free description of previous and present treatment were included. Finally, the participants were asked whether they had experienced dysmenorrhea, pelvic pain, dysuria, and/or dyschezia during the 4 weeks prior to answering the questionnaire.

EHP-30

The responses are based on patient experiences during the 4 weeks prior to answering the questionnaire. The core questionnaire is composed of 30 items grouped into five scales: pain (11 items), control & powerlessness (6 items), emotional well-being (6 items), social support (4 items), and self-image (3 items). The modular questionnaire is composed of 23 items grouped into 6 scales: work life (5 items), relationship with children (2 items), sexual intercourse (5 items), medical profession (4 items), treatment (3 items), and infertility (4 items). The modular questionnaire is characterized by the possibility of responding only to scales which the patient deems relevant to her. All scales can achieve a minimum score of 0, indicating low disability, and a maximum score of 100, indicating high disability. All items of a scale must be answered to be able to calculate a scale score. The only exception is the scale sexual intercourse, where each item may be relevant independently of the other items of the same scale. Thus, the scale score for the scale sexual intercourse is calculated by omitting items which are not relevant.

Translation and cultural adaptation of the Norwegian version EHP-30

The Norwegian language has two distinct written varieties, "bokmål" and "nynorsk."[20] "Bokmål" is the most commonly used variety. The EHP-30 was therefore translated to "bokmål." The translation and cultural adaptation of the NO-EHP-30 was conducted by Oxford outcomes according to recommended guidelines,[21] (Supplementary material 1).

SF-36v2

The Short form-36 is a generic PROM of HRQoL composed of 36 items, one item assessing health change and 35 items assessing eight health concepts representing eight

scales: physical functioning (10 items), role limitations due to physical problems (4 items), bodily pain (2 items), general health perceptions (5 items), vitality (4 items), social functioning (2 items), role limitations due to emotional problems (3 items), and mental health (5 items).[19,22] All scales can achieve a minimum score of 0, indicating worst possible health, and a maximum score of 100, indicating best possible health. QualityMetric Health Outcomes™ Scoring Software 4.5 from OptumInsight Life Sciences, Inc, was used to score SF-36v2.

Sample size calculation

Correlation coefficients play a central role in this study. We used Fisher's z transformation to estimate 95% confidence interval for a correlation coefficient r.[23] The confidence interval for a correlation coefficient r is widest when $r=0.50$. We consider it sufficient with a precision of ±0.10, that is, when the length of the confidence interval for r is at most 0.20.[10] For a correlation coefficient of 0.50 with a sample of 150 patients, this confidence interval will be 0.40–0.60. We therefore decided to include at least 150 women with endometriosis in our study.

Psychometric evaluation and statistical analysis

Construct validity, reliability, and interpretability of the NO-EHP-30 were assessed. We used the taxonomy, terminology, and definitions of measurement properties suggested by the COSMIN study.[24] Hypotheses-testing was specified as assessment of convergent validity where it could be misinterpreted as hypotheses-testing associated with factor analysis. Reliability was specified as test–retest reliability where it was thought to increase clarity. All analyses were performed with IBM SPSS Statistics, version 22.

Construct validity

Structural validity. Exploratory factor analysis was used to assess structural validity.[25] Principal components analysis with varimax rotation was used to identify the different potential components with eigenvalues greater than 1.[26] Items with factor loadings ⩾0.40 in a factor were included in the factor.

Hypotheses-testing. SF-36v2 was used for hypotheses-testing to assess convergent validity.[17,27,28] We hypothesized the strongest correlations between EHP-30 pain and SF-36v2 bodily pain, and EHP-30 emotional well-being and SF-36v2 mental health. We further expected a strong correlation between EHP-30 social support and SF-36v2 social functioning, and EHP-30 work life and SF-36v2 role-physical. After obtaining the results of the factor analyses, we hypothesized a strong correlation between EHP-30 control & powerlessness and SF-36v2 bodily pain, and EHP-30 relationship with children and SF-36v2 role-physical. Associations between scales of the EHP-30 and the SF-36v2 were calculated by Spearman's rho correlation coefficient. There are no widely accepted criteria for defining a strong versus moderate versus weak correlation.[29] Values 0.20–0.39 were considered to indicate weak correlations, values 0.40–0.59 moderate, values 0.60–0.79 strong, and values 0.80–1.00 very strong correlations.

Reliability

Internal consistency. Cronbach's alpha and corrected item-total correlations were used to measure internal consistency. Cronbach's alpha above 0.70 were considered to indicate acceptable internal consistency reliability for group comparisons, and values above 0.90 for individual comparisons.[28] Item-total correlations were corrected for overlap by omitting the item from the parent scale total. Item-total correlations above 0.40 were considered to indicate acceptable internal consistency.[30]

Test–retest reliability. Intraclass correlation coefficients for agreement and paired t-tests were used to measure test–retest reliability. Intraclass correlation coefficients above 0.70 were considered to indicate acceptable reliability for group comparisons, and values above 0.90 for individual measurements over time.[28,31] Significant differences in mean scores ($p<0.05$) were considered to indicate poor reliability. No significant differences in mean scores were considered to indicate acceptable reliability.

Interpretability

Data completeness, mean scores and standard deviations, floor and ceiling effects, and skewness of score distribution were used to describe the distribution of item responses.[17] Floor or ceiling effects were considered present if more than 15% of respondents scored the minimum value of 0 or the maximum value of 100, respectively.[31]

Ethical approval

This study was approved by the Regional Committee for Medical and Health Research Ethics, division south-eastern Norway (trial registration number: 2011/2213/Regional Committee for Medical and Health Research Ethics, division south-eastern Norway B).

Results

Initially, 150 sets of questionnaires were sent to a random sample of members of the Norwegian Endometriosis Association. Of these, 60 questionnaires were successfully completed and returned. Based on this preliminary

Table 1. Basic characteristics of the participants (n = 157).

Age (years), mean ± 1 SD		35.2 ± 6.5
BMI (kg/m²), mean ± 1 SD		24.8 ± 5.2
Diagnostic delay (years), mean ± 1 SD		8.1 ± 6.5
Diagnosis confirmed by surgery (%)		100
	n	%
Organ affected (n = 148)		
Only peritoneum	10	6.8
Ovaries	98	66.2
Bladder	36	24.3
Vagina	28	18.9
Bowels	54	36.5
Previous treatment (n = 146)		
Analgesic	17	11.6
Hormonal	85	58.2
Surgical	122	83.6
Present treatment (n = 138)		
No treatment	45	32.6
Receiving treatment	93	67.4
Analgesic	28	30.1
Hormonal	73	78.5
Awaiting surgery	4	2.9
Pain experienced past 4 weeks		
Dysmenorrhea (n = 135)	97	71.9
Pelvic pain (n = 152)	129	84.9
Dysuria (n = 154)	52	33.8
Dyschezia (n = 155)	83	53.5

BMI: body mass index; SD: standard deviation.

response rate, an additional 225 sets of questionnaires were sent to a second random sample of members of the Norwegian Endometriosis Association not contacted in the first round. In total, 162 of 375 questionnaires were successfully completed and returned. Five of these were from women with endometriosis who reported that their diagnosis had not been confirmed surgically. These were excluded. Among the 157 included respondents, 94 completed and returned a second questionnaire at a later date. Of these, 10 reported change in treatment or starting new treatment since completing the first questionnaire. Excluding these, test–retest reliability of the NO-EHP-30 could be assessed in 84 of the respondents. The median number of days between answering the first and second questionnaire was 34 (range 7–168). Of the 84 respondents, 61 reported either having menstruation when answering both questionnaires or not having menstruation when answering both questionnaires. Of the 84 respondents, 15 reported having menstruation when answering one questionnaire, and not having menstruation when answering the other. The characteristics of the participants are presented in Table 1.

Construct validity

Structural validity. Factor analysis of the 30 items of the core questionnaire suggested three factors, explaining 70.2% of the total variance. The three-factor model resulted in 20 items loading on the hypothesized scales and 10 items loading on alternative scales (Table 2). Factor analysis of the 23 items of the modular questionnaire suggested five factors, explaining 100% of the total variance. The five-factor model resulted in 15 items loading on the hypothesized scales and 8 items loading on alternative scales (Table 3).

Hypotheses-testing. Correlations between scales of the EHP-30 and the SF-36v2 ranged from −0.63 to −0.81 (Table 4). The correlations are negative because the EHP-30 and the SF-36v2 are scored in opposite directions. All hypotheses were confirmed.

Reliability

Internal consistency. Cronbach's alpha ranged from 0.87 to 0.96 for the original scales of the core questionnaire and from 0.78 to 0.94 for the original scales of the modular questionnaire (supplementary material 2). The corrected item-total correlation coefficients ranged from 0.45 (item 23) to 0.91 for the original scales of the core questionnaire and from 0.55 to 0.89 for the original scales of the modular questionnaire.

Test–retest reliability. Intraclass correlation coefficient for test–retest agreement ranged from 0.80 to 0.85 for the scales of the core questionnaire, and from 0.67 to 0.91 for the scales of the modular questionnaire (Table 5). The mean scale scores did not differ significantly between the first and second measurements. Test–retest reliability analysis including only the 61 respondents reporting either having or not having menstruation when answering both questionnaires, did not alter the general findings (data not shown).

Interpretability

The results are presented in Table 6. Data completeness of at least 97.5% was achieved for all EHP-30 scales. The proportion of participants to whom each scale of the modular questionnaire was relevant, varied from 39.4% (the scale infertility) to 87.2% (the scale sexual intercourse). Floor effect was only found for the scale self-image (20.1%) in the core questionnaire, and for the scales work life (33.9%), relationship with children (34.2%), and medical profession (20.5%) in the modular questionnaire. No ceiling effects were observed. Skewness was low for all the scales.

Discussion

Factor analysis suggested a three-factor model for the EHP-30 core questionnaire, in contrast to the original five-factor model. Items of the scales pain and control & powerlessness loaded on the same factor. A similar finding was

Table 2. Factor analysis of the 30 items of the EHP-30 core questionnaire suggesting a three-factor model.

Items of the EHP-30 core questionnaire		Factor 1	Factor 2	Factor 3
1	Been unable to go to social events because of the pain?	0.84		
2	Been unable to do jobs around the home because of the pain?	0.81		
3	Found it difficult to stand because of the pain?	0.77		
4	Found it difficult to sit because of the pain?	0.79		
5	Found it difficult to walk because of the pain?	0.79		
6	Found it difficult to exercise or do the leisure activities you would like to do because of the pain?	0.78		
7	Lost your appetite and/or been unable to eat because of the pain?	0.71		
8	Been unable to sleep properly because of the pain?	0.72		
9	Had to go to bed/lie down because of the pain?	0.80		
10	Been unable to do the things you want to do because of the pain?	0.82		
11	Felt unable to cope with the pain?	0.75		
12	Generally felt unwell?	0.65		
13	Felt frustrated because your symptoms are not getting better?	0.61	0.44	
14	Felt frustrated because you are not able to control your symptoms?	0.60	0.46	
15	Felt unable to forget your symptoms?	0.41	0.57	
16	Felt as though your symptoms are ruling your life?	0.57	0.63	
17	Felt your symptoms are taking away your life?	0.59	0.63	
18	Felt depressed?		0.41	0.70
19	Felt weepy/tearful?		0.40	0.67
20	Felt miserable?	0.61		0.51
21	Had mood swings?			0.77
22	Felt bad tempered or short tempered?			0.80
23	Felt violent or aggressive?			0.65
24	Felt unable to tell people how you feel?		0.68	
25	Felt others do not understand what you are going through?		0.81	
26	Felt as though others think you are moaning?	0.41	0.44	0.40
27	Felt alone?		0.73	
28	Felt frustrated as you cannot always wear the clothes you would choose?		0.62	
29	Felt your appearance has been affected?		0.62	
30	Lacked confidence?		0.48	0.66

EHP-30: Endometriosis Health Profile-30.
Principal components analysis with varimax rotation. Only factor loadings $\geqslant$0.40 are shown.
In the original EHP-30, items 1–11 belong to the scale "pain," items 12–17 to the scale "control & powerlessness," items 18–23 to the scale "emotional well-being," items 24–27 to the scale "social support," and items 28–30 to the scale "self-image."

demonstrated in the original, Portuguese, and French version EHP-30.[9,14,15] As argued by the developers, it is likely that pain has considerable impact on sense of control and powerlessness. In this study, assessment of convergent validity demonstrated strong correlations between each of the EHP-30 scales pain and control & powerlessness and the SF-36v2 scale bodily pain, supporting this interpretation. Strong correlations were also demonstrated between the EHP-30 scales emotional well-being and social support and the corresponding SF-36v2 scales mental health and social functioning. Thus, the findings in this study support construct validity of four of five scales (pain, control & powerlessness, emotional well-being, and social support) of the core questionnaire.

The fifth scale of the core questionnaire, self-image, consists of three items. The first two items concern the effect of endometriosis on choice of clothing and appearance, and the last item concerns the effect of endometriosis on self-confidence. In factor analysis, the first two items loaded on the scale social support, and the last item loaded on the scale emotional well-being. Thus, the construct self-image does not seem to be measured appropriately by the NO-EHP-30. The lack of association between appearance and self-confidence is likely not exclusive to the Norwegian culture. Subtle differences in exploratory factor analysis technique, that is, performed with or without predefinition of five factors for the core questionnaire, may have masked a similar finding in other translated versions.[14,25]

Factor analysis suggested a five-factor model for the EHP-30 modular questionnaire, in contrast to the original six-factor model. Factor analysis of the modular questionnaire has been performed for the original and French version EHP-30.[9,15] In this study, items of the scales work life and relationship with children loaded on the same factor. A similar finding was demonstrated in the original version, but not in the French version.[9,15] These discrepancies may

Table 3. Factor analysis of the 23 items of the EHP-30 modular questionnaire suggesting a five-factor model.

Items of the EHP-30 modular questionnaire		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
A1	Had to take time off work because of the pain?	0.97				
A2	Been unable to carry out duties at work because of the pain?	0.97				
A3	Felt embarrassed about symptoms at work?	0.51			0.73	
A4	Felt guilty about taking time off work?	0.97				
A5	Felt worried about not being able to do your job?	0.95				
B1	Found it difficult to look after your child/children?	0.97				
B2	Been unable to play with your child/children?	0.97				
C1	Experienced pain during or after intercourse?			0.91		
C2	Felt worried about having intercourse because of the pain?		0.63	0.72		
C3	Avoided intercourse because of the pain?	0.52	0.71			
C4	Felt guilty about not wanting to have intercourse?			0.88		
C5	Felt frustrated because you cannot enjoy intercourse?			0.92		
D1	Felt the doctor(s) you have seen is (are) not doing anything for you?		0.95			
D2	Felt the doctor(s) think it is all in your mind?		0.95			
D3	Felt frustrated at the doctor(s) lack of knowledge about endometriosis?		0.94			
D4	Felt like you are wasting the doctor(s) time?		0.95			
E1	Felt frustrated because treatment is not working?	0.77			0.42	
E2	Found it difficult coping with the side effects of treatment?		–0.90			0.41
E3	Felt annoyed at the amount of treatment you have had to have?		–0.60		0.74	
F1	Felt worried about the possibility of not having children/more children?					0.90
F2	Felt inadequate because you may not/have not been able to have children/more children?	0.56				0.78
F3	Felt depressed at the possibility of not having children/more children?	0.51			0.59	0.46
F4	Felt that the possibility of not conceiving/not being able to conceive has put a strain upon your personal relationship?	0.52			0.77	

EHP-30: Endometriosis Health Profile-30.
Principal components analysis with varimax rotation. Only factor loadings ⩾0.40 are shown.
In the original EHP-30, items A1-5 belong to the scale "work life," items B1-2 to the scale "relationship with children," items C1-5 to the scale "sexual intercourse," items D1-4 to the scale "medical profession," items E1-3 to the scale "treatment," and items F1-4 to the scale "infertility."

be due to difference in daily patterns of work life and child care in these three countries. In this study, factor analysis could not support construct validity of the scale treatment. The three items of the scale treatment loaded on three separate factors. A tendency of the first item of the scale treatment to load on a different factor than the two latter items has been demonstrated by factor analysis with larger samples in both the original and French version EHP-30.[9,15]

The NO-EHP-30 demonstrated acceptable test–retest reliability except for the scale relationship with children of the modular questionnaire, which demonstrated an intraclass correlation coefficient of 0.67. Although the time interval between answering the first and second questionnaire likely was long enough to minimize memory effects, it may have allowed changes in the status of the subject.[32] Exclusion of questionnaires from respondents reporting change in treatment or starting new treatment between assessments, probably reduced this effect. Phase of menstruation did not seem to affect the outcome. The scale relationship with children consists of two items. The second item concerns the ability to play with child/children and implies children of younger age. In the case of children of younger age, the score of this scale may depend not only on the health status of the respondent but also on the health status of the child/children. Thus, this particular scale may be less reliable.

This study is the first to evaluate both test–retest reliability and validity of the core questionnaire of the EHP-30 including adequate sample sizes.[18,33] Regarding the modular questionnaire, the varying relevance of scales to participants has likely rendered some sample sizes inadequate. To ensure adequate sample size for the least relevant modular questionnaire scale, the general sample size should have been three times larger. On the other hand, these variations in relevance of the scales of the modular questionnaire, would limit the use of the modular questionnaire in most research settings. Another weakness of this study is the lack of representativeness of

Table 4. Convergent validity. Correlations between some EHP-30 scales and relevant SF-36v2 scales.

EHP-30 scale	SF-36v2 scale	Spearman's rho	Two-tailed test p-value
Core questionnaire			
Pain	Bodily pain	–0.81	<0.001
Control & powerlessness	Bodily pain	–0.73	<0.001
Emotional well-being	Mental health	–0.74	<0.001
Social support	Social functioning	–0.63	<0.001
Modular questionnaire			
Work life	Role-physical	–0.68	<0.001
Relationship with children	Role-physical	–0.75	<0.001

EHP-30: Endometriosis Health Profile-30; SF-36v2: Short Form-36 version 2.

Table 5. Test–retest reliability and intraclass correlation coefficients (ICC) with 95% confidence intervals (CI) for test–retest agreement. Comparison of mean scale scores at time 1 and time 2 with p-values.

EHP-30 scales	n	ICC[a]	95% CI	Mean ± SD Time 1	Mean ± SD Time 2	p-value[b]
Core questionnaire						
Pain	79	0.80	0.71, 0.87	35.5 ± 24.0	35.6 ± 23.2	0.93
Control & powerlessness	80	0.80	0.70, 0.87	49.8 ± 27.2	47.7 ± 27.4	0.27
Emotional well-being	78	0.84	0.75, 0.89	39.0 ± 21.0	39.2 ± 21.5	0.94
Social support	81	0.85	0.78, 0.90	42.3 ± 26.1	43.2 ± 26.6	0.56
Self-image	81	0.80	0.70, 0.86	39.2 ± 29.1	39.7 ± 28.8	0.80
Modular questionnaire						
Work life	63	0.86	0.77, 0.91	26.3 ± 27.4	27.7 ± 26.5	0.46
Relationship with children	42	0.67	0.47, 0.81	28.9 ± 23.2	31.3 ± 24.7	0.43
Sexual intercourse	65	0.91	0.86, 0.95	47.5 ± 30.5	46.8 ± 31.9	0.65
Medical profession	35	0.75	0.56, 0.86	40.4 ± 29.7	37.0 ± 27.9	0.33
Treatment	37	0.71	0.50, 0.84	45.9 ± 27.8	46.8 ± 24.5	0.79
Infertility	23	0.87	0.73, 0.95	63.9 ± 24.0	61.7 ± 25.6	0.41

ICC: intraclass correlation coefficient; EHP-30: Endometriosis Health Profile-30; CI: confidence interval.
[a]Each ICC was significantly different from zero ($p < 0.001$).
[b]Paired samples t-test, significance two-tailed.

the endometriosis patient group. Participants were recruited from a patient association. Thus, participants with severe forms of endometriosis are likely overrepresented.[34] Recruiting a representative sample of women with endometriosis is a challenge in almost all research settings. Most, if not all, of the EHP-30 validation studies have recruited participants from patient associations and/or from secondary or tertiary referral centers.[10–15] Thus, participants with severe forms of endometriosis are likely overrepresented in all studies, although in varying degree. Moreover, patients attending secondary and tertiary referral centers are more likely to be in active disease and treatment settings, making test–retest reliability analysis difficult. Endometriosis registries would have been a preferable recruitment source to endometriosis associations. However, no endometriosis registry is established in Norway. Furthermore, the responsiveness of the NO-EHP-30 was not evaluated.

The construct self-image does not seem to be measured appropriately by the NO-EHP-30, suggesting a lack of cross-cultural validity of the EHP-30. With multinational and multicultural studies increasing, this study underlines the importance of adequate translation, cross-cultural adaptation, and cross-cultural validation of PROMs.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The present study was funded by University of Oslo.

References

1. De Graaff AA, D'Hooghe TM, Dunselman GA, et al. The significant effect of endometriosis on physical, mental and social wellbeing: results from an international cross-sectional survey. *Hum Reprod* 2013; 28(10): 2677–2685.

Table 6. Interpretability. Data completeness, mean scores and standard deviations (SD), floor and ceiling effects, and skewness of score distribution.

EHP-30 scales	n	$n_{N/R}$[a]	$n_{missing}$	Mean	SD	Floor effect (%)	Ceiling effect (%)	Coefficient of skewness
Core questionnaire								
Pain	156	N/A	1	34.8	24.2	12.2	0	0.01
Control & powerlessness	156	N/A	1	49.1	26.9	7.7	0.6	–0.36
Emotional well-being	154	N/A	3	40.0	21.5	7.1	0	–0.28
Social support	154	N/A	3	41.3	25.8	13.6	0	–0.18
Self-image	154	N/A	3	39.9	28.8	20.1	1.3	0.06
Modular questionnaire								
Work life	124	30	3	28.3	28.5	33.9	0.8	0.60
Relationship with children	79	74	4	26.4	25.3	34.2	1.3	0.65
Sexual intercourse	136	20	1	48.0	29.3	8.1	3.7	–0.09
Medical profession	88	68	1	38.3	29.9	20.5	2.3	0.15
Treatment	83	73	1	45.0	26.9	10.8	1.2	–0.18
Infertility	61	94	2	62.7	27.1	4.9	9.8	–0.52

N/R: not relevant; N/A: not applicable; EHP-30: Endometriosis Health Profile-30.
[a]Number of participants for whom the scale was not relevant (only applicable for the modular questionnaire).

2. Bottomley A, Pe M, Sloan J, et al. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol* 2016; 17(11): e510–e514.
3. FDA. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. *Fed Regist* 2009; 74(235): 65132–65133.
4. Hays RD and Reeve BB. *Measurement and modeling of health-related quality of life*. Amsterdam: Elsevier Inc., 2008.
5. European Medicines Agency (EMA). *Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products*. London: EMA, 2005.
6. McHorney CA, Ware JE Jr and Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; 31(3): 247–263.
7. Comans TA, Nguyen KH, Mulhern B, et al. Developing a dementia-specific preference–based quality of life measure (AD-5D) in Australia: a valuation study protocol. *BMJ Open* 2018; 8(1): e018996.
8. Jones G, Kennedy S, Barnard A, et al. Development of an endometriosis quality-of-life instrument: the Endometriosis Health Profile-30. *Obstet Gynecol* 2001; 98(2): 258–264.
9. Jones G, Jenkinson C, Taylor N, et al. Measuring quality of life in women with endometriosis: tests of data quality, score reliability, response rate and scaling assumptions of the Endometriosis Health Profile Questionnaire. *Hum Reprod* 2006; 21(10): 2686–2693.
10. van de Burgt TJ, Hendriks JC and Kluivers KB. Quality of life in endometriosis: evaluation of the Dutch-version Endometriosis Health Profile-30 (EHP-30). *Fertil Steril* 2011; 95(5): 1863–1865.
11. Nojomi M, Bijari B, Akhbari R, et al. The assessment of reliability and validity of Persian version of the Endometriosis Health Profile (EHP-30). *Iran J Med Sci* 2011; 36(2): 84–89.
12. Maiorana A, Scafidi Fonti GM, Audino P, et al. The role of EHP-30 as specific instrument to assess the quality of life of Italian women with endometriosis. *Minerva Ginecol* 2012; 64(3): 231–238.
13. Jia S-Z, Leng J-H, Sun P-R, et al. Translation and psychometric evaluation of the simplified Chinese-version Endometriosis Health Profile-30. *Hum Reprod* 2013; 28(3): 691–697.
14. Nogueira-Silva C, Costa P, Martins C, et al. Validation of the Portuguese Version of EHP-30 (The Endometriosis Health Profile-30). *Acta Med Port* 2015; 28(3): 347–356.
15. Chauvet P, Auclair C, Mourgues C, et al. Psychometric properties of the French version of the Endometriosis Health Profile-30, a health-related quality of life instrument. *J Gynecol Obstet Hum Reprod* 2017; 46(3): 235–242.
16. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine* 2000; 25(24): 3186–3191.
17. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; 19(4): 539–549.
18. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012; 21(4): 651–657.
19. Loge JH, Kaasa S, Hjermstad MJ, et al. Translation and performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. I. Data quality, scaling assumptions, reliability, and construct validity. *J Clin Epidemiol* 1998; 51(11): 1069–1076.
20. The Language Council of Norway, http://www.sprakradet.no/Vi-og-vart/Om-oss/English-and-other-languages/English/norwegian-bokmal-vs.-nynorsk/ (accessed 1 December 2017).
21. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the

ISPOR Task Force for Translation and Cultural Adaptation. *Value Health* 2005; 8(2): 94–104.
22. Jenkinson C, Stewart-Brown S, Petersen S, et al. Assessment of the SF-36 version 2 in the United Kingdom. *J Epidemiol Community Health* 1999; 53(1): 46–50.
23. Machin D, Campbell MJ, Tan S-B, et al. *Sample size tables for clinical studies*. 3rd ed. Chichester: John Wiley & Sons, 2009.
24. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010; 63(7): 737–745.
25. de Vet HC, Ader HJ, Terwee CB, et al. Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Qual Life Res* 2005; 14(5): 1203–1218; discussion 19–21, 23–24.
26. Ware JE Jr, Kosinski M, Gandek B, et al. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; 51(11): 1159–1165.
27. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res* 2013; 22(8): 1889–1905.
28. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002; 11(3): 193–205.
29. Portney LG and Watkins MP. *Foundations of clinical research: applications to practice*. 3rd ed. Harlow: Pearson Education, 2014.
30. Gandek B, Ware JE Jr, Aaronson NK, et al. Tests of data quality, scaling assumptions, and reliability of the SF-36 in eleven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol* 1998; 51(11): 1149–1158.
31. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60(1): 34–42.
32. Fayers PM and Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2nd ed. Chichester: John Wiley & Sons, 2007.
33. The COSMIN checklist with 4-point rating scale, http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20with%204-point%20scale%2022%20juni%202011.pdf (accessed 1 December 2017).
34. De Graaff AA, Dirksen CD, Simoens S, et al. Quality of life outcomes in women with endometriosis are highly influenced by recruitment strategies. *Hum Reprod* 2015; 30(6): 1331–1341.

PAPER 1 – Supplementary material 1

Supplementary material 1: Translation and cultural adaptation of the Norwegian version EHP-30

The translation and cultural adaptation of the Norwegian version EHP-30 (NO-EHP-30) was conducted by Oxford Outcomes and included the following:

1. *Forward translations*: The EHP-30 was translated into Norwegian by two independent translators who are native Norwegian speakers. The two forward translations were reconciled into a third translation by an in-country investigator. Any issues that arose from this stage were discussed with the Oxford Outcomes project manager.
2. *Back translations*: The reconciled translation was back translated into English by two independent translators who are native English speakers and fluent in the Norwegian language and who had no prior knowledge of the EHP-30. The back translations were reviewed against the original EHP-30 by the Oxford Outcomes project manager. Any issues arising from this review were passed to the in-country investigator for comment.
3. *Clinician review*: The translation was passed to a clinician specializing in the appropriate area in Norway. The clinician reviewed the translation to ensure that it was linguistically and culturally appropriate for use in Norway, and that it was acceptable for use with patients. Any suggestions or issues were passed to the in-country investigator who, in conjunction with the Oxford Outcomes project manager, worked to resolve any problems and further refine the translation.
4. *Developer review*: The translation was reviewed by the instrument developer.
5. *Cognitive debriefing*: The translation was given to five women with endometriosis in Norway who are all native Norwegian speakers. They were asked to read through and complete the NO-EHP-30. Following completion, the women were asked a series of questions aimed at gauging their comprehension of the wording of the translation. The answers, along with any other relevant comments and suggestions were summarized in a report, followed by review of results by the Oxford Outcomes project manager. Any issues arising were sent to the in-country investigator for further review or revision.

PAPER 1 – Supplementary material 2

Supplementary material 2 (table): Internal consistency. Cronbach's alpha (α) for the EHP-30 scales and corrected item-total correlations between the EHP-30 items and their scales

Core questionnaire	Corrected item-total correlation[a]
PAIN (α = 0.96, n = 156)	
Item 1	0.87
Item 2	0.88
Item 3	0.81
Item 4	0.78
Item 5	0.80
Item 6	0.85
Item 7	0.72
Item 8	0.79
Item 9	0.87
Item 10	0.91
Item 11	0.81
CONTROL & POWERLESSNESS (α =0.92, n = 156)	
Item 12	0.73
Item 13	0.83
Item 14	0.83
Item 15	0.63
Item 16	0.83
Item 17	0.85
EMOTIONAL WELL-BEING (α = 0.91, n = 154)	
Item 18	0.82
Item 19	0.81
Item 20	0.76
Item 21	0.83
Item 22	0.80
Item 23	0.45
SOCIAL SUPPORT (α = 0.87, n = 154)	
Item 24	0.69
Item 25	0.84
Item 26	0.59
Item 27	0.78
SELF-IMAGE (α = 0.88, n = 154)	
Item 28	0.80
Item 29	0.80
Item 30	0.72

Modular questionnaire	Corrected item-total correlation[a]
WORK LIFE (α = 0.93, n = 124)	
Item A1	0.83
Item A2	0.86
Item A3	0.72
Item A4	0.85
Item A5	0.89
RELATIONSHIP WITH CHILDREN (α = 0.92, n = 79)	
Item B1	0.86
Item B2	0.86
SEXUAL INTERCOURSE (α =0.94, n = 129)	
Item C1	0.77
Item C2	0.87
Item C3	0.84
Item C4	0.82
Item C5	0.87
MEDICAL PROFESSION (α = 0.92, n = 88)	
Item D1	0.82
Item D2	0.75
Item D3	0.81
Item D4	0.87
TREATMENT (α = 0.78, n = 83)	
Item E1	0.55
Item E2	0.63
Item E3	0.69
INFERTILITY (α = 0.91, n = 61)	
Item F1	0.79
Item F2	0.82
Item F3	0.83
Item F4	0.78

[a] Pearson, all correlations were significant ($p < 0.001$)

III

BMJ Open

Development of a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis: cross-sectional study

Nina Julie Verket ,[1,2] Ragnhild Sørum Falk,[3] Erik Qvigstad,[1,4] Tom Gunnar Tanbo,[1,5] Leiv Sandvik[3]

To cite: Verket NJ, Falk RS, Qvigstad E, *et al*. Development of a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis: cross-sectional study. *BMJ Open* 2019;**9**:e030346. doi:10.1136/bmjopen-2019-030346

► Prepublication history and additional material for this paper are available online. To view, please visit the journal (http://dx.doi.org/10.1136/bmjopen-2019-030346).

Received 10 March 2019
Revised 18 October 2019
Accepted 22 October 2019

[1]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[2]Research Center for Obstetrics and Gynecology, Oslo University Hospital, Oslo, Norway
[3]Oslo Center for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway
[4]Department of Gynecology, Oslo University Hospital, Oslo, Norway
[5]Department of Reproductive Medicine, Oslo University Hospital, Oslo, Norway

Correspondence to
Dr Nina Julie Verket;
ninaverket@gmail.com

ABSTRACT

Objectives To identify predictors of disease among a few factors commonly associated with endometriosis and if successful, to combine these to develop a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis.

Design Cross-sectional anonymous postal questionnaire study.

Setting Women aged 18–45 years recruited from the Norwegian Endometriosis Association and a random sample of women residing in Oslo, Norway.

Participants 157 women with and 156 women without endometriosis.

Main outcome measures Logistic and least absolute shrinkage and selection operator (LASSO) regression analyses were performed with endometriosis as dependent variable. Predictors were identified and combined to develop a prediction model. The predictive ability of the model was evaluated by calculating the area under the receiver operating characteristic curve (AUC) and positive predictive values (PPVs) and negative predictive values (NPVs). To take into account the likelihood of skewed representativeness of the patient sample towards high symptom burden, we considered the hypothetical prevalences of endometriosis in the general population 0.1%, 0.5%, 1% and 2%.

Results The predictors *absenteeism from school due to dysmenorrhea* and *family history of endometriosis* demonstrated the strongest association with disease. The model based on logistic regression (AUC 0.83) included these two predictors only, while the model based on LASSO regression (AUC 0.85) included two more: *severe dysmenorrhea in adolescence* and *use of painkillers due to dysmenorrhea in adolescence*. For the prevalences 0.1%, 0.5%, 1% and 2%, both models ascertained endometriosis with PPV equal to 2.0%, 9.4%, 17.2% and 29.6%, respectively. NPV was at least 98% for all values considered.

Conclusions External validation is needed before model implementation. Meanwhile, endometriosis should be considered a differential diagnosis in women with frequent absenteeism from school or work due to painful menstruations and positive family history of endometriosis.

Strengths and limitations of this study

- ► The present study is the first to identify and combine predictors of endometriosis to develop a prediction model that may be used in primary care.
- ► A randomly selected sample from the general population was used to recruit control subjects.
- ► We did not have access to medical records.
- ► Possible recall and selection bias cannot be excluded.
- ► External validation is needed before model implementation.

INTRODUCTION

Endometriosis is a chronic inflammatory gynaecological disease with an estimated prevalence of ~5% among women of childbearing age.[1 2] Tissue similar to the inner lining of the uterus in aberrant locations can cause pain, most frequently painful menstruations and painful intercourse, and infertility.[3] Disease onset can be as early as adolescence, with disease persistence throughout reproductive age until a presumed burnout at menopause. Both disease expression and disease progression can vary markedly.[2] There is no cure, and symptomatic treatment can vary from occasional use of over-the-counter painkillers to multiple extensive surgeries with adhesiolysis and organ resection or removal.[4] Thus, the potential consequences of early-onset progressive endometriosis can be substantial and can last for multiple decades.[5 6]

Endometriosis is difficult to diagnose because painful menstruations, painful intercourse and infertility are common among too many without endometriosis. To date, the only way of diagnosing endometriosis is visual confirmation of abnormal patches of tissue during surgery.[7] Thus, it is not surprising that

for some it may take years before endometriosis is diagnosed, prolonging patient uncertainty and delaying treatment and care.[8–10] It follows from the lack of diagnostic tools that the longest delay takes place in primary care.[5 11]

Screening tools are often developed for screening of general populations. However, in the field of endometriosis, screening tool development has been confined to women attending secondary and tertiary gynaecological surgical units or infertility clinics.[12 13] Even if successful, screening tools developed from such studies would not be applicable in primary care due to the requirement of specialised examinations, such as ultrasound, MRI or surgery.[14] In the present study, we used a control group from the general population. Our objectives were to identify predictors of disease among a few factors commonly associated with endometriosis and available to physicians through medical interview and, if successful, to combine these to develop and internally validate a prediction model to aid primary care physicians in early identification of women at high risk of developing endometriosis.

PARTICIPANTS AND METHODS

Study design and data collection

Cross-sectional data collection was performed from 2012 to 2013. A postal questionnaire for anonymous reply was sent to women with endometriosis and a random sample of women from the general population.

Study populations

Women with endometriosis were recruited from the Norwegian Endometriosis Association. Inclusion criteria were 18–45 years of age and surgically confirmed diagnosis. In total, 162 of 375 women successfully completed and returned the questionnaire. Among these, five reported that their diagnosis had not been confirmed surgically and were excluded. Thus, 157 women with endometriosis were included, representing a response rate of 41.9% (online supplementary flow chart).

Following approval from the Norwegian Tax Administration, the Norwegian Civil Registry provided names and addresses of a random sample of women aged 18–45 years living in Oslo, Norway. Inclusion criteria were 18–45 years of age and no known diagnosis of endometriosis. In total, 159 of 1050 women successfully completed and returned the questionnaire. Although the survey included a letter asking only women without endometriosis to participate, three women reported having endometriosis and were excluded. Thus, 156 women without endometriosis were included, representing a response rate of 14.9% (online supplementary flow chart).

Basic characteristics

Background information included age, height, weight and symptoms (dysmenorrhea, pelvic pain, dysuria, dyschezia, fatigue, nausea, irregular menstrual bleeding and irregular bowel movement) experienced at any time during the 4 weeks prior to answering the questionnaire. For participants with endometriosis, diagnostic delay was recorded as the year receiving diagnosis minus the year the participant started having symptoms. Disease duration was recorded as the year of data collection minus the year receiving diagnosis. Further, the questionnaire included a multiple choice question on organs/anatomical locations affected by endometriosis, and two open questions inviting free description of previous and present treatments.

Candidate predictors

The candidate predictors were chosen based on three criteria: (1) they had to be applicable to most, if not all, female adolescents; by this criterion, variables such as dyspareunia (according to surveys from 99 700 Norwegian high school students from 2016 to 2018, about half have had intercourse by the age of 18), ultrasound/MRI findings, surgical findings, infertility and previous pregnancies were excluded as candidate predictors[15]; (2) they had to be simple and comprehensible to young adolescents, without the need for supplementary explanation; by this criterion, variables such as pelvic pain (eg, we were not confident in adolescents' ability to readily localise symptoms as from the pelvis) and the concept of cyclic versus non-cyclic symptoms were excluded; and (3) they had to be available from early stages of the disease and reasonably frequent; by this criterion, variables such as dysuria and dyschezia were excluded. The following candidate predictors (with the questions (Q) and answer (A) alternatives given in parentheses) were included in the final questionnaire:

1. Age at menarche
   (Q: How old were you when you had your first period?)
2. Severe dysmenorrhea in adolescence
   (Q: Did you have very painful periods as a teenager?)
   (A: never/rarely/sometimes/often/always)
3. Absenteeism from school due to dysmenorrhea
   (Q: Did you have to be absent from school—junior high school/high school—because of painful periods?)
   (A: never/rarely/sometimes/often/always)
4. Use of painkillers due to dysmenorrhea in adolescence
   (Q: Did you use painkillers for painful periods as a teenager?)
   (A: never/rarely/sometimes/often/always)
5. Use of oral contraceptives due to dysmenorrhea in adolescence
   (Q: Did you use oral contraceptives because of painful periods as a teenager?)
   (A: yes/no)
6. Family history of endometriosis
   (Q: Does anyone in your family have endometriosis?)
   (A: yes/no/irrelevant)

Statistical analysis

Data were presented as mean with SD for continuous variables and as frequencies with percentages for categorical variables. Continuous variables were compared using independent samples t-test. Categorical variables were

compared using Pearson's χ^2 test. Ordered categorical variables were compared using linear-by-linear association χ^2 test.

Development of risk indices

Two different approaches were used to develop two risk indices: Endometriosis Risk Index Variant 1 (ERI-1), based on logistic regression analysis, and Endometriosis Risk Index Variant 2 (ERI-2), based on least absolute shrinkage and selection operator (LASSO) regression analysis. Logistic regression analysis is one of the most frequently used methods to develop prediction models by selecting relevant predictors and combining them statistically into a multivariable model.[16] However, logistic regression may overestimate performance. We therefore applied LASSO regression analysis, a penalisation procedure that performs both variable selection and regularisation, during model development, as recommended in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) checklist for developing and validating prediction models.[16]

In the regression analyses, *age at menarche* was included as a continuous variable. To increase test power, the ordered categorical variables *severe dysmenorrhea in adolescence* and *absenteeism from school due to dysmenorrhea* were included as continuous variables based on linearity of the beta coefficients, supporting the assumption of the categories (never/rarely/sometimes/often/always) being equally spaced. The ordered categorical variable *use of painkillers due to dysmenorrhea in adolescence* was recoded into three categories (never/rarely, sometimes and often/always) based on deviations from linearity of the beta coefficients. *Use of oral contraceptives due to dysmenorrhea in adolescence* was included as a dichotomous (yes/no) variable. The categorical variable *family history of endometriosis* was recoded into two categories (yes and no/irrelevant/missing) to be able to handle the real-life response category 'irrelevant' (eg, if adopted). Missing responses were also included in this dichotomous categorisation due to the likelihood of blank responses being comparable to participants simply not knowing. Participants with complete data for the candidate predictors according to the previous mentioned description were included in the analyses (154 cases and 145 controls). Further, a sensitivity analysis was performed, that is, a reanalysis with an alternative dichotomous categorisation (yes/no) for the categorical variable *family history of endometriosis*, excluding the responses irrelevant and missing (142 cases and 130 controls).

First, univariable and multivariable logistic regression analyses were performed to assess the relationship between the six candidate predictors and endometriosis. Backward stepwise variable selection was performed using $p \leq 0.157$ as criterion (corresponding to Akaike information criteria). The results were presented as beta coefficients and ORs with 95% CIs based on 1000 bootstrap samples. ERI-1 was based on the relative ratio between the beta coefficients. Second, LASSO regression analysis was performed with 10-fold cross-validation and 1000 bootstrap samples, as implemented in the R package *mami*. The results were presented as means of the LASSO regression coefficients with 95% CIs. ERI-2 was based on the relative ratios between the LASSO regression coefficients.

Internal validation

The predictive abilities of the two risk indices, ERI-1 and ERI-2, were described by area under the receiver operating characteristic curve (AUC). Sensitivity and specificity for different cut-off values of the risk indices were calculated, as well as positive predictive values (PPVs) and negative predictive values (NPVs). To take into account the likelihood of skewed representativeness of the patient sample towards high symptom burden,[17] we considered the following hypothetical prevalences of endometriosis in the general population: 0.1%, 0.5%, 1% and 2%. Participants with complete data for the predictors included in ERI-1 and ERI-2 (155 cases and 148 controls) were included in the analyses.

A significance level of 5% was used if not otherwise stated. All analyses were performed with IBM SPSS Statistics V.22, STATA/SE V.15 and R V.3.5.

Patient and public involvement

A representative of the Norwegian Endometriosis Association assessed the readability and the respondent burden of the questionnaire prior to survey administration. Patients were not consulted to interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

RESULTS

Basic characteristics of the participants

Basic characteristics of the participants are presented in tables 1–3. All 157 participants with endometriosis reported surgically confirmed diagnosis. Of these, 123 reported previous or present affection of one or both ovaries, bladder, vagina and/or bowels. To an open question inviting free description of previous treatment, 122 reported surgical treatment. Of these, 33 reported specific surgical procedures, including 18 hysterectomies, 12 oophorectomies (11 unilateral and 1 bilateral), 5 cystectomies of endometriomas and 7 partial colectomies.

Candidate predictors

Responses to the candidate predictors are presented in table 2. Blank responses were described as missing. In the control group, six participants skipped an entire page of the questionnaire (including the candidate predictors), most likely by error, and therefore had blank responses for all candidate predictors.

Regarding family history of endometriosis in the endometriosis group, 42 participants reported positive family history; 102 reported negative family history; 5 answered irrelevant; and 8 did not answer at all (however, seven of

Table 1 Recent characteristics of the participants

Variable	Endometriosis group n=157		Control group n=156		P value
Age (years), mean±1 SD	35.2±6.5		32.6±6.5		<0.001 *
Body mass index (kg/m^2), mean±1 SD	24.8±5.2		23.4±4.1		0.02*
Dysmenorrhea,† n (%)	97	(71.9%)	66	(43.4%)	<0.001‡
Pelvic pain,† n (%)	129	(84.9%)	29	(19.2%)	<0.001‡
Dysuria,† n (%)	52	(33.8%)	6	(3.9%)	<0.001‡
Dyschezia,† n (%)	83	(53.5%)	17	(11.0%)	<0.001‡
Fatigue,† n (%)	143	(91.1%)	91	(59.1%)	<0.001‡
Nausea,† n (%)	73	(46.5%)	30	(19.2%)	<0.001‡
Irregular menstrual bleeding,† n (%)	45	(32.4%)	22	(14.7%)	<0.001‡
Irregular bowel movement,† n (%)	105	(68.2%)	37	(24.2%)	<0.001‡

*Independent samples t-test.
†Experienced at any time during the 4 weeks prior to answering the questionnaire.
‡Pearson's χ^2 test. Because of missing values, the calculated percentages may not refer to the total number of participants.

these eight had written 'I don't know' as a comment in the answer field). Of the 42 who reported positive family history, 41 specified nature of kinship (reporting one to three relatives each). Nineteen reported a mother, 13 a sister, 9 one or more aunts, 4 a grandmother, 3 a cousin, 2 a parent's cousin, 1 a niece and 1 a great aunt. In total, 28 of 41 (68.3%) reported one or more first-degree relatives with endometriosis. In the control group, 7 participants reported positive family history; 126 reported negative family history; 8 answered irrelevant; and 15 did not answer at all. Of the seven who reported positive family history, six reported one or more sisters, one a mother and one a cousin. All seven reported one or more first-degree relatives with endometriosis.

Development of ERI-1 using logistic regression analysis

Based on univariable logistic regression analysis, use of painkillers due to dysmenorrhea in adolescence, family history of endometriosis, use of oral contraceptives due to dysmenorrhea in adolescence, absenteeism from school due to dysmenorrhea and severe dysmenorrhea in adolescence were the strongest predictors of endometriosis (table 4). Multivariable logistic regression analysis with backward stepwise variable selection procedure resulted in two predictors: absenteeism from school due to dysmenorrhea (A) and family history of endometriosis (F). Based on the relative ratio between the beta coefficients (A:F ratio was 1.1:2.3, rounded to 1:2), the following risk index was developed and assigned scores from 0 to 6:

ERI-1=A+2F, where

- A=absenteeism from school due to dysmenorrhea (never=0 points, rarely=1 point, sometimes=2 points, often=3 points, always=4 points)
- F=family history of endometriosis (yes=1 point, not yes=0 points).

Development of ERI-2 using LASSO regression analysis

Based on LASSO regression analysis, four predictors were selected: severe dysmenorrhea in adolescence, absenteeism from school due to dysmenorrhea, use of painkillers due to dysmenorrhea in adolescence (the categories often or always) and family history of endometriosis (table 4). Based on the relative ratios between the means of the LASSO regression coefficients, the following risk index was developed and assigned scores from 0 to 44:

ERI-2=D+6A+2P+14F, where

- D: severe dysmenorrhea in adolescence (never=0 points, rarely=1 point, sometimes=2 points, often=3 points, always=4 points).
- A: absenteeism from school due to dysmenorrhea (never=0 points, rarely=1 point, sometimes=2 points, often=3 points, always=4 points).
- P: use of painkillers due to dysmenorrhea in adolescence (never/rarely/sometimes=0 points, often/always=1 point).
- F: family history of endometriosis (yes=1 point, not yes=0 points).

Logistic and LASSO regression analyses, including participants with complete data for the candidate predictors, who only responded 'yes' or 'no' to the candidate predictor 'family history of endometriosis' (142 cases and 130 controls), did not alter the findings (online supplementary table).

Internal validation

The AUC was 0.83 and 0.85 for ERI-1 and ERI-2, respectively. Sensitivities and specificities for different cut-off values for ERI-1 and ERI-2 are presented in tables 5 and 6. Estimated specificities for ERI-1 with a cut-off of ≥5 (ERI-1≥5) and ERI-2 with a cut-off of ≥33 (ERI-2≥33) were 100%. As a true specificity of 100% is highly unlikely, we chose a value of 99.5% when calculating PPV for ERI-1≥5 and ERI-2≥33.

Table 2 Adolescent characteristics and family history of the participants

Variable		Endometriosis group n=157		Control group n=156		P value
Age at menarche (years), mean±1 SD		12.7±1.5		13.0±1.6		0.11*
	Missing, n (%)	1	(0.6%)	7	(4.5%)	
Severe dysmenorrhea in adolescence, n (%)	Never	5	(3.2%)	30	(20.1%)	
	Rarely	13	(8.3%)	36	(24.2%)	
	Sometimes	31	(19.9%)	43	(28.9%)	<0.001†
	Often	45	(28.8%)	21	(14.1%)	
	Always	62	(39.7%)	19	(12.8%)	
	Missing	1	(0.6%)	7	(4.5%)	
Absenteeism from school due to dysmenorrhea, n (%)	Never	28	(17.8%)	99	(66.4%)	
	Rarely	23	(14.6%)	26	(17.4%)	
	Sometimes	52	(33.1%)	17	(11.4%)	<0.001†
	Often	38	(24.2%)	5	(3.4%)	
	Always	16	(10.2%)	2	(1.3%)	
	Missing	0	(0%)	7	(4.5%)	
Use of painkillers for dysmenorrhea in adolescence, n (%)	Never	20	(12.8%)	56	(37.6%)	
	Rarely	15	(9.6%)	30	(20.1%)	
	Sometimes	36	(23.1%)	40	(26.8%)	<0.001†
	Often	39	(25.0%)	10	(6.7%)	
	Always	46	(29.5%)	13	(8.7%)	
	Missing	1	(0.6%)	7	(4.5%)	
Use of oral contraceptives for dysmenorrhea in adolescence, n (%)	Yes	60	(38.2%)	17	(11.5%)	<0.001‡
	No	97	(61.8%)	131	(88.5%)	
	Missing	0	(0%)	8	(5.1%)	
Family history of endometriosis, n (%)	Yes	42	(26.8%)	7	(4.5%)	<0.001‡
	Not yes§	115	(73.2%)	149	(95.5%)	

*Independent samples t-test.
†Linear-by-linear association χ^2 test.
‡Pearson's χ^2 test.
§Not yes: no/irrelevant/missing.

For each hypothetical prevalence, PPV and NPV were calculated for ERI-1 cut-off values of 2, 3, 4 and 5 (table 5) and for ERI-2 cut-off values of 12, 19, 26 and 33 (table 6). The highest cut-off value provided the highest PPV. For the prevalences of 0.1%, 0.5%, 1% and 2%, both prediction models 'ERI-1≥5' (score range 0–6) and 'ERI-2≥33' (score range 0–44) ascertained endometriosis with PPVs equal to 2.0%, 9.4%, 17.2% and 29.6%, respectively. For both indices, PPV was low for the cut-off value that provided the highest sensitivity. NPV was at least 98% for all values considered (tables 5 and 6). In the present dataset, 16 of 155 participants with endometriosis achieved ERI-1≥5 and ERI-2≥33. Among participants without endometriosis, the highest achieved ERI-1 and ERI-2 scores were 4 and 32, respectively.

DISCUSSION

Statement of principal findings

In the present study, regression analysis was used to develop two endometriosis risk indices. The predictors absenteeism from school due to dysmenorrhea and family history of endometriosis demonstrated the strongest association with disease. ERI-1 included these two predictors only. ERI-2 included two more: severe dysmenorrhea in adolescence and use of painkillers due to dysmenorrhea in adolescence. These two predictors had the lowest weight among the predictors included in ERI-2. For the hypothetical prevalences of endometriosis in the general population of 0.1%, 0.5%, 1% and 2%, both prediction models ERI-1≥5 (score range 0–6) and ERI-2≥33 (score range 0–44) ascertained endometriosis with PPVs equal to 2.0%, 9.4%, 17.2% and 29.6%, respectively, and NPV was at least 98% for all values considered. Thus, no

Table 3 Further characteristics of the endometriosis group

Diagnostic delay (years), mean±1 SD		8.1±6.5
Disease duration (years), mean±1 SD		6.6±5.0
Diagnosis confirmed by surgery		100%
Organ affected* (N=148)		
Only peritoneum, n (%)	10	(6.8%)
Ovaries, n (%)	98	(66.2%)
Bladder, n (%)	36	(24.3%)
Vagina, n (%)	28	(18.9%)
Bowels, n (%)	54	(36.5%)
Previous treatment† (N=146)		
Analgesic, n (%)	17	(11.6%)
Hormonal, n (%)	85	(58.2%)
Surgical, n (%)	122	(83.6%)
Present treatment† (N=138)		
No treatment, n (%)	45	(32.6%)
Receiving treatment, n (%)	93	(67.4%)
Analgesic, n (%)	28	(30.1%)
Hormonal n (%)	73	(78.5%)
Awaiting surgery, n (%)	4	(2.9%)

*Multiple choice question.
†Open question inviting free description.

apparent additional value was observed for ERI-2 relative to ERI-1. However, this issue should be investigated in an external validation study. For the predictor family history of endometriosis, comments from participants suggest that 'I don't know' should be included as a response category (in addition to 'yes', 'no' and 'irrelevant') in future studies.

Strengths and weaknesses of the study

A major strength of the present study is that it is the first to identify predictors of endometriosis which may be used in primary care. When developing prediction models, high PPV is preferable to high sensitivity and specificity. Thus, cut-off values for the risk indices providing the highest PPV were chosen. Depending on the prevalence, the prediction models may identify women at high risk of developing endometriosis with PPVs comparable to that of mammography screening, where PPVs close to 15% are common.[18] However, a sensitivity close to 10% is lower than we would prefer. Still, our patient sample has previously been demonstrated to carry a high disease burden, with marked pain and low health-related quality of life, comparable to or worse than women with rheumatoid arthritis, but with the disease hitting them at a much younger age.[17] Thus, we have a patient sample representing a subtype of endometriosis that would undoubtedly benefit from early diagnosis and treatment. Hence, a screening tool with a sensitivity of 10% seems much better

Table 4 Logistic and LASSO regression analyses of candidate predictors of endometriosis

	Univariable logistic regression		Multivariable logistic regression		Logistic regression with backward stepwise selection‡		LASSO regression	
Candidate predictors	**B**	**OR (95% CI)**	**B**	**OR (95% CI)**	**B**	**OR (95% CI)**	**B**	**(95% CI)**
Intercept			–2.6	0.1 (0.0 to 0.9)	–1.5	0.2 (0.1 to 0.3)	–1.5	(–4.3 to –0.5)
Age at menarche (years)	–0.1	0.9 (0.8 to 1.0)	0.1	1.1 (0.9 to 1.3)				
Severe dysmenorrhea* (cont.)	0.8	2.2 (1.8 to 2.8)	0.2	1.2 (0.9 to 1.8)			0.1	(0.0 to 0.5)
Absenteeism from school† (cont.)	1.1	3.0 (2.2 to 3.9)	0.9	2.5 (1.6 to 3.7)	1.1	3.0 (2.3 to 4.1)	0.8	(0.5 to 1.2)
Use of painkillers† (ref. never/rarely)								
Sometimes	0.9	2.3 (1.2 to 4.5)	–0.2	0.8 (0.4 to 2.0)				
Often/always	2.3	9.8 (5.2 to 18.7)	0.2	1.3 (0.5 to 3.5)			0.3	(0.0 to 1.0)
Use of oral contraceptives†	1.6	4.8 (2.6 to 8.8)	0.1	1.1 (0.5 to 2.6)				
Family history of endometriosis	2.2	8.7 (3.2 to 23.5)	2.2	9.4 (2.9 to 30.6)	2.3	9.5 (3.1 to 29.2)	1.7	(1.0 to 3.0)

Only participants with complete data for the candidate predictors (154 cases and 145 controls) were included in the analyses.
*Experienced in adolescence.
†Due to dysmenorrhea in adolescence.
‡Backward stepwise variable selection was performed using Wald test statistics with p≤0.157 as the criterion for inclusion.
cont., continuous; LASSO, least absolute shrinkage and selection operator; ref., reference.

Table 5 PPVs and NPVs for ERI-1 (score range 0–6) with cut-off values 2, 3, 4 and 5 for different possible prevalences of endometriosis

	ERI-1≥2		ERI-1≥3		ERI-1≥4		ERI-1≥5	
	Sensitivity 76.8% Specificity 79.7%		Sensitivity 45.2% Specificity 92.6%		Sensitivity 24.5% Specificity 98.0%		Sensitivity 10.3% Specificity 100.0%	
Possible prevalences (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)
2.0	7.2	99.4	11.1	98.8	20.0	98.5	29.6*	98.2
1.0	3.7	99.7	5.8	99.4	11.0	99.2	17.2*	99.1
0.5	1.9	99.9	3.0	99.7	5.8	99.6	9.4*	99.5
0.1	0.4	100.0	0.6	99.9	1.2	99.9	2.0*	99.9

Only participants with complete data for the predictors included in ERI-1 and ERI-2 (155 cases and 148 controls) were included in the analyses.
*PPV for ERI-1 cut-off ≥5 was calculated using a specificity of 99.5%, not 100.0%.
ERI-1, Endometriosis Risk Index Variant 1; ERI-2, Endometriosis Risk Index Variant 2; NPV, negative predictive value; PPV, positive predictive value.

than the alternative of no screening tool. Cut-offs giving a sensitivity and a specificity of ~80% provided an unacceptable PPV of ~3%.

Our study has several weaknesses. First, we did not have access to medical records. Thus, severity of endometriosis could not be assessed. A second weakness is that we cannot exclude the possibility of recall bias. Women with endometriosis may be more liable to recall symptoms suggestive of endometriosis experienced in adolescence compared with women without endometriosis. A third weakness is the low response rate from the general population, following an overall international trend of declining response rates to postal surveys.[19] Thus, the control group may not be completely randomly selected even though random procedures were used for selection. However, the prevalences of absenteeism from school due to dysmenorrhea and family history of endometriosis in the control group in the present study were comparable to those found in a Finnish survey involving 1103 adolescent girls from the general population, in which 2.7% reported having a first degree relative with endometriosis, and 5% reported regular absenteeism from school or voluntary activities because of painful menstruation.[20]

Comparison with other studies

Previous studies on screening tool development have not included control groups from the general population and have not been intended for use in primary care settings, making comparisons of findings difficult.[12 13 21–23] In general, reporting of pain, such as frequency of dysmenorrhea, is subject to substantial individual variation and is expected to be of limited predictive value. However, interference of pain with daily life, such as absenteeism from school due to dysmenorrhea, is less common and likely less subject to individual variation. The choice of the response options 'never', 'rarely', 'sometimes', 'often' and 'always' to the question on frequency of absenteeism from school, although seldom used in other studies, has most likely been suitable. Endometriosis has an estimated total heritability of about 50%.[24 25] It is therefore not surprising that a positive family history of endometriosis

Table 6 PPVs and NPVs for ERI-2 (score range 0–44) with cut-off values 12, 19, 26 and 33 for different possible prevalences of endometriosis

	ERI-2≥12		ERI-2≥19		ERI-2≥26		ERI-2≥33	
	Sensitivity 78.1% Specificity 79.7%		Sensitivity 45.2% Specificity 92.6%		Sensitivity 24.5% Specificity 98.0%		Sensitivity 10.3% Specificity 100.0%	
Possible prevalences (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)	PPV (%)	NPV (%)
2.0	7.3	99.4	11.1	98.8	20.0	98.5	29.6*	98.2
1.0	3.7	99.7	5.8	99.4	11.0	99.2	17.2*	99.1
0.5	1.9	99.9	3.0	99.7	5.8	99.6	9.4*	99.5
0.1	0.4	100.0	0.6	99.9	1.2	99.9	2.0*	99.9

Only participants with complete data for the predictors included in ERI-1 and ERI-2 (155 cases and 148 controls) were included in the analyses.
*PPV for ERI-2 cut-off ≥33 was calculated using a specificity of 99.5%, not 100.0%.
ERI-1, Endometriosis Risk Index Variant 1; ERI-2, Endometriosis Risk Index Variant 2; NPV, negative predictive value; PPV, positive predictive value.

is required for both prediction models to identify women at high risk of developing endometriosis.

The predictors identified in the current study are in line with a French study, however more so for advanced endometriosis than for endometriosis in general.[26] In a cross-sectional study comparing adolescent markers among women with endometriosis, women with deeply infiltrating endometriosis were found to have a more positive family history of endometriosis (OR 3.2) and higher absenteeism from school during menstruation (OR 1.7) than women with superficial peritoneal endometriosis and/or ovarian endometriomas.[26] In a genome-wide association study regarding heredity of endometriosis, moderate and severe endometrioses showed greater genetic burden than minimal or mild endometriosis.[27] Thus, our models may be more predictive of advanced endometriosis than of endometriosis in general. The prevalence of deep endometriosis is assumed to be ~2%,[2 28] which may be a bit overstated according to some prevalence studies.[29–32] Thus, the chosen range of hypothetical prevalences in the present study seems appropriate.

Future research

More studies on screening tool development for endometriosis including control groups from the general population are needed. Register studies should be encouraged. However, newer candidate predictors such as absenteeism from school due to dysmenorrhea with suitable response options may not always be available. In view of the diversity of endometriosis, different subtypes may require different prediction models.

CONCLUSIONS AND CLINICAL IMPLICATIONS

The developed prediction models need to be validated in future studies before use. Meanwhile, endometriosis should be considered a differential diagnosis in women with frequent absenteeism from school or work due to dysmenorrhea and positive family history of endometriosis.

Persevering or increasing interference of pain with daily life should prompt referral to secondary or tertiary care clinics experienced in handling endometriosis patients.

Dissemination declaration

We aim to disseminate the results in the Norwegian Endometriosis Association newsletter. If the prediction models are validated, primary care physicians will be informed through national health care and primary care physician websites. School nurses will be informed through school nurse networks, including presentation at the annual national school nurse conference.

Acknowledgements We gratefully acknowledge the contribution of Karen Bertelsen of the Norwegian Endometriosis Association and the association itself.

Contributors Study concept and design: NJV, RSF and LS. Acquisition of data: NJV. Analysis and interpretation of data: NJV, RSF, EQ, TGT and LS. Drafting of manuscript: NJV, RSF and LS. The final manuscript was critically revised and approved by all authors.

Funding The present study was funded by the University of Oslo.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was approved by the Regional Committee for Medical and Health Research Ethics, division south-eastern Norway (trial registration number: 2011/2213/Regional Committee for Medical and Health Research Ethics, division south-eastern Norway B).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The data used in the present study is part of a larger dataset. Due to ongoing data analysis, the data used in the present study will not be available until all data analysis is completed. The corresponding author can be contacted for details.

ORCID iD

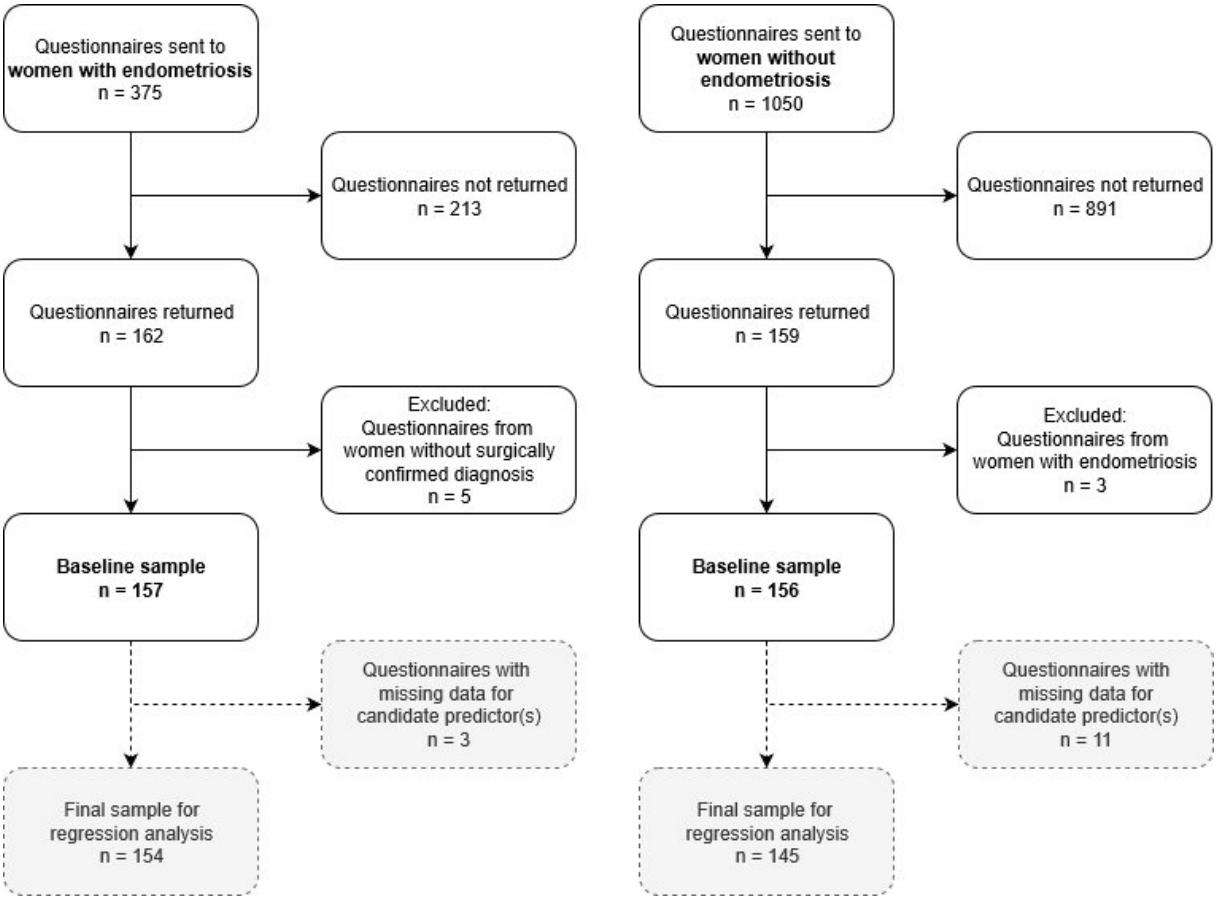
Nina Julie Verket http://orcid.org/0000-0002-2337-9653

REFERENCES

1 Fuldeore MJ, Soliman AM. Prevalence and symptomatic burden of diagnosed endometriosis in the United States: national estimates from a cross-sectional survey of 59,411 women. *Gynecol Obstet Invest* 2017;82:453–61.
2 Zondervan KT, Becker CM, Koga K, *et al*. Endometriosis. *Nat Rev Dis Primers* 2018;4.
3 Vercellini P, Viganò P, Somigliana E, *et al*. Endometriosis: pathogenesis and treatment. *Nat Rev Endocrinol* 2014;10:261–75.
4 Giudice LC. Clinical practice. endometriosis. *N Engl J Med* 2010;362:2389–98.
5 Nnoaham KE, Hummelshoj L, Webster P, *et al*. Impact of endometriosis on quality of life and work productivity: a multicenter study across ten countries. *Fertil Steril* 2011;96:366–73.
6 De Graaff AA, D'Hooghe TM, Dunselman GAJ, *et al*. The significant effect of endometriosis on physical, mental and social wellbeing: results from an international cross-sectional survey. *Hum Reprod* 2013;28:2677–85.
7 Nisenblat V, Prentice L, Bossuyt PMM, *et al*. Combination of the non-invasive tests for the diagnosis of endometriosis. *Cochrane Database Syst Rev* 2016;7.
8 Husby GK, Haugen RS, Moen MH. Diagnostic delay in women with pain and endometriosis. *Acta Obstet Gynecol Scand* 2003;82:649–53.
9 Matsuzaki S, Canis M, Pouly J-L, *et al*. Relationship between delay of surgical diagnosis and severity of disease in patients with symptomatic deep infiltrating endometriosis. *Fertil Steril* 2006;86:1314–6.
10 Hudelist G, Fritzer N, Thomas A, *et al*. Diagnostic delay for endometriosis in Austria and Germany: causes and possible consequences. *Hum Reprod* 2012;27:3412–6.
11 Staal AHJ, van der Zanden M, Nap AW. Diagnostic delay of endometriosis in the Netherlands. *Gynecol Obstet Invest* 2016;81:321–4.
12 Calhaz-Jorge C, Mol BW, Nunes J, *et al*. Clinical predictive factors for endometriosis in a Portuguese infertile population. *Hum Reprod* 2004;19:2126–31.
13 Nnoaham KE, Hummelshoj L, Kennedy SH, *et al*. Developing symptom-based predictive models of endometriosis as a clinical screening tool: results from a multicenter study. *Fertil Steril* 2012;98:692–701.
14 Surrey E, Carter CM, Soliman AM, *et al*. Patient-Completed or symptom-based screening tools for endometriosis: a scoping review. *Arch Gynecol Obstet* 2017;296:153–65.
15 Bakken A. *Ungdata. Nasjonale resultater 2018, nova Rapport 8/18*. Oslo: Norwegian Social Research (NOVA), 2018.
16 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
17 Verket NJ, Uhlig T, Sandvik L, *et al*. Health-Related quality of life in women with endometriosis, compared with the general population

and women with rheumatoid arthritis. *Acta Obstet Gynecol Scand* 2018;97:1339–48.
18 Domingo L, Hofvind S, Hubbard RA, *et al*. Cross-National comparison of screening mammography accuracy measures in U.S., Norway, and Spain. *Eur Radiol* 2016;26:2520–8.
19 Morton LM, Cahill J, Hartge P. Reporting participation in epidemiologic studies: a survey of practice. *Am J Epidemiol* 2006;163:197–203.
20 Suvitie PA, Hallamaa MK, Matomäki JM, *et al*. Prevalence of pain symptoms suggestive of endometriosis among Finnish adolescent girls (TEENMAPS study). *J Pediatr Adolesc Gynecol* 2016;29:97–103.
21 Chapron C, Barakat H, Fritel X, *et al*. Presurgical diagnosis of posterior deep infiltrating endometriosis based on a standardized questionnaire. *Hum Reprod* 2005;20:507–13.
22 Ballard K, Lane H, Hudelist G, *et al*. Can specific pain symptoms help in the diagnosis of endometriosis? a cohort study of women with chronic pelvic pain. *Fertil Steril* 2010;94:20–7.
23 Lafay Pillet MC, Huchon C, Santulli P, *et al*. A clinical score can predict associated deep infiltrating endometriosis before surgery for an endometrioma. *Hum Reprod* 2014;29:1666–76.
24 Treloar SA, O'Connor DT, O'Connor VM, *et al*. Genetic influences on endometriosis in an Australian twin sample. sueT@qimr.edu.au. *Fertil Steril* 1999;71:701–10.
25 Saha R, Pettersson HJ, Svedberg P, *et al*. Heritability of endometriosis. *Fertil Steril* 2015;104:947–52.
26 Chapron C, Lafay-Pillet M-C, Monceau E, *et al*. Questioning patients about their adolescent history can identify markers associated with deep infiltrating endometriosis. *Fertil Steril* 2011;95:877–81.
27 Sapkota Y, Attia J, Gordon SD, *et al*. Genetic burden associated with varying degrees of disease severity in endometriosis. *Mol Hum Reprod* 2015;21:594–602.
28 Koninckx PR, Ussia A, Adamyan L, *et al*. Deep endometriosis: definition, diagnosis, and treatment. *Fertil Steril* 2012;98:564–71.
29 Moen MH, Schei B. Epidemiology of endometriosis in a Norwegian County. *Acta Obstet Gynecol Scand* 1997;76:559–62.
30 Abbas S, Ihle P, Köster I, *et al*. Prevalence and incidence of diagnosed endometriosis and risk of endometriosis in patients with endometriosis-related symptoms: findings from a statutory health insurance-based cohort in Germany. *Eur J Obstet Gynecol Reprod Biol* 2012;160:79–83.
31 von Theobald P, Cottenet J, Iacobelli S, *et al*. Epidemiology of endometriosis in France: a large, nation-wide study based on hospital discharge data. *Biomed Res Int* 2016;2016:1–6.
32 Eisenberg VH, Weil C, Chodick G, *et al*. Epidemiology of endometriosis: a large population-based database study from a healthcare provider with 2 million members. *BJOG* 2018;125:55–62.

PAPER 3 – Supplementary flow chart



PAPER 3 – Supplementary table

Supplementary table: Logistic and lasso regression analyses of candidate predictors of endometriosis among observation with complete data for the candidate predictors, who only responded "Yes" or "No" to the candidate predictor family history of endometriosis" (142 cases and 130 controls)

	Univariable logistic regression		Multivariable logistic regression		Logistic regression with backward stepwise selection [c]		Lasso regression	
Candidate predictors	**B**	**OR (95%CI)**	**B**	**OR (95% CI)**	**B**	**OR (95% CI)**	**B**	**(95% CI)**
Intercept			-2.5	0.1 (0.0, 1.4)	-1.5	0.2 (0.1, 0.3)	-1.5	(-4.1, -0.5)
Age at menarche (years)	-0.2	0.9 (0.8, 1.0)	0.1	1.1 (0.9, 1.3)				
Severe dysmenorrhea [a] (cont.)	0.8	2.2 (1.8, 2.8)	0.2	1.2 (0.8, 1.8)			0.1	(0.0, 0.5)
Absenteeism from school [b] (cont.)	1.1	2.9 (2.2, 3.8)	0.9	2.4 (1.6, 3.6)	1.1	3.0 (2.2, 4.0)	0.8	(0.5, 1.2)
Use of painkillers [b] (ref. never/rarely)								
Sometimes	0.8	2.3 (1.2, 4.2)	-0.1	0.9 (0.4, 2.0)				
Often/Always	2.3	10.5 (5.4, 20.3)	0.4	1.5 (0.5, 4.2)			0.4	(0.0, 1.1)
Use of oral contraceptives [b]	1.5	4.5 (2.4, 8.4)	-0.1	0.9 (0.4, 2.2)				
Family history of endometriosis	2.2	8.7 (3.5, 21.2)	2.3	9.5 (3.5, 26.1)	2.3	9.6 (3.6, 26.0)	1.8	(1.0, 3.1)

OR: Odds ratio. CI: Confidence interval based on 1000 bootstrap samples. cont.: Continuous. [a] Experienced in adolescence. [b] Due to dysmenorrhea in adolescence. [c] Backward stepwise variable selection was performed using Wald test statistics $p \leq 0.157$ as the criterion for inclusion.