A systematic review of Patient-Reported Outcome Measures used to evaluate pelvic organ prolapse severity

by

Sarah Jane Conrad

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Rehabilitation Science

Faculty of Rehabilitation Medicine

University of Alberta

Abstract

Introduction and Hypothesis: In measuring intervention impact, patient-reported outcome measures (PROMs) are recommended as they aim to capture what is most meaningful to patients. To accurately reflect the impact of pelvic organ prolapse (POP) from the patient's perspective, PROMs must have strong measurement properties such as validity, reliability, and responsiveness. The aim was to conduct a systematic review of studies reporting on the measurement properties of PROMs used in both surgical and conservative management settings for evaluating POP.

Methods: Medline, EMBASE, CINAHL, Scopus, and Cochrane databases were searched from inception to May 2022 for studies that evaluated female adults with a diagnosis of or seeking treatment for POP. Studies were eligible if they involved 1) at least one group of female adults diagnosed with or presenting with symptoms of POP; 2) a self-reported outcome measure (PROMs, questionnaires) to evaluate POP-related domains; and 3) at least one measurement property including PROM development, content validity, internal consistency, reliability, measurement error, concurrent validity, discriminative validity, and/or responsiveness. Methodological quality was assessed using the COSMIN risk of bias (ROB) checklist, and measurement quality was determined using the COSMIN criteria for good measurement properties.

Results: 2,495 abstracts were screened, and 68 full-text articles were reviewed, from which 12 studies met the selection criteria and were included in this review. The most frequently evaluated measurement properties were internal consistency (6 PROMs, five studies) and responsiveness

(11 PROMs, ten studies). Most PROMs received a *sufficient* rating for the measurement quality of responsiveness in a surgical setting. Only five studies evaluated PROM responsiveness in the conservative management setting, and while most showed sufficient evidence of responsiveness; due to small number of studies and small sample sizes, confidence in the reported quality of this measurement property is low for conservative settings.

Conclusion: This original work identified a gap in evidence regarding the measurement qualities of identified PROMs used in the POP population. Few PROMs have empirical evidence supporting their content validity and responsiveness for evaluating the effectiveness of conservative interventions. Further research is needed to assess the full spectrum of measurement properties identified by COSMIN when considering existing PROMs used for those with POP.

Keywords: Pelvic Organ Prolapse, POP, Patient-Reported Outcome Measures (PROMs), COSMIN

Preface

This thesis is a manuscript-based thesis and is original work completed by Sarah J. Conrad. Chapter One includes the background information for the thesis. Chapter Two contains the full systematic review, which will be submitted for publication in a peer-reviewed journal. The results of Chapter Two have been presented at the Canadian Society for Pelvic Medicine's Annual General Meeting (Canmore, 2023), the Canadian Physical Therapy Congress (Quebec City, 2023) and at the International Continence Society's Annual General Meeting (Toronto, 2023). The authors are Sarah J. Conrad (University of Alberta), Dr. Stéphanie Bernard (Université Laval), Dr. Douglas P. Gross (University of Alberta), and Dr. Linda McLean (University of Ottawa).

Dedication

This thesis is dedicated to my family, who supported this journey and believed in me when I needed it the most.

To my husband and best friend Sam – thank you for loving a challenge (me) and choosing a path not yet taken. I have loved and appreciated every moment.

To my boys- Nicolas, Oliver, and Benjamin, thank you for pushing yourself to be the best you can be and inspiring me to do the same.

Acknowledgement

First, I would like to acknowledge the unwavering support from my advisors, Dr. Douglas Gross, and Dr. Linda McLean. Doug, I truly would not have had this opportunity without your belief in me and this project. You have taught me to me be bold in the unknown. I have appreciated our discussions and gained so many new perspectives from the moment I arrived at your office, thank you. Linda, the opportunity to be your student has been extraordinary, and I still can't believe it. Thank you for welcoming me into your incredible research lab and team at the University of Ottawa. I appreciate you sharing your wisdom, your expertise, and challenging me to think in new ways.

This project was truly elevated by the involvement of Dr. Stéphanie Bernard. Her patience, guidance and attention to detail were above and beyond in making this project the best it could be. I will be forever grateful for your involvement and support.

Thank you to the lab members of Mclean Functional Movement (MFM) and the lab students of Dr. Douglas Gross. Your motivation and passion for your research was contagious and gave me the energy to keep going when I needed it most.

I am grateful to my fantastic team at Royal Alexandra Hospital/Lois Hole Hospital for Women's Urogynecology clinic for the tremendous encouragement and support to complete this work while continuing in my clinical position.

To my parents, in-laws, and siblings, thank you for encouraging every step and giving me the space to make this dream into a reality. And cheers to friendships and my physio-girls who make me laugh and remind me of who I am at the end of the day. Lastly, I would like to thank my patients who inspired me to pursue this research project.

This thesis was made possible through the assistance of the Alberta Health Services Learning and Professional Development Fund, Women's Auxiliary Employee Education Assistance, University of Alberta Graduate Student Association, Faculty of Graduate Studies and Research Awards, and the Canadian Physiotherapy Association's Women's Health Division Jodi Boucher Advanced Practice Bursary.

Table of Contents

List of Tables

List of Figures

Chapter One- Introduction

Clinical Vignette

A 52-year-old female waits nervously in the changing room of her fitness facility. Three weeks previous, she had attended a fitness class with her friends and daughter. She noticed an unexpected bulging sensation in her vagina during the fitness class. After the class, she went to the washroom and saw a protruding tissue sitting at the opening of her vagina- she was horrified. Horrible thoughts had gone through her mind. "Is that my bladder? Is that a cancerous mass? After the class, she returned home and took it easy over the weekend. Instead of playing a co-ed volleyball game with her husband, she said she was not feeling well and stayed home. After looking up these symptoms online, sadness, frustration, and despair filled her thoughts when she looked at images of pelvic organ prolapse.

Today, she is sitting in the changeroom contemplating speaking to her friends about this experience. So many thoughts have gone through her mind: what happened to cause this? Will it come back? Will her husband be able to feel this or notice this bulge during sex? Thinking over these thoughts, she feels worse and does not want to bother her friends with these concerns. She decides not to stay for the fitness class and tells her friends she is not feeling well. She returns home and books an appointment with her physician.

After waiting for six months to see a specialist, she now waits in the examination room. A urogynecologist has given her a diagnosis of pelvic organ prolapse. Specifically, she has a stage two uterine prolapse. Her treatment options are surgery, pelvic floor muscle training, and a medical device called a pessary. Her impression from the urogynecologist is the severity of her prolapse is mild to moderate, but she feels her severity is much higher. These past six months she has stopped attending fitness classes with her friends, quit her volleyball team, and has been avoiding her husband. She feels like she has become a different person.

What is Pelvic Organ Prolapse?

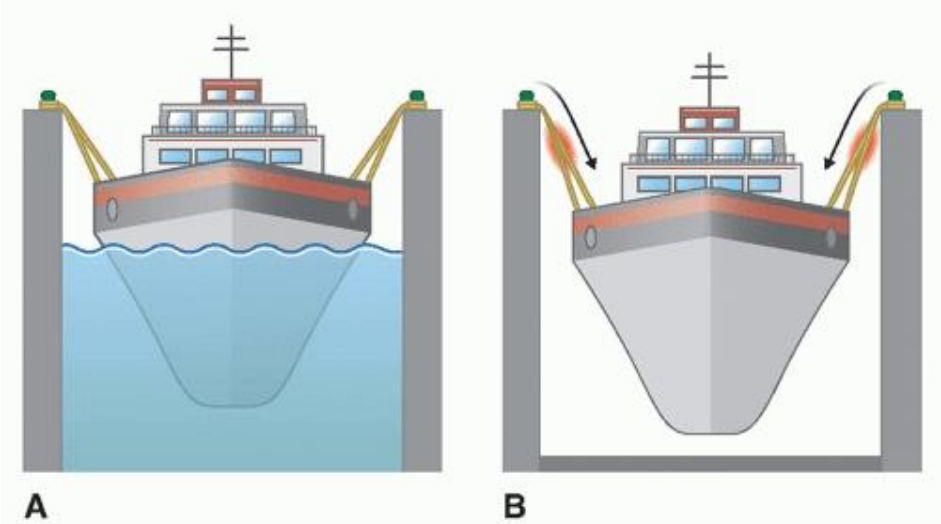
Pelvic Organ Prolapse (POP) is defined in the joint report on terminology by the International Urogynecology Association (IUGA) and the International Continence Society (ICS) as the downward displacement or herniation of the pelvic organs beyond the vaginal walls[1]. The most prominent symptom of POP is a vaginal bulge or feeling of heaviness, often with a combination of bladder and bowel dysfunction, including retention and incontinence[1].

This disorder significantly impacts patients, particularly those who participate in lifestyles or occupations with physically demanding tasks. Although POP is unlikely to result in mortality, there is urgency in addressing the dramatic disruption and adverse effect of POP symptoms on the performance of daily living, working activities, and quality of life[2].

Pathophysiology of Pelvic Organ Prolapse

The pathophysiology of POP is primarily related to weakness of the pelvic floor fascia. The main component of pelvic floor fascia and the ligaments is collagen. When our bodies perform tasks, contraction of the pelvic floor muscles places tension on the pelvic floor fascia. This tension pulls on the fascia in three directions, adding additional strength to support the pelvic organs[3]. To better understand this complex relationship, Dr. Norton[3] introduced the analogy of the "boat theory." Imagine a boat docking in a marina *(see Figure 1)*. The **boat** represents the **pelvic organs**, the **water** represents the **pelvic floor muscles** supporting the boat, and the **ropes** stabilizing the boat are the **pelvic ligaments**. If the boat was moved to a dry dock and the water had been drained there would be increased tension on the ropes (ligaments), and if a storm (injury, age-related change) occurs, more tension will be placed on the ropes (ligaments) and will cause the boat to sink (prolapse)[3].

Figure 1 POP Analogy



"Boat in dry dock" analogy for support of the cervix and vaginal apex. The water (analogous to the levator ani muscle complex) supports the boat. The lines (analogous to the uterosacral and cardinal ligaments) maintain the position of the boat. If the water is removed, the lines cannot support the weight of the boat. Thus, a loss of pelvic muscle function may inevitably lead to failure of the ligaments. (From Lammers, K., Prokop, M., Vierhout, M.E. et al. A pictorial overview of pubovisceral muscle avulsions on pelvic floor magnetic resonance imaging. Insights Imaging, 431–441 (2013). https://doi.org/10.1007/s13244-013-0261-9

Many risk factors are associated with the disruption of the fascia and the pelvic floor muscles, which impede their capacity to withstand stresses. However, the relationship between these risk factors and prolapse is not entirely understood. The known risk factors affecting collagen weakness are: menopause, parity, age-related changes, injury or trauma (vaginal delivery), obesity, mechanical loading (lifting), straining (constipation) and genetic factors[4].

Prevalence

Regrettably, the definition of POP from the international associations IUGA & ICS is too broad, recently Brown et.al, stated the lack of a commonly accepted POP definition limits the ability to provide adequate assessment[5]. Thus, the prevalence, incidence, and pathology of POP may be poorly understood. In a narrative report on the epidemiology of POP, Brown et al., report that if the prevalence of POP was based on *symptoms,* it affects 1% to 31% of females. If based on *pelvic examination (anatomical),* it affects 10% to 50%, and if *both symptoms* and *anatomical* impairments are considered, it affects 20% to 65% of the female population.

Measurement of POP

The evaluation of POP is complex. In the past, only anatomical markers were considered; however, anatomical measurements do not always correlate to patient experiences. Capturing patient experience through questionnaires called Patient-Reported Outcome Measurements (PROMs) has become increasingly important for accurately measuring the burden and impact of POP and providing a patient-centred approach to care[6].

Anatomical Measurement of POP - Measurement of Organ Displacement

For anatomical assessment of prolapse, one of the most commonly recommended tool is the Pelvic Organ Prolapse Quantification (**POP-Q**) scale. The POP-Q provides specific measurements of pelvic organ descent along the vaginal walls. This tool was developed in 2002 by a subcommittee of the ICS[7]. It offers a unique way to objectively measure, record and communicate anatomical findings regarding pelvic support, especially before or after an intervention.

The POP-Q has been validated to be reproducible in describing pelvic organ position[7-10]. Although a practical clinical tool, it has some limitations. The tool can be difficult to learn how to use, patient position affects reproducibility, and the tool is unable to identify asymmetrical or unilateral defects[7] as could be done on ultrasound or through Magnetic Resonance Imaging (MRI) where an overview of the entire pelvic floor is available[11].

Symptomatic Measurement - Patient-Reported Outcome Measures (PROMs)

PROMs provide important information beyond the anatomical measurements derived from the POP-Q. These tools measure **constructs** from the patient's experience. A construct is a broad topic or phenomenon that a tool is designed to measure, and it is not necessarily observable. For instance, measuring self-esteem or quality of life. Examples of constructs for POP include vaginal bulge symptoms, impaired physical function, and body image perception[12].

PROMs also provide a platform for patients to describe their symptoms, encouraging a way to disclose embarrassing information to their healthcare providers and capture what is most meaningful to the patient[6].

PROM use in Pelvic Health/ Urogynecology

Due to the discrepancy between anatomic and symptomatic presentation of POP, there has been increased recognition in the field of urogynecology to measure prolapse impact (i.e. symptom severity on the quality of life) from the patient's perspective[13]. This recognition led to the development of numerous PROMs for use in clinical and/or research settings[14]. However, not all PROMs in use have demonstrated robust measurement properties for use in the POP population[15], particularly when used to measure the impact of conservative management interventions. This is important as conservative management is the recommended first-line intervention for POP[16].

The strength of the measurement properties for a PROM is integral to ensuring the interpretation of results is accurate. In other words, the more robust the measurement properties, the more confidence one can have in adequately measuring the issue and in the observed results following the intervention.

Measurement Properties of PROMs

In deciding which PROM to use, researchers/clinicians should consider the following questions:

- Does the PROM measure what it is meant to be measuring? (validity)
- Will the PROM measure the same way each time it is used? (reliability)
- Does the PROM detect change over time in the construct? (responsiveness)

Consensus-based Standards for the selection of health Measurement Instruments (COSMIN)

To improve the selection of PROMs used in healthcare settings, an international multidisciplinary team named COSMIN was founded in 2005. This team of researchers with backgrounds in epidemiology, psychometrics, medicine, and qualitative research composed and developed steps for consideration when designing or when choosing a measurement tool[17-19].

The COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) manual was published in 2018.

Problem Statement

Clinicians/researchers need measurement tools that demonstrate robust measurement properties, including evidence of validity, reliability, and responsiveness to measure constructs related to POP, to evaluate the effectiveness of interventions and to enable advancements in patient-centred care. The current state of evidence for the measurement properties of PROMs used in the management of POP is largely unknown when considering both surgical and conservative management settings.

Thesis Objective

The aims of this research are to *identify* which PROMs have been evaluated through research on the POP population, to *assess* the methodological quality of studies evaluating the measurement properties of these PROMs, and to evaluate the *quality of evidence* for the measurement properties identified in the selected studies. A secondary aim is to determine which PROMs have been studied and evaluated to detect change after intervention in conservative management settings.

To address these aims, a systematic review following the recommended rigorous methodology recommended by COSMIN for systematic reviews of Patient-Reported Outcome Measures (PROMs) was completed[18].

Research Questions

Primary Research Question: What are the measurement properties of PROMs used for adult women with a diagnosis of POP?

Secondary Research Question: What is the responsiveness of PROMs used in adult women with a diagnosis of POP undergoing conservative management?

Chapter Two: Patient-Reported Outcome Measures for Pelvic Organ Prolapse: A Systematic Review using the COnsensus based Standards for health Measurement INstruments (COSMIN).

Introduction

Pelvic Organ Prolapse (POP) is a prevalent condition that, although not life-threatening, affects an individual's quality of life, physical function, sexual function, body image and social function[12]. The prevalence of POP in the literature is widely variable. It has been estimated that approximately 40-50% of women older than 40 years will present with *anatomical* prolapse, and one in ten women will report being *symptomatic*[20]. The most prominent symptom of POP is a vaginal bulge or heaviness, often with a combination of bladder and bowel dysfunction, including retention and incontinence. This disorder dramatically impacts women, particularly those who participate in lifestyles or occupations with physically demanding tasks. In measuring the impact of a health condition or intervention effectiveness, patient-reported outcome measures (PROMs) are recommended as they aim to capture what is most meaningful to patients. However, PROMs must have strong measurement properties such as validity, reliability, and responsiveness to accurately reflect the impact of this condition as perceived by the patient.

Previous reviews[6, 15] have summarized available PROMs and provided a synthesis of the evidence found regarding their utility in assessing pelvic organ prolapse symptoms. However, the advantage of using the COnsensus based Standards for health Measurement INstruments (COSMIN) guidelines over traditional evidence-based reviews is that the focus is not only on the reporting of data provided by those studies, but also on the *methodological quality* of the studies reporting on the measurement properties of the PROMs. Furthermore, the results of the measurement properties in the selected studies are then compared and rated against the criteria for good measurement properties (See Appendix 4). Each result is rated as either having a sufficient (+), insufficient (–), or indeterminate (?) level of evidence for good *measurement quality* according to this table. This ultimately provides a synthesis of the level of evidence for different PROMS in terms of their measurement quality according to COSMIN. Thus, it enables a more profound reporting and interpretation of the evaluated investigation results and the selected PROM[21]. The aim of this thesis was to conduct the first systematic review of studies

reporting on the psychometric properties of PROMs used in the evaluation of pelvic organ prolapse following the COSMIN guidelines.

Materials and Methods

This systematic review was registered on the PROSPERO website (reference number CRD42022309335). COSMIN guidelines[17] were followed based on details found on the COSMIN website (https://www.cosmin.nl).

Search Strategy

An electronic search was completed with the assistance of a research librarian (Elizabeth Dennett) from the University of Alberta. The five databases Medline, EMBASE, CINAHL, Scopus, and Cochrane were searched from inception to May 2022. The search strategy included controlled vocabulary and free-form text related to three concepts: *Pelvic Organ Prolapse*, *Survey and Questionnaires/PROMs, Validation Studies/Psychometrics*, and used the search terms and filters recommended by COSMIN[22]. (*See Appendix 2*)

Search Selection Criteria

Studies were eligible if they involved 1) **at least one group or subgroup of female adults diagnosed with or presenting with symptoms of POP**; 2) **a self-reported outcome measure (PROMs, questionnaires) to evaluate POP-related domains**; and 3) **measurement of at least one measurement property including *PROM development*, *content validity*, *internal consistency*, *reliability*, *measurement error*, *concurrent validity*, *discriminative validity*, and/or *responsiveness***. Studies were excluded if their analysis included patients with mixed conditions (i.e., patients with urinary incontinence and POP together) and/or if the PROM was not used in its original form. In addition, as no gold standard exists for PROMs, studies on criterion validity were excluded from this review, as well as cross-cultural validity. Studies involving cross-cultural validity were deliberately excluded because the inclusion of the many studies assessing this property would have made the current review unwieldy. The research team agreed this measurement property would be better served if examined in isolation as a separate systematic review. *(See Figure 2, PRISMA Flow Chart)*

Two researchers (S.C., S.B.) worked independently to screen titles and abstracts for inclusion, and then independently screened the full-text using the Covidence software. Disagreements in the selection process were resolved through discussion, while any remaining conflicts were resolved by discussion with a third reviewer (D.G. or L.M.).

Data Extraction

The following information was extracted from each included paper: author and year of the study, study population, number of patients, patient demographic information, PROMs used, constructs measured, the content of the PROM and the number of items/domains, and psychometric information. *(See Table 1).* In addition, relevant information for completing the COSMIN checklist was extracted. (*See Appendix 3*)

Data Analysis

Data for each study were categorized according to the measurement property boxes from the COSMIN checklist in this recommended order: *Content Validity*, *Structural Validity*, *Internal Consistency*, *Reliability*, *Measurement Error*, *Hypothesis Testing for construct validity* (convergent and discriminative), and *Responsiveness*. For further definitions of terms of the COSMIN taxonomy, please refer to the COSMIN website: https://www.cosmin.nl/tools/cosmin-taxonomy-measurement-properties/ or Appendix 1.

Data analysis occurred in *two* steps. The COSMIN Risk of Bias checklist was used to complete the first step. The first step aimed to assess the methodological quality of the studies. COSMN recommends the above hierarchy when reviewing PROMs, beginning with content validity. The evaluation of content validity has a separate manual[18] and was referenced for this section of the analysis. The content validity manual has a checklist of standards and recommendations for PROM development (*Box 1)* followed by standards for content validity (*Box 2)*. These standards are elaborate, and the PROM is explored in terms of relevance, comprehensibility, and comprehensiveness from the perspective of both patients and professionals. (*See Appendix 3*)

The remaining measurement properties were assessed and compared to the standards outlined in the COSMIN Risk of Bias Checklist *(Box 3 to Box 10)*.

According to the standards met, *each study* was then given a rating of either **V** ″very good," **A** "Adequate," **D "**Doubtful," or **I** "Inadequate". An **overall score of each study** was determined by taking the lowest rating among the checklist criteria (i.e. 'the worst score counts'), reported as **V** ″very good," **A** "Adequate," **D "**Doubtful," or **I** "Inadequate".

The second step was to compare the measurement property results of each study to the COSMIN *Criteria of Good Measurement Properties* (*See Appendix 4*) to determine the *quality of evidence* for the measurement property. The two researchers (S.C., S.B.) independently assessed the methodological quality of the included studies using the COSMIN risk of bias (ROB) checklist, as well as independently compared the study results to the measurement quality indicated in the COSMIN "Criteria for Good Measurement Properties" table. Each researcher (S.C., S.B.) provided a quality rating on the level of evidence as per the checklist recommendations[17, 18, 22] (*See Appendix 4)* Disagreements were resolved through discussion, while any remaining conflicts were resolved by discussion with a third reviewer (D.G. or L.M.).

In summary, each of the included studies was given a **methodological quality rating**. This methodological quality rating provides information about the standards of the study's methodology*,* indicating if any *significant* methodological flaws were present that may lead to biased results or conclusions. The rating categories were **V,** ″very good," **A,** "Adequate," **D, "**Doubtful," or **I**, "Inadequate." Next, the study's results on **measurement quality** were given a rating of sufficient evidence **(+)**, indeterminate evidence **(?)**, or insufficient evidence **(-)**. These two steps were completed for all 12 studies selected for our review.

Of note, most PROMs retained in this review had only *a single* study of evidence. Therefore, a pooling of results for each PROM was not completed. This affected two elements of the review. Specifically, in evaluating the measurement property *content validity,* we did not complete the third step of the COSMIN methodology. This final step in assessing content validity is to provide a measurement *quality rating* of sufficient, insufficient, or indeterminate for the measurement

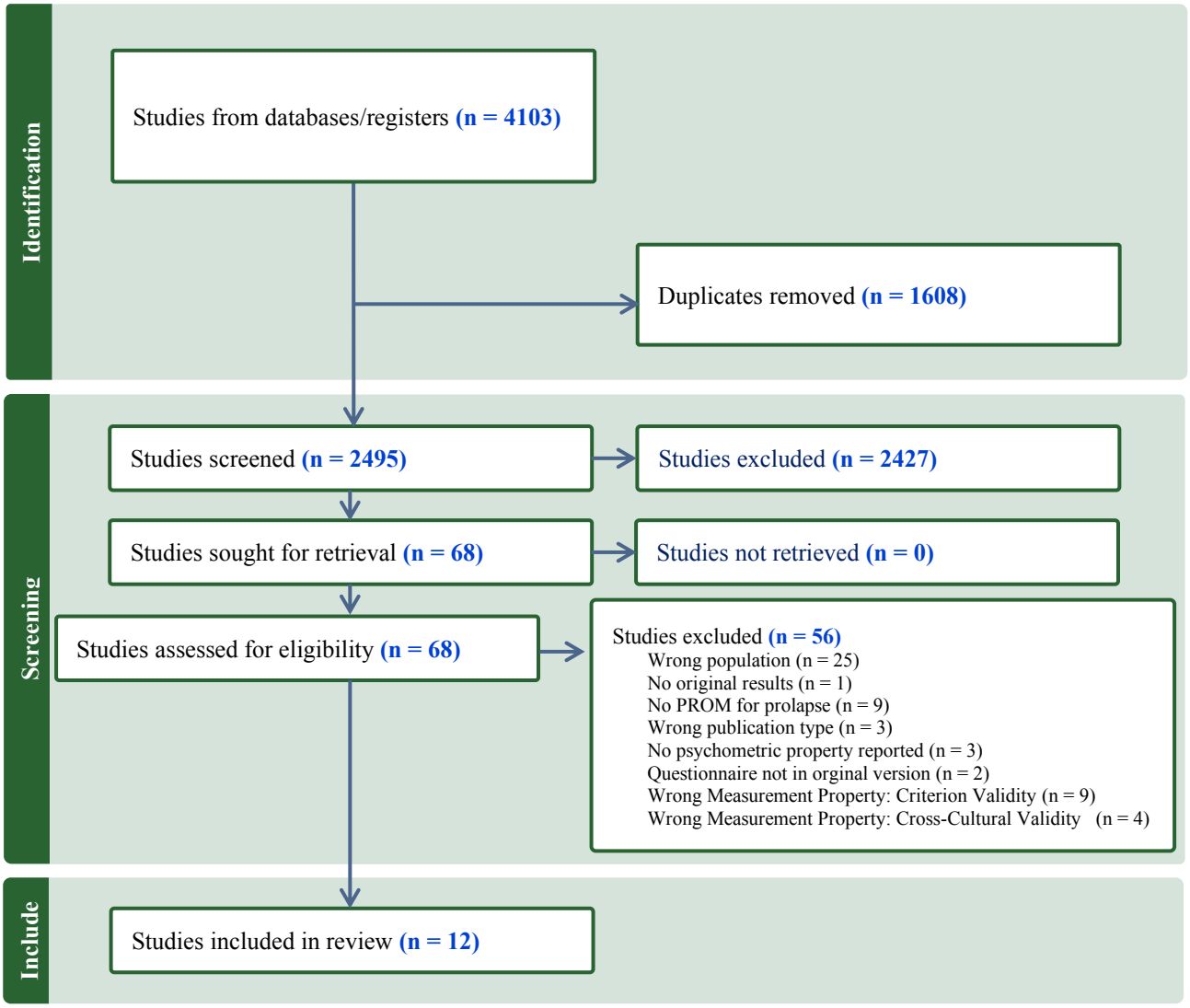
property. This measurement property is unique to content validity as it is to be completed by pooling the results on the relevance rating, the comprehensiveness rating, and the comprehensibility rating of the studies for each PROM. Therefore, only the *study's methodological quality* is reported in this section.

Furthermore, part four of the COSMIN methodology for systematic reviews[18] was not completed for this review. This section aims to provide an overall *G*rading of *R*ecommendations *A*ssessment, *D*evelopment, and *E*valuation (*GRADE*) for each PROM. This decision was based on the lack of adequate numbers of studies and small sample sizes included within the studies retained.

Results

There were 2,495 abstracts screened and 68 full-texts reviewed, from which 12 studies met the inclusion criteria and were included in this review. (*See Figure 2*)

Figure 2 PRISMA FLOW CHART



The 12 identified PROMs were: the Australian Pelvic Floor Questionnaire (***APFQ***), the Body Image in Pelvic Organ Prolapse Questionnaire ***(BIPOP)***, the electronic Personal Assessment Questionnaire-Pelvic Floor ***(ePAQ-PF)***, the International Consultation on Incontinence Questionnaire-Vaginal Symptoms ***(ICIQ-VS)*** module, Pelvic Floor Distress Inventory ***(PFDI)*** and its *short form* ***(PFDI-20)***, the Pelvic Floor Impact Questionnaire ***(PFIQ)*** and its *short form* ***(PFIQ-7)***, the Pelvic Organ Prolapse Symptom Score ***(POP-SS)***, the Pelvic Organ Prolapse-Urinary Incontinence ***(PISQ-IR)*** questionnaire, the Prolapse Quality of Life Questionnaire ***(P-QOL),*** and the Sheffield Prolapse Symptoms Questionnaire ***(SPS-Q)***. *(See Table 1)*

Measurement Properties Overview

The COSMIN measurement properties studied (n=number or articles) were as follows: *Content Validity* (n= 5), *Structural Validity* (n=2*)*, *Internal Consistency* (n=5), *Reliability* (n=4), *Hypothesis Testing for construct validity* (convergent and discriminative) (n=7), and *Responsiveness* (n=10). No studies reported on the measurement property of *Measurement Error.* Only two PROMs were evaluated by two separate research groups. They were the PFDI-20 & PFIQ-7[23, 24]. All other PROMs had only one study evaluating their measurement properties. The study by Price et al.,[25] on the ICIQ-VS provided the most detail and was evaluated for all seven measurement properties included in this review.

Table 1 Overview of Included Studies, PROMs, and Construct Measured.

First Author (year)	Patient Demographics	n	PROM	Construct	Content
Baessler **(2019)**[26]	Women who opted for pelvic floor surgery from urogynaecological clinics.	103	APFQ	4 scales (bladder, bowel, prolapse, and sexual function	42 items
Barber (2006)[27]	Two studies: 1) multicenter comparing two pessaries and 2) stage III or greater prolapse who underwent vaginal reconstructive surgery	42 (pessary) 64 (surgery)	PFDI	Symptom Severity: 3 scales symptom severity (Urinary, POP, Colorectal-Anal Distress Inventory)	46 items
	Same study		PFIQ	QOL- 3 scales (Urinary, POP, Colorectal- Anal Impact	93 items
Barber (2011)[23]	Four groups: two groups for POP surgery, a UI surgery group, and cohort group for FI	406	PFDI-20	Symptom Severity- 3 scales (Urinary, POP, Colorectal- Anal)	20 items
Gelhorn (2012)[24]	Two Studies: 1) Surgical POP repair (transvaginal placement of lightweight mesh system) 2) Surgical with trocar-guided repair system using a partially absorbable mesh)	1)148 2) 127	PFDI-20	(See above)	
Weigersma (2017)[28]	RCT Two Groups: 1) PFMT compared with watchful waiting in women with a prolapse above the hymen. 2) PFMT compared with pessary treatment in women with a prolapse above the hymen.	110 (PFMT) 39 (pessary)	PFDI-20	(See above)	
Barber (2011)[23]	Four groups: two groups for POP surgery, a UI surgery group, and cohort group for FI	406	PFIQ-7	QOL- 3 scales (Urinary, POP, Colorectal- Anal impact)	21 items
Gelhorn (2012)[24]	(Gelhorn above)		PFIQ-7	(See above)	

Bradshaw (2006)[29]	Study Group -referred for POP	203	SPS-Q	impact of uterovaginal POP on QOL.	25 items
Digesu (2005)[30]	Referred to urogynecology clinic for symptomatic POP	140	P-QOL	QOL- 9 scales (general health perceptions; prolapse impact; role; physical, social, and personal limitations; emotions; sleep/energy; and severity measures)	20 items
Hagen (2009)[31]	3 different settings: 1) New Zealand- participants of survey of postnatal women at 12 year follow up (ProLONG) 2) 2 outpatient gyne clinic- Scotland (new patients presenting with prolapse symptoms for PFMT) (POPPY) 3) Gyne Sx department- Scotland (women having any/post repair randomised to mesh or no mesh) (IMPRESS)	1) 435 2) 47 3) 66	POP-SS	Symptom Severity - presence and extent of key prolapse symptoms	7 items
Jones (2009)[32]	Women undergoing surgery for pelvic floor disorders	47	(ePAQ-PF)	Quality of Life- web- based questionnaire consisting of urinary, bowel, vaginal and sexual health	20 items
Lowder (2014)[33]	Women seeking care for POP	211	BIPOP	Sexual Function - identify the impact of POP on body image. Two versions one for sexually active and non-sexually population.	21 items
Price (2006)[25]	Patients from urogynecology clinic seeking care for varying degrees of POP	141	ICIQ-VS	Frequency, severity and impact of vaginal symptoms and related sexual matters	14 items

Pruijssera (2021)[34]	Multicentre prospective cohort study comparing the effect of pessary therapy vs surgery in women suffering from a symptomatic pelvic organ prolapse.	199 (pessary) 127 (surgery)	PISQ-IR	Sexual Function - (developed from the PISQ-12). Addresses sexual function specifically for women.	12 items

LEGEND:

APFQ Australian Pelvic Floor Questionnaire, **PFDI** Pelvic Floor Distress Inventory, **PFIQ** Pelvic Floor Impact Questionnaire, **PFDI-20** Pelvic Floor Distress Inventory Short Form, **PFIQ- 7** Pelvic Floor Impact Questionnaire short form, **SPS-Q** Sheffield Prolapse Symptoms Questionnaire, **P-QOL** Prolapse Quality of Life Questionnaire, **Sx** Surgery , **POP-SS** Pelvic Organ Prolapse Symptom Score, , **ePAQ-PF** electronic Personal Assessment Questionnaire- Pelvic Floor, **BIPOP** Body Image in the Pelvic Organ Prolapse Questionnaire, **ICIQ-VS** International Consultation on Incontinence Questionnaire- Vaginal Symptoms, **PISQ-IR** Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire- IUGA Revised

Measurement Properties by Category

PROM Development

Five of the included studies reported on PROM development: BIPOP[33], ICIQ-VS[25], POP-SS[31], P-QOL[30], SPS-Q[29]. The ROB was as follows: One study was rated *Adequate*[33], two were rated *Doubtful*[30, 31], and two were rated *Inadequate*[25, 29]. *(See Table 2)*

Content Validity

Content validity emphasizes how well the content of the PROM reflects the construct to be measured[17]. The evaluation of content validity is the most rigorous of all the measurement properties, it has a separate manual pertaining specifically to PROM development and content validity[18]. This measurement property was evaluated in four studies on four PROMs: BIPOP[33], ICIQ-VS[25], P-QOL[30], SPS-Q[29]. Three studies received ratings of *Doubtful*[25, 30, 33] and one study received *Inadequate*[29]. *(See Table 2)*

Table 2 PROM Development, Content Validity and Structural Validity

PROM	BIPOP	ICIQ-VS	POP-SS	P-QOL	SPS-Q
First Author	Lowder, 2014	Price, 2006	Hagen, 2009	Digesu,2005	Bradshaw, 2006
PROM Development					
Risk of Bias	A	I	D	D	I
Content Validity					
Risk of Bias	D	D		D	I
Structural Validity					
Risk of Bias	A	A			
Measurement Quality	+ (EFA completed, no CTT or IRT)	+ (EFA completed, no CTT or IRT)			

Legend

Methodological Quality Risk of Bias (ROB): V Very Good, A Adequate, D Doubtful, I Inadequate

Measurement Quality: sufficient evidence (+), indeterminate evidence (?), insufficient evidence (-)

EFA Exploratory factor analysis, CTT Classical Test Theory, IRT Item Response Theory

Black square represents not evaluated

Structural Validity

The standards for structural validity focus on how well the scores of the PROM adequately reflect the construct to be measured[19]. Most often this is completed by factor analysis. Both the Lowder et al., (BIPOP)[33] and the Price et al., (ICIQ-VS)[25] studies reported conducting exploratory analyses. As few details were given about the statistical methods used in both studies a third reviewer (D.G.) was consulted. The consensus was made to "give the benefit of the doubt" regarding statistical methods completed, and a ROB rating of *Adequate* was given to both studies. Additionally, the BIPOP and the ICIQ-VS PROMs had *sufficient* evidence for the measurement property structural validity. *(See Table 2)*

Internal Consistency

This measurement property describes the level of interrelatedness among the items in the PROM[19]. The studies on internal consistency demonstrated a low risk of bias among all five studies with a rating of *Very Good*. All five studies[24, 25, 30, 31, 33] used Cronbach's alpha as a statistical method. However, caution should be taken with interpretation of these results as there was limited reporting on the dimensionality of the scales. Cronbach's alpha is intended for unidimensional measures, that is, when the items in the scale (or subscale) are intended to measure a single construct[17]. In our analysis of scales and domains, it was not always clear how the scales were presented and whether the authors had completed factor analysis. Therefore, we reported on this measurement to the best of our ability with the evidence reported in each study. Only the BIPOP and ICIQ-VS had a measurement quality rating that was deemed *sufficient* for this property. *(See Table 3)*

Reliability

Reliability evaluates the extent to which the scores for patients who have not changed are consistent for repeated measurement under several conditions[17]. All four studies reporting on reliability[25, 29, 30, 33] for the PROMs: BIPOP, ICIQ-VS, P-QOL, and SPS-Q received a ROB rating of *Doubtful*. Only the BIPOP and ICIQ-VS PROMs had *sufficient evidence* for measurement quality of reliability as they used the recommend statistical methods. Test-retest reliability was the only type of reliability reported in the four studies. *(See Table 3)*

Table 3 Internal Consistency and Reliability

PROM	**BIPOP**	**ICIQ-VS**	**PFDI-20**	**PFIQ-7**	**POP-SS**	**P-QOL**	**SPS-Q**
First Author	Lowder 2014	Price 2006	Gelhorn 2012	Gelhorn 2012	Hagen 2009	Digesu 2005	Bradshaw 2006
Internal Consistency							
ROB	V	V	V	V	V	V	
Results (Cronbach's alpha)	0.93*	0.79-0.84*	Study 1 (0.68-0.85) * Study 2 (0.64-0.82) *	Study 1 (0.89-0.95) * Study 2 (0.90-0.96) *	0.823 * (control) 0.79-0.72* (PFMT) 0.81-0.82 * (Sx)	0.80*	
Quality	(+)	(+)	(?)	(?)	(?)	(?)	
Reliability							
Risk of Bias	D	D				D	D
Results	ICC: 0.77-0.80	Weighted Kappa: 0.58 -1.0				SCC: 0.64-0.87	Kappa: 0.55-1.0 Rep* 0.74-1.4
Quality	(+)	(+)				(-)	(?)

Legend

Methodological Quality Risk of Bias (ROB): **V** Very Good, **A** Adequate, **D** Doubtful, **I** Inadequate

Measurement Quality: sufficient evidence (+), indeterminate evidence (?), insufficient evidence (-)

* Caution of interpretation due to clarity of unidimensionality scale reporting, **PFMT** Pelvic Floor Muscle Training, **Sx** Surgery, **ICC** Intra-class Correlation Coefficient, **SCC** Spearman's rho nonparametric correlation coefficient, **Rep*** Repeatability coefficients

Black square represents not evaluated**.**

Hypotheses Testing for Construct Validity (convergent)

Hypotheses testing for construct validity refers to the degree to which the scores of the PROM are consistent with the hypotheses[17]. Hypotheses testing can occur by comparison to other outcome measures (convergent) or differences in scores between known groups (discriminative).

Four studies reported on convergent validity[24, 29, 30, 33]. The ROB was *Adequate* for Digesu[30] and Lowder[33], all others were rated *Doubtful.* The comparative instruments used to evaluate the category of **pelvic floor symptoms and severity were** the Pelvic Organ Prolapse Quantification (POP-Q) instrument, the Pelvic Floor Distress Inventory short form (PFDI-7), Kings Health Questionnaire (KH-Q), and the Birmingham Bowel and Urinary Symptoms Questionnaire (BBUSQ). For the category **impact on quality of life,** instruments such as the Pelvic floor

Impact Questionnaire short form (PFIQ-7), EuroQol (EQ-5D), and the SPS-Q and the SF36) were chosen. In the category of **impact on body image**, the compared instruments were Body Exposure during Sexual Activity Questionnaire, Body Image Quality of Life inventory, Pelvic Floor Distress Inventory-20, Pelvic Floor Impact Questionnaire -7, Patient Health Questionnaire-9, and the Self-esteem Scale. Lastly, for the category of **impact on sexual function,** the Pelvic Organ Prolapse/Urinary Incontinence Sexual Function (PISQ-12), and the Female Sexual Function Index (FSF-I) were used. *(See Table 4)* Studies used Spearman's rank correlation coefficient in their statistical methods. The BIPOP[33], PFDI-20[24], PFIQ-7[24], SPS-Q[29] and the P-QOL[30] had *sufficient evidence* for construct validity (convergent) measurement quality.

Hypotheses Testing for construct validity (discriminative validity)

Six studies reported on discriminative validity[24-26, 29-31]. The groups in all studies were mainly compared between a surgical cohort and a community cohort. Overall, the risk of bias for most of the studies was high, with most studies receiving a *Doubtful* rating for methodological quality. The PROMs: APFQ, ICIQ-VS, POP-SS, P-QOL, and the SPS-Q had *sufficient levels* of evidence for measurement quality. The PROM POP-SS[31] had the lowest risk of bias for study methodology with a rating of *Very Good* and *sufficient evidence* for construct validity(discriminative validity) quality. *(See Table 4)*

Table 4 Hypothesis Testing for Construct Validity

PROM	APFQ	BIPOP	ICIQ-VS	PFDI-20	PFIQ-7	POP-SS	P-QOL	SPS-Q
First Author, year	Baessler 2019	Lowder 2014	Price 2006	Gelhorn 2012	Gelhorn 2012	Hagen 2009	Digesu 2005	Bradshaw 2006
Construct Validity (convergent)								
ROB		A		D	D		A	D
Results r[A]		[B]0.77-0.79		[C] -0.31 to 0.46	[C]0.37-0.52		[E]0.11-0.59	[D]0.3-0.8
Quality		(+)		(+)	(+)		(+)	(+)
Construct Validity (discriminative)								
Risk of Bias	D		D	D	D	V	D	D
Results r [A]	-		-	[E]-0.11,0.20	[E] -0.08, -0.15	-	[E]0.11-0.59	-
Quality	(+)		(+)	(-)	(-)	(+)	(+)	(+)

Legend

Methodological Quality Risk of Bias (ROB): **V** Very Good, **A** Adequate, **D** Doubtful, **I** Inadequate

Measurement Quality: sufficient evidence (+), indeterminate evidence (?), insufficient evidence (-)

A Spearman's rank correlation coefficients

B Comparison Outcome Measures: *Body Exposure during Sexual Activity Questionnaire, Body Image Quality of Life inventory, Pelvic Floor Distress Inventory-20, Pelvic Floor Impact Questionnaire -7, Patient Health Questionnaire-9, and the Self-esteem Scale*

C Comparison Outcome Measures: *PISQ-12 Pelvic Organ Prolapse/Urinary Incontinence Sexual Function, EQ-5D EuroQol, PGIC Patient Global Impression of Change*

D Comparison Outcome Measures: *BBUSQ Birmingham Bowel and Urinary Symptoms Questionnaire, KHQ Kings Health Questionnaire, FSF-I Female Sexual Function Index, SF-36 Short From 36*

E Comparison Outcome Measures: ***POP-Q*** *Pelvic Organ Prolapse Quantification Scale*

Black square represents not evaluated**.**

Responsiveness

Responsiveness refers to the ability of a PROM to detect change over time in the construct to be measured[17]. There are four approaches to test hypotheses of change of score for responsiveness; they are: **(a)** criterion approach (comparison to a gold standard), **(b)** construct approach (comparison to other outcome instruments), **(c)** construct approach (comparison between subgroups), and **(d)** construct approach (pre-and post-intervention). For clarity this labelling was used in Table 5 and 6. The level of evidence for this measurement property refers to how well the result is in accordance with the hypothesis[17].

Responsiveness in Surgical Settings

Nine studies evaluated responsiveness in the surgical setting. Four studies demonstrated low risk of bias with a *Very Good* methodological quality rating for the construct approach (hypotheses testing pre- and post-intervention)[23, 24, 26, 32]. Two studies received a rating of *Doubtful*[27, 34], and three studies received a rating of *Inadequate*[25, 29, 31]. Only two studies evaluated responsiveness with the construct approach (comparison between subgroups). One study received a rating of *Very Good*[27] and the other received rating of *Inadequate*[31] for methodological quality. (*See Table 5*)

The PROMs with *sufficient evidence* for good measurement quality were the APFQ[26], PFDI[27], PFDI-20[23, 24], PFIQ-7[24], ePAQ-PF[32], POP-SS[31]. PROMs with *indeterminate* quality of evidence were the PFIQ-7[23], the ICIQ-VS[25], and the SPS-Q[29], and lastly, the PFIQ[27] had *insufficient* measurement quality evidence. *(See Table 5)*

Responsiveness in Conservative Management Settings

Only four studies[27, 28, 31, 34] evaluated PROM responsiveness in a conservative management setting (PROMs: PFDI, PFDI-20, PFIQ, POP-SS, PSIQ-IR). (*See Table 6*) The risk of bias varied across studies. The only study to examine responsiveness with the construct approach (comparison with another outcome measure) was Weigersma et al.,[28] who compared the outcome measures of two instruments, the POP-Q and the Global Perception of Improvement

Scale (GPI), with the PFDI-20. This study was rated *Doubtful* for methodological quality. Two studies used the construct approach (comparison between subgroups). One study had a low risk of bias with a *Very Good*[27] and the other was rated *Inadequate*[31]. Lastly, three studies evaluated responsiveness using the construct approach (comparison pre- and post-intervention). Two of these received a rating of *Doubtful*[27, 34] and one received a rating *Inadequate*[31] for methodological quality.

The PROMs: PFDI, PFDI-20, PFIQ, and POP-SS all had *sufficient evidence* for good measurement quality for responsiveness in conservative management setting (*See Table 6)*

Table 5 Responsiveness in the Surgical Setting

PROM	ICIQ-VS	APFQ	PFDI	PFDI-20	PFIQ	PFIQ -7	POP-SS	ePAQ-PF	SPS-Q	PSIQ-IR
First Author, year	Price, 2006	Baessler, 2018	Barber, 2006	Barber, 2011 Gelhorn, 2012	Barber, 2006	Barber, 2011 Gelhorn, 2012	Hagen, 2009	Jones, 2009	Bradshaw, 2006	Pruijssers, 2021
Surgical Setting										
ROB	d) **I**	d) **V**	c) **V**, d) **D**	d) **V**	c) **V** d) **D**	d) **V**	c) **I** d) **I**	d) **V**	d) **I**	d) **D**
Results	d)*	d) POP scale: ES: 2.2 SRM: 2.1	b) Sx vs Pessary group d)POPDI: ES:1.23 SRM:1.20	d)POPDI-6 SRM: -1.35 to-1.68	b) Sx vs Pessary group d)POPIQ: ES:0.56 SRM:0.64	d)POPIQ-7 SRM: -0.21 to -0.90	c)Surgery vs PFMT d) *	d)Vaginal scale: ES:07to1.0 SRM:0.6 to 1.0	d) ES:0.2 to 0.8	ES:0.6
Quality	**(?)**	**(+)**	**(+), (+)**	**(+)**	**(+), (-)**	**(?)**	**(+), (?)**	**(+)**	**(?)**	**(+)**
ROB				d)**V**		d)**V**				
Results				d) POPDI-6 ES: -1.49 to -1.58 SEM:9.12 to 6.33		d)POPIQ-7 ES: -0.71 to-0.82 SEM:4.40 to1.79				
Quality				(+)		(+)				

Legend

Methodological Quality Risk of Bias (ROB): **V** Very Good, **A** Adequate, **D** Doubtful, **I** Inadequate

Measurement Quality: sufficient evidence (+), indeterminate evidence (?), insufficient evidence (-)

POP Pelvic Organ Prolapse domain, **ES** Effect Size, **SRM** Standardized Response Mean, **Sx** Surgery, **POPDI** (long form) Pelvic Organ Prolapse Distress Inventory, **POPDI-6** (short form) Pelvic Organ Prolapse Distress Inventory, **POPIQ** (long form) Pelvic Organ Prolapse Impact Questionnaire, **POPIQ-7** (short form) Pelvic Organ Prolapse Impact Questionnaire, **SEM** Standard Error of Measurement ***** Statistical method not recommended by COSMIN, **PFMT** Pelvic Floor Muscle Training.

Black square represents not evaluated**.**

Table 6 Responsiveness in Conservative Management settings

PROM	**PFDI**	**PFDI-20**	**PFIQ**	**POP-SS**	**PSIQ-IR**
First Author, year	Barber, 2006	Weigersma, 2017	Barber, 2006	Hagen, 2009	Pruijssers, 2021
Conservative Management					
ROB	c) ***V***, d) ***D***	b) ***D***	c) ***V*** d) ***D***	c) **I** d) ***I***	d) ***D***
Results	c) Sx vs Pessary group d)POPDI: ES:0.68 SRM:0.69	b) POP-Q with PFDI-20 small correlation Spearman's (ρ)=0.001 b) GPI and the PFDI-20 moderate correlation Spearman's (ρ)=0.35	c) Sx vs Pessary group d)POPIQ: ES:0.34 SRM:0.30	c)Surgery vs PFMT d) *	ES:0.0
Quality	**(+), (+)**	**(+)**	**(+), (-)**	**(+), (?)**	**(-)**

Legend

Methodological Quality Risk of Bias (ROB): **V** Very Good, **A** Adequate, **D** Doubtful, **I** Inadequate

Measurement Quality: sufficient evidence (+), indeterminate evidence (?), insufficient evidence (-)

Sx Surgery, **POPDI** (long form) pelvic organ prolapse distress inventory, **ES** effect size, **SRM** Standardized Response Mean**, POP- Q** Prolapse Quantification Scale, **GPI** Global Patient Improvement Scale, **POPIQ** (long form) pelvic organ prolapse impact questionnaire, **PFMT** Pelvic Floor Muscle Training,

***** Statistical method not recommended by COSMIN

Discussion

To the best of the authors' knowledge, this is the first systematic review reporting on both the methodological quality of studies and the measurement property quality for PROMs used in the POP population. Our findings highlight a gap in evidence for PROMs with strong content validity, raising a serious concern about the validity of PROMs[35] currently used in the prolapse population. Content validity is considered the most critical measurement property of a PROM, and a lack of content validity may affect the quality of other measurement properties[18]. For instance, irrelevant items in the PROM would decrease its internal consistency and structural validity[17]. The COSMIN content validity standards focus on the relevance of a PROM, its comprehensiveness, and its comprehensibility from two perspectives: patients and healthcare professionals.

The evaluation for content validity includes the standards for PROM development and for content validity. (*See Appendix 3)* In this review, PROMs were developed primarily through authors completing a literature search, interviewing professionals, and speaking with patients. However, most studies[25, 29, 30] did not clearly explain how the patients or the healthcare professionals (physicians) were consulted. Interestingly, no studies consulted other relevant health disciplines (i.e., physical therapy, nursing) for input. Studies on the BIPOP[33] and the POP-SS[31] PROMs addressed most PROM design standards. They were the only two studies that recorded having received input from patients for the categories relating to the comprehensibility and comprehensiveness of the PROM.

Unique to the BIPOP PROM development process was the application of body image themes taken from women with POP who had participated in a previous study[33] Interestingly, the Lowder et al., authors referenced the work by Sung et al.,[12] which also offered supportive evidence for similar themes and meaningful outcomes identified by patients with POP. These outcomes are notable as they represent constructs distinctive to the prolapse population. In order of importance, these outcomes were listed as: 1) resolution of vaginal bulge symptoms, 2) improvement of physical function, 3) improvement in sexual function, 4) improvement of body image perception, and 5) improvement of social function.

Importantly, all studies in the *content validity* category *(See Table 2)* were given a rating of *Doubtful or Inadequate* predominately because they did not conform with the COSMIN standard of including at least two researchers in the analysis of relevance, comprehensiveness, and comprehensibility. Of note, the Lowder et al.,[33] study provided the most detail regarding relevance, comprehensibility, and comprehensiveness and would have scored much higher compared to the other studies if this standard had been met. The Bradshaw et al.,[29] study was rated *Inadequate* because the reported methods were not in accordance with the recommended standards as per the COSMIN checklist for content validity.

According to the COSMIN hierarchy, the next evaluation item is the internal structure of the PROM *(See Table 3)*. Internal structure includes the measurement properties of *structural validity* and *internal consistency*. Only the BIPOP[33] and ICIQ-VS[25] PROMs had *sufficient evidence* for internal consistency for measurement quality as both these PROMs had reported some evidence for structural validity (i.e. exploratory factor analysis) and a Cronbach's alpha greater than 0.70. Notably, the small sample sizes, the scarcity of studies on structural validity, and the considerations for the reporting on internal consistency (i.e.unidimensional scale) suggest that caution should be applied when considering the reported results on the internal structure of these PROMs[17]. Of note, a third reviewer was consulted on the evaluation of the Digesu et al., [30] paper due to an identified discrepancy between the definition of internal consistency not matching the statistical methods used, as well as inconsistency between the statistical tests described in the methods and the statistics reported in the results section of the paper. The team reached the consensus that this was a typographical error, and the decision was based on the definition of terms in the statistical methods section.

The remaining properties (*reliability*, hypothesis testing for *construct validity* and *responsiveness*) depend on the performance of the above measurement properties (content validity, structural validity, and internal consistency), primarily content validity. This is because the measurement properties of reliability, construct validity, and responsiveness mainly assess the quality of the *PROM scale* rather than the items (constructs) in the PROM[19]. As mentioned

previously, if constructs included in the PROM are irrelevant, this could decrease it's internal consistency and structural validity[18].

For *reliability*, the PROMs BIPOP[33] and the ICIQ-VS[25] used the intraclass correlation coefficient (ICC) and weighted kappa calculations as recommended by COSMIN, and provided satisfactory evidence for test-retest reliability. One critical assumption in reliability studies is the assumption that all patients were stable in terms of the construct being measured between the initial test and the retest[17]. However, no study provided evidence of such stability, which decreased the ratings when assessing the risk of bias. The lack of clarity surrounding the test and retest conditions also affected the risk of bias. It was often unclear if the retest was completed in the same setting (at the clinic) or followed up by mail. This is relevant as patients may answer differently when the environment changes[17]. The Digesu et al.,[30] paper required a third reviewer (D.G.) for consensus due to a discrepancy in reporting of statistical results as reliability was defined by referring to Cronbach's alpha, consensus was reached this was a typed error and was referring to internal consistency. The article later referred to reliability statistics using Spearman's rho nonparametric correlation coefficient, and this was the statistic evaluated in our review (*See Table 3)*

In this review, hypothesis testing for *construct validity* (discriminative) was most robust in the PROM POP-SS[31] as it had the lowest risk of bias and demonstrated *sufficient evidence* by comprehensively describing the characteristics between the comparison groups. The study also discussed the expected direction and magnitude of the correlation between the different study groups (PFMT and surgery). *(See Table 4)*

*Responsiveness i*s the measurement property most linked to clinical practice, as it indicates an ability to detect a change in the construct being measured over time. Importantly, although responsiveness is a separate measurement property from validity, the only difference between construct validity and responsiveness is that validity refers to the validity of a single score, while responsiveness refers to the validity of a change in that score[17], and so the COSMIN responsiveness standards were similar to those for hypotheses testing for construct validity.

In this review, nine studies evaluated the responsiveness of PROMs in the surgical setting, and although there *is high-level evidence supporting PFMT as a first-line treatment intervention for POP* in the general female population[16], only **four** studies evaluated responsiveness of PROMs in the conservative management setting.

The most common approach to testing the responsiveness of the PROMs was the construct approach, both through hypotheses testing pre- and post-intervention and through comparison with other outcome measurement instruments. Effect size was the most used statistical approach. The methodological quality of the studies assessing responsiveness varied across studies, with the high ROB ratings being primarily due to the incorrect or incomplete reporting of interventions, statistical analyses, or results. For the **surgical setting** the APFQ[26], ePAQ-PF[32], PFDI-20[23, 24], PFDI[27], PFIQ-7[24], PSIQ-IR[34] and the POP-SS[31] PROMs all had sufficient evidence *(See Table 5)*; and **for conservative management** the PFDI[27], PFDI-20[28], PFIQ[27], and the POP-SS[31] PROMs had sufficient measurement quality. *(See Table 6)*.

Although these PROMs demonstrated sufficient evidence, the low sample sizes and the small numbers of studies led us to conclude that caution should be considered when applying these results. Interestingly, only the POP-SS PROM had sufficient evidence of quality for the measurement properties of construct validity and responsiveness in both the surgical and conservative management settings. This finding emphasizes the need for more high-quality studies evaluating and examining PROMs used in conservative management settings.

Relevance to Future Research and Clinical Practice

These findings build upon the work and recommendations from the International Urogynecology Consultation in 2022 by Cichowksi et al.,[6] by completing the subsequent steps of assessing methodological quality of the selected studies using the COSMIN guidance ***and*** by including conservative management settings. All PROMs evaluated in this systematic review were referenced in this consultation. However, the recommendations of Cichowski et al., included patient-reported goals, and had a different inclusion criterion as they used broader search terms (*pelvic floor disorder/pelvic floor dysfunctions,* etc.). Whereas the search terms in the current

review were limited to *pelvic organ prolapse* and *self-reported outcome measures evaluating a POP-related domain.* Therefore, some of the PROMs (i.e.: *Patient Global Impression of Improvement (PGI-I), Patient global impression of change (PGI-C), Pelvic Floor Bother (PFBQ), Surgical Satisfaction Questionnaire (SSQ), Improvement Satisfaction Scale (ISS)*) were not retrieved for inclusion in this review.

Although the results of this systematic review did not allow us to provide a GRADE level recommendation for the selected PROMs, they can be used to inform PROM use and recommended next steps. For instance, the review highlighted the relevance and significance of the measurement property content validity. There is a need for current PROMs to either be updated and re-examined using current standards for recommendations (i.e., COSMIN) or perhaps a new PROM should be developed considering these standards.

Specific to the second research question, the findings suggest that the Pelvic Organ Prolapse-Symptom Score (POP-SS)[31] is the only PROM with sufficient evidence of adequate construct validity and responsiveness to be used in both surgical and conservative management settings.

This systematic review is timely; there is an emerging recommendation by the International Urogynecology Consultation commitee[5] that a validated PROM should be considered along with the pelvic examination of the POP-Q as the *Gold Standard* for evaluating POP. Our results support this recommendation and are a call to action for more high-quality studies to validate PROM use in the prolapse population.

Strengths and Limitations

A strength of this review was the use of the rigorous COSMIN recommended guidelines to produce robust findings on the quality and strength of the current PROMs used in the POP population. However, this may have also been a limitation. Many of the retained studies were published before the COSMIN standards and criteria were published in 2018. As such, the reporting on these measurement properties was held to a higher standard than was expected at the time of publication for many of these studies.

A further limitation is that we included studies that *reported results evaluating the PROMs on individuals with POP*. Notably, the condition of pelvic organ prolapse is complex and co-existing symptoms of bladder and bowel may also be present[1]. As such, the generalizability of the results is limited.

Lastly, only those studies which used the original form of a PROM were included. This meant that most studies were completed in English-speaking populations and were thus over-representative of white females from high-income countries. For instance, in the data extraction phase on *patient demographics,* it was notable that all 12 of the included studies had ethnicity (when recorded) of over 90% identifying as white women. Importantly, if this measurement property had been included, we may have had more ethnicity representation in our review.

Conclusion

The results from this review advocate for future research evaluating PROM measurement properties in the literature, especially for content validity. Additionally, the results emphasize how few studies have evaluated these measurement properties in conservative management settings for the pelvic organ prolapse population. Furthermore, our findings complement the recommendations recently proposed by the International Urogynecology Consultation group[6] on PROMs use in the evaluation of patients with pelvic organ prolapse.

References

1. Haylen BT, Maher CF, Barber MD, Camargo S, Dandolu V, Digesu A, et al. An International Urogynecological Association (IUGA) / International Continence Society (ICS) Joint Report on the Terminology for Female Pelvic Organ Prolapse (POP). Neurourology and Urodynamics. 2016;35(2):137-68.
2. Bordeianou LG, Anger J, Boutros M, Birnbaum E, Carmichael JC, Connell K, et al. Measuring pelvic floor disorder symptoms using patient-reported instruments: proceedings of the consensus meeting of the pelvic floor consortium of the American Society of Colon and Rectal Surgeons, the International Continence Society, the American Urogynecologic Society, and the Society of Urodynamics, Female Pelvic Medicine and Urogenital Reconstruction. Tech Coloproctol. 2020;24(1):5-22.
3. Norton PA. Pelvic Floor Disorders: The Role of Fascia and Ligaments. Clinical Obstetrics and Gynecology. 1993;36(4):926-38.
4. Weintraub AY, Glinter H, Marcus-Braun N. Narrative review of the epidemiology, diagnosis and pathophysiology of pelvic organ prolapse. Int Braz J Urol. 2020;46(1):5-14.
5. Brown HW, Hegde A, Huebner M, Neels H, Barnes HC, Marquini GV, et al. International urogynecology consultation chapter 1 committee 2: Epidemiology of pelvic organ prolapse: prevalence, incidence, natural history, and service needs. Int Urogynecol J. 2022;33(2):173-87.
6. Cichowski S, Grzybowska ME, Halder GE, Jansen S, Gold D, Espuna M, et al. International Urogynecology Consultation: Patient Reported Outcome Measures (PROs) use in the evaluation of patients with pelvic organ prolapse. Int Urogynecol J. 2022;33(10):2603-31.
7. Persu C CC, Cauni V, Gutue S, Geavlete P. Pelvic Organ Prolapse Quantification System (POP-Q) - a new era in pelvic prolapse staging. J Med Life. 2011(Jan-Mar):75-81.
8. Hall AF, Theofrastous JP, Cundiff GW, Harris RL, Hamilton LF, Swift SE, et al. Interobserver and intraobserver reliability of the proposed International Continence Society, Society of Gynecologic Surgeons, and American Urogynecologic Society pelvic organ prolapse classification system. American Journal of Obstetrics and Gynecology. 1996;175(6):1467-71.
9. Lemos NLDBM, Auge APF, Lunardelli JL, Carramão SDS, Faria ALA, Aoki T. Validation of the Pelvic Organ Prolapse Quantification Index (POP-Q-I): a novel interpretation of the POP-Q system for optimization of POP research. International Urogynecology Journal. 2008;19(7):995-7.
10. Bump RC. The POP-Q system: two decades of progress and debate. International Urogynecology Journal. 2014;25(4):441-3.
11. Lammers K, Prokop M, Vierhout ME, Kluivers KB, Fütterer JJ. A pictorial overview of pubovisceral muscle avulsions on pelvic floor magnetic resonance imaging. Insights into Imaging. 2013;4(4):431-41.
12. Sung VW, Rogers RG, Barber MD, Clark MA. Conceptual framework for patient-important treatment outcomes for pelvic organ prolapse. Neurourol Urodyn. 2014;33(4):414-9.
13. Habashy E, Mahdy AE. Patient-Reported Outcome Measures (PROMs) in Pelvic Floor Disorders. Curr Urol Rep. 2019;20(5):22.
14. Grigoriadis T, Athanasiou S, Rizk D. Female pelvic floor dysfunction questionnaires: the modern Tower of Babel? International Urogynecology Journal. 2020;31(6):1059-61.

15. Gray TG, Vickers H, Krishnaswamy P, Jha S. A systematic review of English language patient-reported outcome measures for use in urogynaecology and female pelvic medicine. International Urogynecology Journal. 2021;32(8):2033-92.
16. Bo K, Angles-Acedo S, Batra A, Braekken IH, Chan YL, Jorge CH, et al. International urogynecology consultation chapter 3 committee 2; conservative treatment of patient with pelvic organ prolapse: Pelvic floor muscle training. Int Urogynecol J. 2022;33(10):2633-67.
17. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;27(5):1171-9.
18. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Quality of Life Research. 2018;27(5):1159-70.
19. Mokkink LB, Terwee CB, Stratford PW, Alonso J, Patrick DL, Riphagen I, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Quality of Life Research. 2009;18(3):313-33.
20. Maxwell M, Semple K, Wane S, Elders A, Duncan E, Abhyankar P, et al. PROPEL: implementation of an evidence based pelvic floor muscle training intervention for women with pelvic organ prolapse: a realist evaluation and outcomes study protocol. BMC Health Services Research. 2017;17(1).
21. Mason SJ, Catto JWF, Downing A, Bottomley SE, Glaser AW, Wright P. Evaluating patient-reported outcome measures (PROMs) for bladder cancer: a systematic review using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. BJU Int. 2018;122(5):760-73.
22. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research. 2018;27(5):1147-57.
23. Barber MD, Chen Z, Lukacz E, Markland A, Wai C, Brubaker L, et al. Further validation of the short form versions of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ). Neurourol Urodyn. 2011;30(4):541-6.
24. Gelhorn HL, Coyne KS, Sikirica V, Gauld J, Murphy M. Psychometric evaluation of health-related quality-of-life measures after pelvic organ prolapse surgery. Female Pelvic Med Reconstr Surg. 2012;18(4):221-6.
25. Price N, Jackson SR, Avery K, Brookes ST, Abrams P. Development and psychometric evaluation of the ICIQ Vaginal Symptoms Questionnaire: the ICIQ-VS. BJOG. 2006;113(6):700-12.
26. Baessler K, Mowat A, Maher CF. The minimal important difference of the Australian Pelvic Floor Questionnaire. Int Urogynecol J. 2019;30(1):115-22.
27. Barber MD, Walters MD, Cundiff GW, Group PT. Responsiveness of the Pelvic Floor Distress Inventory (PFDI) and Pelvic Floor Impact Questionnaire (PFIQ) in women undergoing vaginal surgery and pessary treatment for pelvic organ prolapse. Am J Obstet Gynecol. 2006;194(5):1492-8.
28. Wiegersma M, Panman CM, Berger MY, De Vet HC, Kollen BJ, Dekker JH. Minimal important change in the pelvic floor distress inventory-20 among women opting for conservative prolapse treatment. Am J Obstet Gynecol. 2017;216(4):397 e1- e7.

29. Bradshaw HD, Hiller L, Farkas AG, Radley S, Radley SC. Development and psychometric testing of a symptom index for pelvic organ prolapse. J Obstet Gynaecol. 2006;26(3):241-52.
30. Digesu GA, Khullar V, Cardozo L, Robinson D, Salvatore S. P-QOL: a validated questionnaire to assess the symptoms and quality of life of women with urogenital prolapse. Int Urogynecol J Pelvic Floor Dysfunct. 2005;16(3):176-81; discussion 81.
31. Hagen S, Glazener C, Sinclair L, Stark D, Bugge C. Psychometric properties of the pelvic organ prolapse symptom score. BJOG. 2009;116(1):25-31.
32. Jones GL, Radley SC, Lumb J, Farkas A. Responsiveness of the electronic Personal Assessment Questionnaire-Pelvic Floor (ePAQ-PF). Int Urogynecol J Pelvic Floor Dysfunct. 2009;20(5):557-64.
33. Lowder JL, Ghetti C, Oliphant SS, Skoczylas LC, Swift S, Switzer GE. Body image in the Pelvic Organ Prolapse Questionnaire: development and validation. Am J Obstet Gynecol. 2014;211(2):174 e1-9.
34. Pruijssers B, van der Vaart L, Milani F, Roovers JP, Vollebregt A, van der Vaart H. Minimal Clinically Important Difference (MCID) for the Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire - IUGA Revised (PISQ-IR). J Sex Med. 2021;18(7):1265-70.
35. De Arruda GT, Dos Santos Henrique T, Virtuoso JF. Pelvic floor distress inventory (PFDI)—systematic review of measurement properties. International Urogynecology Journal. 2021;32(10):2657-69.

Appendix 1

Description of the COSMIN checklist measurement properties and standards for reporting each measurement property[21]

Measurement Property	What the measurement property is	COSMIN standards for reporting the measurement property
Content Validity	Does the questionnaire include items relevant to the underlying outcome (construct) of interest? Does it include items covering the whole scope of the outcome? The validity is assessed by examining how the items for inclusion in the questionnaire were generated.	Evidence should be presented of an assessment concerning item relevance and scope. Development and pilot work with experts, clinicians, and patients is typically undertaken and reported.
Structural Validity	The degree to which the scores of the PROM instrument are an adequate reflection of the dimensionality of the construct to be measured.	Factor analysis should be reported for Classical Test Theory. Rasch analysis should be reported for IRT.
Internal consistency	Internal Consistency is the degree of interrelatedness in the construct being measured.	Following initial factor analysis to check scale unidimensionality, Cronbach's alpha should be reported.

Reliability	For a questionnaire to be reliable it should result in the same or similar responses or scores every time, if the circumstances of the people completing the questionnaire remain the same. If the scale is reliable the scores will stay the same when the PROM is completed twice by patients whose health is stable.	Test-retest should be calculates using ICC for continuous scores. Evidence of at least two independent measurements, with an appropriate time interval during which the participants were stable should be reported.
Measurement Error	Checks if changes in PROM score are due to reasons other than genuine changes in the construct being measured (an error in measurement).	Standard Error of Measurement (SEM), Smallest Detectable Chance (SDC) or Limits of Agreement (LoA) should be calculated.
Hypothesis Testing	A reliable and valid questionnaire will detect differences between groups of patients who are known to be different in terms of the construct of interest.	Evidence should be presented that hypotheses were formulated a priori, with the directions of mean differences or relative magnitude of correlations stated.
Criterion Validity	Compares whether PROM are an adequate reflection of a 'gold standard'.	Evidence should be presented that the criterion was an adequate 'gold standard' (in the case which of PROM, the full version of the short form measure)
Responsiveness	Responsiveness (or sensitivity to change) refers to the ability of an outcome measure to detect change over time in the construct to be measured.	Appropriate statistical methods should be used. Reporting statical significance with P values is not encouraged.

		Test should measure the change of the PROM scores, not of health status or magnitude of an intervention.
Cross- Cultural Validity	Measures whether the performance of the questions on a translated or culturally adapted PROM are similar or comparable to the performance of the questions in the original version of the PROM.	The process of translating the PROM should be adequately described. Factor analysis should have been performed.

Appendix 2
Search Strategy

Outcome Measures for Pelvic Organ Prolapse used in conservative management: a systematic review of measurement properties.

Research Questions
1. What are the psychometric properties of PROMs used in adult women with a diagnosis of POP?
2. What is the responsiveness of PROMs used in adult women with a diagnosis of POP undergoing conservative management?

Methods
Systematic Review

Following PRISMA guidelines
Develop a flowchart to capture studies where women with pelvic organ prolapse have completed a PROM which assesses prolapse symptomology and the psychometric properties have been reported.

The study population will be women who are presenting for care for symptoms of prolapse or already have a *diagnosis* of pelvic organ prolapse.

The intervention studied will be any PROM which assesses POP symptomology and has reported psychometric properties.

Ovid Medline, CINAHL, EMBASE, Cochrane, Scopus databases were searched

Controlled vocabulary for Questionnaires:

1. exp pelvic organ prolapse/ or exp cystocele/ or exp rectal prolapse/ or uterine/ or uterine prolapse/ or visceral prolapse/

Free Form Text related to the condition:

2. (Cystocele* OR "bladder descen*" OR "fall" bladder"* OR ""ulg*"bladder" OR ""ladde" protru*" OR "ante" ior wall protru"" OR" "anterior vaginal wal" pro" ru*" OR "anterior wall descen"" OR" "anterior vaginal wal" des" en*" OR "anterior wall bulg*""OR ""nterior vaginal wal" bul"*" OR "anterior vaginal wal" pro" aps*" OR "anterior wall prolap"*" O" "bladder prolaps*" OR "rect" cele* OR "rectal" descen*" OR "falle" rectum" OR "b" lg* "ectum" OR "re "tum "rotru*" OR ""ecta" protru*" OR ""oste" ior wall protr"*" O" "posterior vaginal wa" l pr "tru*" OR "posterior vaginal wa" l pr "laps*" OR "posterior wall prola" s*" "R "posterior wall

desce"*" O" "posterior vaginal wa" l de" cen*" OR "posterior wall bulg*" OR "posterior vaginal wa" l bu" g*" OR "rectal prolap*" OR E "tero" ele* OR "visce" al descen*" OR "vis" eral protru*" OR "hyst" rocele* OR "Uter" ne descen*" OR "apic" l descen*" OR ""alle" uterus" OR "b" lg* "terus" OR "ut "rine" protru*" OR "vagi" al vault protru"" OR" "vaginal vault descen"" OR" "vaginal vault bulg*""OR ""aginal vault prolap"*" O" "protru* pelvic organ" OR "pelvic organ dysfunc" ion""

3. 1 OR 2

Free Form T" xt related to Questionnaires:

4. "Surveys and Questionnaire""/ OR Patient Reported Out "ome Measures/ OR Health Status/ OR Health Impact Assessment/ OR Data Collection/ OR Self-Assessment/

"Surveys and Questionnaire""/ OR Patient Reported Out "ome Measures/ OR Health Impact Assessment/ OR Self-Assessment/

Free Form Text related to Questionnaires:

5. ("self-report*" or "self-as" ess*" or "se" f-ev" luat*" or "s" lf-a "ministered" o" "se" f-rated measure*"" or ""elf-rated outcome*"" or ""atient-reported out "ome*" or "PROM*" or diary or d "arie" or q "estionnaire* or "outcome profil*" or inven "ory or inventor" es or (index not (body mass index or kappa index or similarity index)) or indices or scale* or survey* or "health status" or instrum" nt* or checkl "st* or "health assess*" or "heal" h-evaluat*" or ""heal" h outcome" or ""utco" e* measure*" o" "sy" ptom* measure*" o" "sy" ptom* assess*" or ""sym" tom* evaluat*" o" sco" e*).ti,ab,kf. Or "test or tests or tool*).ti.

6. 4 OR 5

Controlled vocabulary for Psychometric properties

7. Validation Studies/ OR Psychometrics/ OR Reproducibility of results/

Free-form text related to psychometric properties.
Validation Studies/ OR Psychometrics/ OR Reproducibility of results/ or (valid* OR responsive* OR reproduci* OR reliab* OR sensitivity OR specificity OR psychometric* or clinimetric* OR "minimal important differe "ce" OR "minimal clinically i "port" nt difference" or "test-retest" or "int" rnal "consistency" or "measurement error" o" int" rpretability or ""easurement propert*").mp"

Medline Ovid Search Strategy
April 8th
Ovid MEDLINE(R) ALL 1946 to April 08, 2022 (Article # 1399)

1. exp pelvic organ prolapse/ or exp cystocele/ or exp rectal prolapse/ or uterine/ or uterine prolapse/ or visceral prolapse/

2. (prolaps* or Cystocele* or "bladder descen*" or "fall" bladder*" or ""ulg*"bladder" or "b" adde" protru*" or "wall" protru*" or "wa" l de" cen*" or "wa" l bu" g*" or recto "ele*" or "rectal" descen*" or "falle" rectum" or "b" lg* "ectum" or "re "tum "rotru*" or ""ecta" protru*" or E "tero" ele* or "visce" al descen*" or "vis" eral protru*" or "hyst" rocele* or "Uter" ne descen*" or "apic" l descen*" or ""alle" uterus" or "b" lg* "terus" or "ut "rine" protru*" or "vagi" al vault protru"" or ""vaginal vault descen"" or ""vaginal vault bulg*"" or ""rotru* pelvic organ" or "pelvic organ dysfunc" ion"" or procidentia*).mp.

3. "1 or 2

4. "Surveys and Questionnaire""/ or Patient Reported Out "ome Measures/ or Health Impact Assessment/ or Self-Assessment/

5. ("self-report*" or "self-as" ess*" or "se" f-ev" luat*" or "s" lf-a "ministered" o" "se" f-rated measure*"" or ""elf-rated outcome*"" or ""atient-reported out "ome*" or "PROM*" or diary or d "arie" or q "estionnaire* or "outcome profil*" or inven "ory or inventor" es or (index not (body mass index or kappa index or similarity index)) or indices or scale* or survey* or "health status" or instrum" nt* or checkl "st* or "health assess*" or "healt"-evaluat*" or "heal" h outcome" or ""utco" e* measure*" o" "sy" ptom* measure*" o" "sy" ptom* assess*" or ""sym" tom* evaluat*" o" sco" e*).ti,ab,kf. or "tool* or test or tests).ti.

6. 4 or 5

7. Validation Studies/ or Psychometrics/ or Reproducibility of results/ or (valid* or responsive* or reproduci* or reliab* or sensitivity or specificity or psychometric* or clinimetric* or "minimal important differe "ce" or "minimal clinically i "port" nt difference" or "test-retest" or "int" rnal "consistency" or "measurement error" o" int" rpretability or ""easurement propert*").mp"

8. 3 and 6 and 7

Translation to EMBASE -
EMBASE 1946 to April 08, 2022 (Article # 1564)

1. prolapse/ or pelvic floor prolapse/ or exp pelvic organ prolapse/ or visceral prolapse/
2. exp rectum prolapse/
3. (prolaps* or Cystocele* or "bladder descen*" or "fall" bladder*" or ""ulg*"bladder" or "b" adde" protru*" or "wall" protru*" or "wa" l de" cen*" or "wa" l bu" g*" or recto "ele*" or "rectal" descen*" or "falle" rectum" or "b" lg* "ectum" or "re "tum "rotru*" or ""ecta" protru*" or E "tero" ele* or "visce" al descen*" or "vis" eral protru*" or "hyst" rocele* or "Uter" ne descen*" or "apic" l descen*" or ""alle" uterus" or "b" lg* "terus" or "ut "rine" protru*" or "vagı" al vault protru"" or ""vaginal vault descen"" or ""vaginal vault bulg*"" or ""rotru* pelvic organ" or "pelvic organ dysfunc" ion"" or procidentia*).tw,kf.
". 1 or 2 or 3
5. questionnaire/ or open-ended questionnaire/ or structured questionnaire/
6. health status indicator/ or disease activity score/ or "severity of illness index"/
7. health survey/
8. pa" ient-reported outcome/
9. health impact assessment/
10. ("self-report*" or "self-as" ess*" or "se" f-ev" luat*" or "s" lf-a "ministered" o" "se" f-rated measure*"" or ""elf-rated outcome*"" or ""atient-reported out "ome*" or "PROM*" or diary or d "arie" or q "estionnaire* or "outcome profil*" or inven "ory or inventor" es or (index not (body mass index or kappa index or similarity index)) or indices or scale* or survey* or "health status" or instrum" nt* or checkl "st* or "health assess*" or "healt"-evaluat*" or "heal" h outcome" or ""utco" e* measure*" o" "sy" ptom* measure*" o" "sy" ptom* assess*" or ""sym" tom* evaluat*" o" sco" e*).ti,ab,kf. or "tool* or test or tests).ti.
11. 5 or 6 or 7 or 8 or 9 or 10
12. validation study/
13. psychometry/
14. reproducibility/ or measurement precision/
15. (valid* or responsive* or reproduci* or reliab* or sensitivity or specificity or psychometric* or clinimetric* or "minimal important differe "ce" or "minimal clinically i "port" nt difference" or "test-retest" or "int" rnal "consistency" or "measurement error" o" int" rpretability or ""easurement propert*").mp"
16. 12 or 13 or 14 "r 15
17. 4 and 11 and

Translation to CINAL
April 18th

1. pelvic organs prolapse or cystocele or rectal prolapse or uterine prolapse or visceral prolapse

(MH "Pelvic Organ Prolapse+") "R (MH "Cystocele") OR "MH "Recta" Prolapse") OR (MH "Rectocele") OR "MH "Uteri"e Prolaps"") OR (MH" "Vaginal Vault P "olapse") "

2. (prolaps* or Cysto "ele* or "bladder descen*" or "fall" bladder*" or ""ulg*"bladder" or "b" adde" protru*" or "wall" protru*" or "wa" l de" cen*" or "wa" l bu" g*" or recto "ele*" or "rectal" descen*" or "falle" rectum" or "b" lg* "ectum" or "re "tum "rotru*" or ""ecta" protru*" or E "tero" ele* or "visce" al descen*" or "vis" eral protru*" or "hyst" rocele* or "Uter" ne descen*" or "apic" l descen*" or ""alle" uterus" or "b" lg* "terus" or "ut "rine" protru*" or "vagi" al vault protru"" or ""vaginal vault descen"" or ""vaginal vault bulg*"" or ""rotru* pelvic organ" or "pelvic organ dysfunc" ion"" or procidentia*).tw,kf.
"
Tw includes (TI- title word, AB abstract)
Keyword Heading Word (**KF**)
TI (prolaps* or Cystocele* or "bladder descen*" or "fall" bladder*" or ""ulg*"bladder" or "b" adde" protru*" or "wall" protru*" or "wa" l de" cen*" or "wa" l bu" g*" or recto "ele*" or "rectal" descen*" or "falle" rectum" or "b" lg* "ectum" or "re "tum "rotru*" or ""ecta" protru*" or E "tero" ele* or "visce" al descen*" or "vis" eral protru*" or "hyst" rocele* or "Uter" ne descen*" or "apic" l descen*" or ""alle" uterus" or "b" lg* "terus" or "ut "rine" protru*" or "vagi" al vault protru"" or ""vaginal vault descen"" or ""vaginal vault bulg*"" or ""rotru* pelvic organ" or "pelvic organ dysfunc" ion"" or procidentia*)

3. S1 "r S2

4. "Surveys and Questionnaire"" or Patient Reported Outc "me Measures or Health Impact Assessment or Self-Assessment
(MH "Surveys+") OR (MH "Struct" red Ques" ionnaires") OR (MH "Patient-Report" d Outcomes" ") OR (MH "Health Impact A "sessment"" OR (MH "Self Assessment")

S7 - "4 OR S5 OR S6
". ("self-report*" or "self-as" ess*" or "se" f-ev" luat*" or "s" lf-a "ministered" o" "se" f-rated measure*"" or ""elf-rated outcome*"" or ""atient-reported out "ome*" or "PROM*" or diary or d "arie" or q "estionnaire* or "outcome profil*" or inven "ory or inventor" es or (index not (body mass index or kappa index or similarity index)) or indices or scale* or survey* or "health status" or instrum" nt* or checkl "st* or "health assess*" or "healt"-evaluat*" or "heal" h outcome" or ""utco" e* measure*" o" "sy" ptom* measure*" o" "sy" ptom* assess*" or ""sym" tom* evaluat*" o" sco" e*).ti,ab,kf. or "tool* or test or tests).ti.

TI ("self-report*" or "self-as" ess*" or "se" f-ev" luat*" or "s" lf-a "ministered" o" "se" f-rated measure*"" or ""elf-rated outcome*"" or ""atient-reported out "ome*" or "PROM*" or diary or d "arie" or q "estionnaire* or "outcome profil*" or inven "ory or inventor" es or (index not (body mass index or kappa index or similarity index)) or indices or scale* or survey* or "health status" or instrum" nt* or checkl "st* or "health assess*" or "healt"-evaluat*" or "heal" h outcome" or ""utco" e* measure*" o" "sy" ptom* measure*" o" "sy" ptom* assess*" or ""sym" tom* evaluat*" o" "to "l*" or "test" or "test"" or "scor"*")
"

6. S" or S" S" or S8"

7. (S10)Validation Studies/ or (S11) Psychometrics/ or (S12)Reproducibility of results/ or (S13)(valid* or responsive* or reproduci* or reliab* or sensitivity or specificity or psychometric* or clinimetric* or "minimal important differe "ce" or "minimal clinically i "port" nt difference" or "test-retest" or "int" rnal "consistency" or "measurement error" o" int" rpretability or ""easurement propert*").mp"
.mp - multipurpose "ranslate to CINAL- open
(valid* or responsive* or reproduci* or reliab* or sensitivity or specificity or psychometric* or clinimetric* or "minimal important differe "ce" or "minimal clinically i "port" nt difference" or "test-retest" or "int" rnal "consistency" or "measurement error" o" int" rpretability or ""easurement propert*")

"10 OR S11 OR S12 OR "13

Translation to Scopus
April 20, 2022

1. N/A

2. TITLE-ABS-KEY(prolaps* or Cystocele* or "bladder descen*" or "fall" bladder*" or ""ulg*"bladder" or "b" adde" protru*" or "wall" protru*" or "wa" l de" cen*" or "wa" l bu" g*" or recto "ele*" or "rectal" descen*" or "falle" rectum" or "b" lg* "ectum" or "re "tum "rotru*" or ""ecta" protru*" or E "tero" ele* or "visce" al descen*" or "vis" eral protru*" or "hyst" rocele* or "Uter" ne descen*" or "apic" l descen*" or ""alle" uterus" or "b" lg* "terus" or "ut "rine" protru*" or "vagi" al vault protru"" or ""vaginal vault descen"" or ""vaginal vault bulg*"" or ""rotru* pelvic organ" or "pelvic organ dysfunc" ion"" or procidentia*)

3. N/A"

4. N/A

5. TITLE-ABS-KEY("self-report*" or "self-as" ess*" or "se" f-ev" luat*" or "s" lf-a "ministered" o" "se" f-rated measure*"" or ""elf-rated outcome*"" or ""atient-reported out "ome*" or "PROM*" or diary or d "arie" or q "estionnaire* or "outcome profil*" or inven "ory or inventor" es or (index not (body mass index or kappa index or similarity index)) or indices or scale* or survey* or "health status" or instrum" nt* or checkl "st* or "health assess*" or "healt"-evaluat*" or "heal" h outcome" or ""utco" e* measure*" o" "sy" ptom* measure*" o" "sy" ptom* assess*" or ""sym" tom* evaluat*" o" sco" e*) or TITLE(tool" or test or tests)

6. N/A

7. TITLE-ABS-KEY(valid* or responsive* or reproduci* or reliab* or sensitivity or specificity or psychometric* or clinimetric* or "minimal important differe "ce" or "minimal clinically i "port" nt difference" or "test-retest" or "int" rnal "consistency" or "measurement error" o" int" rpretability or ""easurement propert*")

8" 3 and 6 and 7

Scop" s Advanced Search April 22 #1996

TITLE-ABS-KEY (prolaps* OR cystocele* OR "bladder descen*" OR "fa"l* bladder*" O" "bul"* bladder" OR" "blad"er protru*" "R "wa"l protru*" OR ""wall "escen*" OR ""wall "ulg*" OR r"ctocel"* OR "re"tal descen*" OR "fal"en rectum" OR" "bulg" rectum" OR ""rectu" protru*" O" "rec"al protru*" O" ente"ocele* OR "v"sceral descen*" OR "v"sceral protru*" "OR hy"terocele* OR ""terine descen*" OR "ap"cal descen*" O" "fal"en uterus" OR" "bulg" uterus" OR ""uteri"e protru*" "R "va"inal vault prot"u*" O" "vaginal vault desc"n*" O" "vaginal vault bulg"" OR ""protru* pelvic org"n" OR" "pelvic organ dysfu"ction"" OR procidentia*) AND" (TITLE-ABS-KEY ("self-report*" OR "self"ssess*" OR ""self-"valuat*" OR" "self"administered"" OR ""elf-rated

measure"" OR ""self-rated outcome"" OR ""patient-reported
o"tcome*" OR "PROM*" OR diary "OR di"ries "OR questionnaire* OR "outcome
profil*" OR inv"ntory OR inve"tories OR (index AND NOT ("body mass
index" OR "k"pa index" OR "simila"ity index" ") OR"
indices OR sc"e* OR survey* OR "health
status" OR instr"ment* OR ch"cklist* OR "health assess*" OR "hea"th-
evaluat*" "R "he"lth outcome" O" "out"ome* measure*"" OR ""ymptom* measure*""
OR ""ymptom* assess*" "OR "s"mptom* evaluat*"" OR s"ore*) OR TITLE"(
tool* OR test OR tests)) AND TITLE-ABS-KEY (
valid* OR responsive* OR reproduci* OR reliab* OR sensitivity OR specificity OR psyc
hometric* OR clinimetric* OR "minimal important differe"ce" OR "minimal
clinically"import"nt difference" OR "test-retest" OR "intern"l consisten"y" OR"
"measurement error"" OR i"terpretability O" "measurement propert*")

Cochrane Results
April 22, 2022

1. [mh "pelvic org" n prolapse"] o" [mh "cystocele"] or "mh "recta" prolapse"] or [mh ""uterine prolap "e"] or [mh "^" visceral prola "se"]

2. ("Prolaps* or Cystoc"le* or (bladder NEXT descen*) or (fall NEXT bladder*) or (bulg* NEXT bladder) or (bladder NEXT protru*) or (wall NEXT protru*) or (wall NEXT descen*) or (wall NEXT bulg*) or rectocele* or (rectal NEXT descen*) or (fallen NEXT rectum) or (bulg* NEXT rectum) or (rectum NEXT protru*) or (rectal NEXT protru*) or Enterocele* or (visceral NEXT descen*) or (visceral NEXT protru*) or hysterocele* or (Uterine NEXT descen*) or (apical NEXT descen*) or "fallen uterus" or (bulg*N "XT uterus) or" (uterine NEXT protru*) or ("vaginal vault" NEXT protr"*) or ("vagin" l vault" NEXT desce"*) or ("vagin" l vault" NEXT bulg*" or (protru* "EXT "pelvic organ") or "pelvic" organ dysfun "tion"" or procidentia*):ti,ab,k"

3. #1 or #2

4. [mh ^ "Surveys and Questionnaire""] or [mh ^ "Patient Report" d Outcome "easures"] or [mh ^ "Health Impact "ssessment"" or [mh ^ "Self-Assessmen""]

5. ((s "lf NEXT report*" or (self NEXT assess*) or (self NEXT evaluat*) or "self administered" or "se" f rated measure" "r "s" lf rated outcome" "r "p" tient reported out "ome"" or "PROM" or diary or di "ries" or q "estionnaire* or (outcome NEXT profil*) or inventory or inventories or (index not ("body mass index" or "kapp" index" or "sim" lari" y index")) "r in "ices or scale* o" survey* or "health status" or instrum "nt or checkli "t or (health NEXT assess*) or (health NEXT evaluat*) or "health outcome" or "outco" e measure" or "symp" om NEXT measure") or (symptom NEXT assess*) or (symptom* NEXT evaluat*) or score*):ti,ab,kw or (tool* or test or tests):ti

6. #4 or #5

7. [mh^ "Validation Studies"] or ["h ^ "Psychometrics"" or [mh ^" "eproducibilit" of result""] or (valid* or responsiv"* or reproduci* or reliab* or sensitivity or specificity or psychometric* or clinimetric* or "minimal important differe "ce" or "minimal clinically

Covidence - all databases uploaded April 25th.
Article total: #2494

Inclusion criteria:
1) included at least one group of adult women presenting for care for prolapse symptoms or pelvic organ prolapse diagnosis
2) evaluated at least one patient-reported outcome measure for pelvic organ prolapse
3) reported on at least one psychometric property
4) Original results only
5) Publication types: full-texted, peer-reviewed published articles of any study design

Exclusion Criteria:
1) Language for which translation is not available to our research team

Updated May 30,2022
Inclusion Criteria
1) include at least one group of adult women presenting for care for prolapse symptoms or pelvic organ prolapse diagnosis

2) evaluated at least one patient-reported outcome measure for pelvic organ prolapse

3) reported on at least one psychometric property (Internal consistency, reliability, measurement error, content validity, face validity, construct validity, structural validity, hypothesis testing, and responsiveness).

4) Original results only

5) Publication types: full-texted, peer-reviewed published articles of any study design

Exclusion Criteria:
1) Language for which translation is not available to our research team

2) Measurement Properties: cross-cultural validity and criterion validity.
Cross-cultural validity - Cross-cultural validity, that is the process of examining the items of the PROM and how they translate to adequately reflect a language or culture, is beyond the scope of this paper and will not be included. Only original versions of the PROMs were included in this review.

Criterion Validity (including concurrent validity & predictive validity)

3) Patient-reported outcome measurement tools used as a predictor tool. The use of a PROM as a predictor is not a measurement property and will therefore be excluded in this review.

Exclusion Reasons for Full Text

1. Wrong Population
2. No PROM for Prolapse
3. No psychometric property was reported
4. No original results
5. Wrong publication type (Conference reports and theses are excluded)
6. Wrong language
7. Animal studies

Appendix 3

COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) User Manual
https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf

COSMIN methodology for assessing the content validity of PROMs User Manual
https://cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf

Appendix 4
Table of Update Criteria for Good Measurement Properties[17]

Measurement property/Rating	Criteria
Structural Validity	
Sufficient (+)	**CTT**: CFA: CFI or TLI or comparable measure >0.95 OR RMSEA ≤0.06 OR SRMR < 0.08 **IRT/Rasch:** no violation of unidimensionality: CFI or TLI or comparable measure >0.95 OR RMSEA ≤0.06 OR SRMR < 0.08 AND No violation of local independence: residual correlations among the items after controlling for the dominant factor <0.20 or Q3's <0.37 ANQ3'so violation of monotonicity: adequate looking graphs OR item scalability >0.30 AND adequate model fit: IRT: $\chi 2$ >0.01. Rasch: infit and outfit mean squares ≥ 0.5 and ≤ 1.5 OR Z-standardized values > -2 and <2
Indeterminate (?)	CTT: Not all information for '+' reported. 'R'/Rasch: Model fit not reported
Insufficient (**-**)	Criteria for '+' not met
Internal consistency	
Sufficient (+)	At least low evidence for structural validity AND Cronbach's alpCronbach's0 for each unidimensional scale or subscale
Indeterminate (?)	Criteria for "at least low "evidence for structural validity" not met
Insufficient (-)	At least low evidence for structural validity AND Cronbach's alpCronbach's70 for each unidimensional scale or subscale
Reliability	
Sufficient (+)	ICC or weighted Kappa ≥0.70
Indeterminate (?)	ICC or weighted Kappa not reported
Insufficient (-)	ICC or weighted Kappa <0.70
Hypothesis Testing for Construct Validity	
Sufficient (+)	The result is in accordance with the hypothesis
Indeterminate (?)	No hypotheses defined (by the review team)
Insufficient (-)	The result is not in accordance with the hypotheses
Responsiveness	
Sufficient (+)	The result is in accordance with the hypothesis OR AUC ≥0.70
Indeterminate (?)	No hypotheses defined (by the review team)
Insufficient (-)	The result is not in accordance with the hypothesis OR AUC <0.70