National Institute for Health and Care Excellence

Draft for consultation

Pelvic floor dysfunction: prevention and non-surgical management

Methods

NICE guideline number tbc

Supplement 1

June 2021

Draft for Consultation

Supplementary material was developed by the National Guideline Alliance which is part of the Royal College of Obstetricians and Gynaecologists ts

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and/or their carer or guardian.

Local commissioners and/or providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN:

Contents

Development of the guideline

Remit

To see “What this guideline covers” and “What this guideline does not cover” see the guideline scope Pelvic floor dysfunction: prevention and non-surgical management.

Methods

This guideline was developed using the methods described in the 2018 NICE guidelines manual. Declarations of interest were recorded according to the NICE conflicts of interest policy.

Developing the review questions and outcomes

The review questions developed for this guideline were based on the key areas identified in the guideline scope. They were drafted by the NGA technical team, and refined and validated by the guideline committee.

The review questions were based on the following frameworks:

- population, intervention, comparator and outcome (PICO) for reviews of interventions
- prognostic reviews – using population, exposure to a risk or prognostic factor, confounders and outcome (PECO)
- qualitative reviews – using population, phenomenon of interest and context (PICo)

Full literature searches, critical appraisals and evidence reviews were completed for all review questions.

The review questions and evidence reviews corresponding to each question (or group of questions) are summarised below.

Table 1: Summary of review questions and index to evidence reviews

Evidence review	Review question(s)	Type of review
[A] Community information strategies	What information strategies are effective in raising awareness about prevention of pelvic floor dysfunction?	Intervention
[B] Risk factors for pelvic floor dysfunction	What are the non-obstetric risk factors (for example age, ethnicity and family history, diet [including caffeine and alcohol], weight, smoking, physical activity) for pelvic floor dysfunction? What are the obstetric risk factors for pelvic floor dysfunction?	Prognostic
[C] Co-existing long-term conditions	What co-existing long-term conditions (for example chronic respiratory disorders) are associated with a higher risk of pelvic floor dysfunction?	Prognostic
[D] Prediction tools for pelvic floor dysfunction	What is the effectiveness of prediction tools for identifying women at risk of PFD?	Intervention
[E] Lifestyle factors for the prevention of pelvic floor dysfunction	What is the effectiveness of modifying lifestyle factors (diet [including caffeine and alcohol], weight loss, stopping smoking, physical activity) for preventing pelvic floor dysfunction?	Intervention
[F] Pelvic floor muscle training for the prevention of pelvic floor dysfunction	What is the effectiveness of pelvic floor muscle training for preventing pelvic floor dysfunction?	Intervention*

Evidence review	Review question(s)	Type of review
[G] Information provision related to the management of pelvic floor dysfunction (people's views and experiences)	What information is valued by women with symptoms associated with pelvic floor dysfunction and their partners or carers?	Qualitative
[H] Information provision about management of pelvic floor dysfunction (most effective ways)	What information provision strategies are effective for women with symptoms associated with pelvic floor dysfunction?	Intervention
[I] Assessment in non-specialist care	What assessments should be conducted in non-specialist care to identify whether the signs and symptoms at presentation are associated with pelvic floor dysfunction?	Intervention
[J] Weight loss interventions	What is the effectiveness of weight loss interventions for improving symptoms of pelvic floor dysfunction?	Intervention
[K] Dietary factors for the management of symptoms	What dietary factors can increase or decrease symptoms of pelvic floor dysfunction?	Intervention
[L] Physical activity for the management of symptoms	What types of physical activity can increase or decrease symptoms of pelvic floor dysfunction?	Intervention
[M] Pelvic floor muscle training for the management of symptoms	What is the effectiveness of pelvic floor muscle training (including Kegel exercises, biofeedback, weighted vaginal cones, and electrical stimulation) for improving symptoms of pelvic floor dysfunction?	Intervention
[N] Physical devices for the management of symptoms	What is the effectiveness of physical devices (including support garments, pessaries and dilators) for improving symptoms of pelvic floor dysfunction?	Intervention*
[O] Psychological interventions	What is the effectiveness of psychological interventions for women with symptoms associated with pelvic floor dysfunction?	Intervention
[P] Behavioural approaches to the management of symptoms	What is the effectiveness of behavioural approaches (for example toilet training, seating, splinting) for improving symptoms of pelvic floor dysfunction?	Intervention
[Q] Pharmacological management	What is the effectiveness of pharmacological management for urinary incontinence associated with pelvic floor dysfunction?	Intervention
[R] Community based multidisciplinary teams	What competencies should be represented in a community-based multidisciplinary team for the management of symptoms associated with pelvic floor dysfunction?	Intervention

[1] *Original health economic analysis conducted*

The COMET database was searched for core outcome sets relevant to this guideline. No core outcome sets were identified and therefore the outcomes were chosen based on committee discussions.

Additional information related to development of the guideline is contained in:

- Supplement 2 (Economics)
- Supplement 3 (NGA staff list).

Searching for evidence

Scoping search

During the scoping phase, searches were conducted for previous guidelines, economic evaluations, health technology assessments and systematic reviews.

Systematic literature search

Systematic literature searches were undertaken to identify published evidence relevant to each review question.

Databases were searched using subject headings, free-text terms and, where appropriate, study type filters. Where possible, searches were limited to retrieve studies published in English. All the searches were conducted in the following databases: Medline, Medline-in-Process, Cochrane Central Register of Controlled Trials (CCTR), Cochrane Database of Systematic Reviews (CDSR), Database of Abstracts of Reviews of Effects (DARE), Health Technology Assessments (HTA) and Embase. For qualitative review questions or those questions which covered multidisplinary working, CINAHL or Emcare and PsycINFO were also searched.

Searches were run once for all reviews during development. Searches for evidence reviews E, F and J-N were updated in February 2021, six weeks in advance of the final committee meeting before consultation on the draft guideline.

Details of the search strategies, including the study-design filters used and databases searched, are provided in Appendix B of each evidence review.

Economic systematic literature search

Systematic literature searches were also undertaken to identify published economic evidence. Databases were searched using subject headings, free-text terms and, where appropriate, an economic evaluations search filter.

A single search, using the population search terms used in the evidence reviews, was conducted to identify economic evidence in the NHS Economic Evaluation Database (NHS EED) and HTA. Another single search, using the population search terms used in the evidence reviews combined with an economic evaluations search filter, was conducted in Medline, Medline in Process and Embase. Where possible, searches were limited to studies published in English.

As with the general literature searches, the economic literature searches were updated in February 2021, six weeks in advance of the final committee meeting before consultation on the draft guideline.

Details of the search strategies, including the study-design filter used and databases searched, are provided in Appendix B of each evidence review.

Quality assurance

Search strategies were quality assured by cross-checking reference lists of relevant studies, analysing search strategies from published systematic reviews and asking members of the committee to highlight key studies. The principal search strategies for each search were also quality assured by a second information scientist using an adaptation of the PRESS 2015 Guideline Evidence-Based Checklist (McGowan 2016). In addition, all publications highlighted by stakeholders at the time of the consultation on the draft scope were considered for inclusion.

Reviewing research evidence

Systematic review process

The evidence was reviewed in accordance with the following approach.

- Potentially relevant articles were identified from the search results for each review question by screening titles and abstracts. Full-text copies of the articles were then obtained.
- Full-text articles were reviewed against pre-specified inclusion and exclusion criteria in the review protocol (see Appendix A of each evidence review).
- Key information was extracted from each article on study methods and results, in accordance with factors specified in the review protocol. The information was presented in a summary table in the corresponding evidence review and in a more detailed evidence table (see Appendix D of each evidence review).
- Included studies were critically appraised using an appropriate checklist as specified in Developing NICE guidelines: the manual. Further detail on appraisal of the evidence is provided below.
- Summaries of evidence by outcome were presented in the corresponding evidence review and discussed by the committee.

Review questions selected as high priorities for economic analysis (and those selected as medium priorities and where economic analysis could influence recommendations) and complex review questions were subject to dual screening and study selection through a 10% random sample of articles. Any discrepancies were resolved by discussion between the first and second reviewers or by reference to a third (senior) reviewer. For the remaining review questions, internal (NGA) quality assurance processes included consideration of the outcomes of screening, study selection and data extraction and the committee reviewed the results of study selection and data extraction. The review protocol for each question specifies whether dual screening and study selection was undertaken for that particular question. Drafts of all evidence reviews were quality assured by a senior reviewer.

Type of studies and inclusion/exclusion criteria

Inclusion and exclusion of studies was based on criteria specified in the corresponding review protocol.

Pelvic floor dysfunction covers a variety of symptoms including: urinary incontinence, emptying disorders of the bladder, faecal incontinence, emptying disorders of the bowel, pelvic organ prolapse, sexual dysfunction and chronic pelvic pain syndromes. Interventions in this area are usually directed at specific symptoms so for most of the

intervention evidence reviews studies were included even if they only considered a single symptom (such as urinary incontinence, emptying disorders of the bladder, emptying disorders of the bowel, faecal incontinence, sexual dysfunction, pelvic organ prolapse and pelvic pain). This was not the case for evidence report [Q] pharmacological management, given existing NICE guidance for pharmacological management of specific symptoms (for example Urinary incontinence and pelvic organ prolapse in women [NG123] and Faecal incontinence in adults: management [CG49]) the evidence review was restricted to studies specifically in women with pelvic floor dysfunction. For evidence reviews [B] risk factors for pelvic floor dysfunction and [C] co-existing long-term conditions it became clear during screening that studies were available on pelvic floor dysfunction as a condition, so for these questions any studies focused on single symptoms were excluded.

Systematic reviews with meta-analyses were considered to be the highest quality evidence that could be selected for inclusion.

For intervention reviews, randomised controlled trials (RCTs) were prioritised for inclusion because they are considered to be the most robust type of study design that could produce an unbiased estimate of intervention effects. Where there was limited evidence from RCTs, non-randomised studies (NRS) were considered for inclusion.

For prognostic reviews, prospective and retrospective cohort and case–control studies and case series were considered for inclusion. Studies that included multivariable analysis were prioritised.

For qualitative reviews, studies using focus groups, structured interviews or semi-structured interviews were considered for inclusion. Where qualitative evidence was sought, data from surveys or other types of questionnaire were considered for inclusion only if they provided data from open-ended questions, but not if they reported only quantitative data.

The committee was consulted about any uncertainty regarding inclusion or exclusion of studies. A list of excluded studies for each review question, including reasons for exclusion is presented in Appendix J of the corresponding evidence review.

Narrative reviews, posters, letters, editorials, comment articles, unpublished studies and studies published in languages other than English were excluded. Conference abstracts were not considered for inclusion because conference abstracts typically do not have sufficient information to allow for full critical appraisal.

Methods of combining evidence

When planning reviews (through preparation of protocols), the following approaches for data synthesis were discussed and agreed with the committee.

Data synthesis for intervention studies

Pairwise meta-analysis

Meta-analysis to pool results from comparative intervention studies was conducted where possible using Cochrane Review Manager (RevMan5) software.

For dichotomous outcomes, such as mortality, the Mantel–Haenszel method with a fixed effect model was used to calculate risk ratios (RRs). For all outcomes with zero

events in both arms the risk difference was presented. For outcomes in which the majority of studies had low event rates (<1%), Peto odds ratios (ORs) were calculated as this method performs well when events are rare (Bradburn 2007).

For continuous outcomes, measures of central tendency (mean) and variation (standard deviation; SD) are required for meta-analysis. Data for continuous outcomes, such as quality of life, were meta-analysed using an inverse-variance method for pooling weighted mean differences (WMDs). Where SDs were not reported for each intervention group, the standard error (SE) of the mean difference was calculated from other reported statistics (p values or 95% confidence intervals; CIs) and then meta-analysis was conducted as described above.

If a study reported only the summary statistic and 95% CI the generic-inverse variance method was used to enter data into RevMan5. If the control event rate was reported this was used to generate the absolute risk difference in GRADEpro. If multivariable analysis was used to derive the summary statistic but no adjusted control event rate was reported, no absolute risk difference was calculated.

When evidence was based on studies that reported descriptive data or medians with interquartile ranges or p values, this information was included in the corresponding GRADE tables (see below) without calculating relative effects. Consequently the imprecision of the effect estimate could not be assessed as per standard methods so the evidence was downgraded by one level in these cases.

For some reviews, evidence was either stratified from the outset or separated into subgroups when heterogeneity was encountered. The stratifications and potential subgroups were pre-defined at the protocol stage (see the protocols for each review for further detail). Where evidence was stratified or subgrouped the committee considered on a case by case basis if separate recommendations should be made for distinct groups. Separate recommendations may be made where there is evidence of a differential effect of interventions in distinct groups. If there is a lack of evidence in one group, the committee considered, based on their experience, whether it was reasonable to extrapolate and assume the interventions will have similar effects in that group compared with others

When meta-analysis was undertaken, the results were presented visually using forest plots generated using RevMan5 (see Appendix E of relevant evidence reviews).

When case series were included, descriptive data from the studies were included and no further analysis was performed.

Data synthesis in evidence review [M] Pelvic floor muscle training for the management of symptoms

No meta-analysis was done for evidence review [M] Pelvic floor muscle training for the management of symptoms. Given there were existing high quality systematic reviews with meta-analyses for the comparisons of interest, the committee were presented with a summary of the results of these systematic reviews. RCTs published since the systematic reviews were also included if they reported outcomes covered by the reviews, or comparisons not covered by the reviews. The committee made subjective judgements as to whether any additional evidence from RCTs affected their confidence in the effects reported in the existing systematic reviews.

Data synthesis for prognostic reviews

ORs, RRs or hazard ratios (HRs) with 95% CIs reported in published studies were extracted or calculated by the NGA technical team to examine relationships between risk factors and outcomes of interest. Ideally analyses would have adjusted for key confounders (such as age or parity) to be considered for inclusion. Recognising variation across studies in terms of populations, risk factors, outcomes and statistical analysis methods (including adjustments for confounding factors), prognostic data were not meta-analysed, but results from individual studies were presented in the evidence reviews.

Data synthesis for qualitative reviews

Where possible, a meta-synthesis was conducted to combine evidence from qualitative studies. Whenever studies identified a qualitative theme relevant to the protocol, this was extracted and the main characteristics were summarised. When all themes had been extracted from studies, common concepts were categorised and tabulated. This included information on how many studies had contributed to each theme identified by the NGA technical team.

Themes from individual studies were integrated into a wider context and, when possible, overarching categories of themes with sub-themes were identified. Themes were derived from data presented in individual studies. When themes were extracted from 1 primary study only, theme names used in the guideline mirrored those in the source study. However, when themes were based on evidence from multiple studies, the theme names were assigned by the NGA technical team. The names of overarching categories of themes were also assigned by the NGA technical team.

Emerging themes were placed into a thematic map representing the relationship between themes and overarching categories. The purpose of such a map is to show relationships between overarching categories and associated themes.

Appraising the quality of evidence

Intervention studies

Pairwise meta-analysis

GRADE methodology for intervention reviews

For intervention reviews, the evidence for outcomes from included RCTs and comparative non-randomised studies was evaluated and presented using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology developed by the international GRADE working group.

When GRADE was applied, software developed by the GRADE working group (GRADEpro) was used to assess the quality of each outcome, taking account of individual study quality factors and any meta-analysis results. Results were presented in GRADE profiles (GRADE tables).

The selection of outcomes for each review question was agreed during development of the associated review protocol in discussion with the committee. The evidence for each outcome was examined separately for the quality elements summarised in

Table 2. Criteria considered in the rating of these elements are discussed below. Each element was graded using the quality ratings summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having a 'serious' or 'very serious' quality issue. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

The initial quality rating was based on the study design: RCTs and NRS assessed by ROBINS-I start as 'high' quality evidence, other non-randomised studies start as 'low' quality evidence. The rating was then modified according to the assessment of each quality element (Table 2). Each quality element considered to have a 'serious' or 'very serious' quality issue was downgraded by 1 or 2 levels respectively (for example, evidence starting as 'high' quality was downgraded to 'moderate' or 'low' quality). In addition, there was a possibility to upgrade evidence from non-randomised studies (provided the evidence for that outcome had not previously been downgraded) if there was a large magnitude of effect, a dose–response gradient, or if all plausible confounding would reduce a demonstrated effect or suggest a spurious effect when results showed no effect.

Table 2: Summary of quality elements in GRADE for intervention reviews

Quality element	Description
Risk of bias ('Study limitations')	This refers to limitations in study design or implementation that reduce the internal validity of the evidence
Inconsistency	This refers to unexplained heterogeneity in the results
Indirectness	This refers to differences in study populations, interventions, comparators or outcomes between the available evidence and inclusion criteria specified in the review protocol
Imprecision	This occurs when a study has few participants or few events of interest, resulting in wide confidence intervals that cross minimally important thresholds
Publication bias	This refers to systematic under- or over-estimation of the underlying benefit or harm resulting from selective publication of study results

Table 3: GRADE quality ratings (by quality element)

Quality issues	Description
None or not serious	No serious issues with the evidence for the quality element under consideration
Serious	Issues with the evidence sufficient to downgrade by 1 level for the quality element under consideration
Very serious	Issues with the evidence sufficient to downgrade by 2 levels for the quality element under consideration

Table 4: Overall quality of the evidence in GRADE (by outcome)

Overall quality grading	Description
High	Further research is very unlikely to change the level of confidence in the estimate of effect

Overall quality grading	Description
Moderate	Further research is likely to have an important impact on the level of confidence in the estimate of effect and may change the estimate
Low	Further research is very likely to have an important impact on the level of confidence in the estimate of effect and is likely to change the estimate
Very low	The estimate of effect is very uncertain

Assessing risk of bias in intervention reviews

Bias is a systematic error, or consistent deviation from the truth in results obtained. When a risk of bias is present the true effect can be either under- or over-estimated.

Risk of bias in RCTs was assessed using the Cochrane risk of bias tool version 2 (see Appendix H in Developing NICE guidelines: the manual).

The Cochrane risk of bias tool assesses the following possible sources of bias:

- risk of bias arising from the randomization process
- risk of bias due to deviations from the intended interventions
- risk of bias due to missing outcome data
- risk of bias due to measurement of the outcome
- risk of bias in selection of the reported result

A study with a poor methodological design does not automatically imply high risk of bias; the bias is considered individually for each outcome and it is assessed whether the chosen design and methodology will impact on the estimation of the intervention effect.

More details about version 2 of the Cochrane risk of bias tool can be found in Section 8 of the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2011).

For systematic reviews the ROBIS checklist was used (see Appendix H in Developing NICE guidelines: the manual).

For non-randomised studies the ROBINS-I checklist was used (see Appendix H in Developing NICE guidelines: the manual).

Assessing inconsistency in intervention reviews

Inconsistency refers to unexplained heterogeneity in results of meta-analysis. When estimates of treatment effect vary widely across studies (that is, there is heterogeneity or variability in results), this suggests true differences in underlying effects. Inconsistency is, thus, only truly applicable when statistical meta-analysis is conducted (that is, results from different studies are pooled). When outcomes were derived from a single study the rating 'no serious inconsistency' was used when assessing this domain, as per GRADE methodology (Santesso 2016).

Inconsistency was assessed visually by inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis (for example if the point estimates of the individual studies consistently showed benefits or harms). This was supported by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating serious heterogeneity, and more than 80% indicating very serious heterogeneity. When serious or very

serious heterogeneity was observed, possible reasons were explored and subgroup analyses were performed as pre-specified in the review protocol where possible. In the case of unexplained heterogeneity, sensitivity analyses were planned based on the quality of studies, eliminating studies at high risk of bias (in relation to randomisation, allocation concealment and blinding, and/or missing outcome data).

When no plausible explanation for the serious or very serious heterogeneity could be found, the quality of the evidence was downgraded in GRADE for inconsistency and the meta-analysis was re-run using the Der-Simonian and Laird method with a random effects model and this was used for the final analysis.

Assessing indirectness in intervention reviews

Directness refers to the extent to which populations, interventions, comparisons and outcomes reported in the evidence are similar to those defined in the inclusion criteria for the review and was assessed by comparing the PICO elements in the studies to the PICO defined in the review protocol. Indirectness is important when such differences are expected to contribute to a difference in effect size, or may affect the balance of benefits and harms considered for an intervention.

Assessing imprecision and importance in intervention reviews

Imprecision in GRADE methodology refers to uncertainty around the effect estimate and whether or not there is an important difference between interventions (that is, whether the evidence clearly supports a particular recommendation or appears to be consistent with several candidate recommendations). Therefore, imprecision differs from other aspects of evidence quality because it is not concerned with whether the point estimate is accurate or correct (has internal or external validity). Instead, it is concerned with uncertainty about what the point estimate actually represents. This uncertainty is reflected in the width of the CI.

The 95% CI is defined as the range of values within which the population value will fall on 95% of repeated samples, were the procedure to be repeated. The larger the study, the smaller the 95% CI will be and the more certain the effect estimate.

Imprecision was assessed in the guideline evidence reviews by considering whether the width of the 95% CI of the effect estimate was relevant to decision making, considering each outcome independently. This is illustrated in Figure 1, which considers a positive outcome for the comparison of two treatments. Three decision-making zones can be differentiated, bounded by the thresholds for minimal importance (minimally important differences; MIDs) for benefit and harm.

When the CI of the effect estimate is wholly contained in 1 of the 3 zones there is no uncertainty about the size and direction of effect, therefore, the effect estimate is considered precise; that is, there is no imprecision.

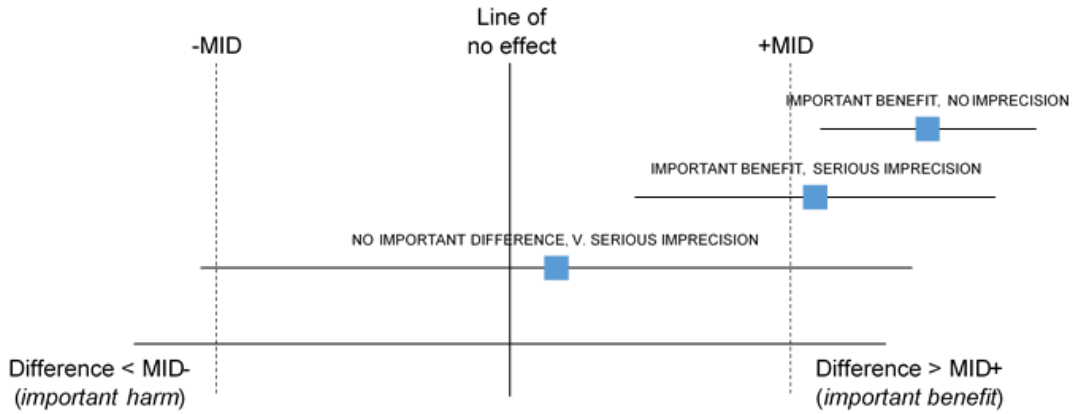
When the CI crosses 2 zones, it is uncertain in which zone the true value of the effect estimate lies and therefore there is uncertainty over which decision to make. The CI is consistent with 2 possible decisions, therefore, the effect estimate is considered to be imprecise in the GRADE analysis and the evidence is downgraded by 1 level ('serious imprecision').

When the CI crosses all 3 zones, the effect estimate is considered to be very imprecise because the CI is consistent with 3 possible decisions and there is

therefore a considerable lack of confidence in the results. The evidence is therefore downgraded by 2 levels in the GRADE analysis ('very serious imprecision').

Implicitly, assessing whether a CI is in, or partially in, an important zone, requires the guideline committee to estimate an MID or to say whether they would make different decisions for the 2 confidence limits.

Figure 1: Assessment of imprecision and importance in intervention reviews using GRADE



MID, minimally important difference

Defining minimally important differences for intervention reviews

The committee was asked whether there were any recognised or acceptable MIDs in the published literature and community relevant to the review questions under consideration. The MIDs identified in the literature are summarised in Table 5.

Table 5: MIDs identified in the literature

Tool	MID	Population	Source
PPBC - Patient Perception of Bladder Condition	≥1-point or ≥2-point improvement	OAB (males and females)	Abrams 2017
OAB-q – Overactive Bladder Questionnaire (made up of two scales, one on Symptom Bother and other on total HRQoL both with same MID).	≥10-point improvement		
FSFI desire domain (Female Sexual Function Index)	+0.6	Premenopausal women with hypoactive sexual desire disorder (HSDD) and mixed HSDD/female sexual arousal disorder (FSAD)	Althof 2019
FSDS-DAO item 13 (feeling bothered by low sexual desire) (Female Sexual Distress Scale Desire/ Arousal/ Orgasm)	-1		
FSFI arousal domain (Female Sexual Function Index)	+0.6-+0.9		
FSDS-DAO item 14 (concerned by difficulty with sexual arousal)	-1		

Tool	MID	Population	Source
(Female Sexual Distress Scale Desire/ Arousal/ Orgasm)			
FSFI total score (Female Sexual Function Index)	+2.1		
FSDS-DAO total score (Female Sexual Distress Scale Desire/ Arousal/ Orgasm)	-7		
Number of satisfying sexual events (SSE) per 28 days/4 weeks.	+1		
APFQ - Australian Pelvic Floor Questionnaire: Global PFD	1.3	Women SUI	Baessler 2019
APFQ - Australian Pelvic Floor Questionnaire: Global PFD	1.0	Women with POP	
UDI (urinary distress inventory)	11	Women with SUI	Barber 2009
UDI-stress subscales (subscale of the pelvic floor distress inventory)	8		
UIQ (urinary impact questionnaire – of the pelvic floor impact questionnaire)	16		
Vaizey scores	-5	Faecal incontinence	Bols 2010
Renzi Obstructed Defecation Syndrome	2	Men and women with ODS diagnosis	Caetano 2017
UDI	-30 to -14	Women with SUI undergoing continence surgery	Chan 2013
UIQ	-28 to -14		
POPDI	-44to -21	Women with SUI undergoing pelvic floor repair	
POPIQ	-40—27		
UDI	-22 to -16		
UIQ	-37 to -31		
CRADI	-37 to -14		
CRAIQ	-34 to -6		
POPDI	-16	Women with SUI who received vaginal pessary	
POPIQ	-29		
UDI	-28		
UIQ	-17		
CRADI	-25		
CRAIQ	-31		
OAB-Q (AND ALL SUBSCALES)	10	Continent and incontinent patients with OAB and nocturia	Coyne 2006
UDI	-35	Women with urge-predominant UI	Dyer 2011
UDI irritative	-25		

Tool	MID	Population	Source
OBSS - Overactive Bladder Symptom Score	-3	Men and women with OAB	Gotoh 2011
I-QOL incontinence Quality of Life questionnaire	4.74	Women with involuntary urine loss	Halme 2015
SF-6D	0.0126		
SF-6D	0.026	Women with POP	Harvie 2019
EQ-5D	0.025		
PRAFAB-questionnaire	2.5 to 4.6 (non severe stress UI) 4.5 to 7.0 (sever stress UI) 2.5 to3.4 (non-severe urgency UI) 4.0 to 4.4 (sever urgency UI)	Women with primary or recurrent UI	Hendricks 2007
PRAFAB-Q	1.9 to 2.7 (non-severe) 3.6-4.1 (severe)	Women with stress UI	Hendricks 2008I
Incontinence episodes	3/week decrease	Men and Women OAB	Homma 2006
FISI long	-4	Women with faecal incontinence	Jelovsek 2014
CRADI long	-11		
CRADI short	-5		
CRAIQ long	-18		
CRAIQ short	-8		
MMHQ	-3		
Kings Health Questionnaire	5-6 points for small effect 10-15 points for medium effect	Men and women OAB / lower UI dysfunction	Kelleher 2004
Incontinence Questionnaire-Urinary Incontinence Short Form (ICIQ-UI SF)	4	Women with SUI	Lim 2019
Incontinence Questionnaire-Lower Urinary Tract Symptoms Quality of Life (ICIQ-LUTSqol)	6		
PFDI-20	50	Chinese women with symptomatic pelvic floor dysfunction	Ma 2019
PISQ	6	Women with OAB, UI or prolapse	Mamik 2014
I-QOL	2 to 5 %	Incontinent women	Patrick 1999

Tool	MID	Population	Source
satisfactory sexual events (SSEs) per week	0.04 to 0.46 (range)	Women with hypoactive sexual desire disorder (HSDD).	Symonds 2007
PFDI-20	48	POP (both surgical and non-surgical patients)	Teig 2017
PFDI-7	47		
PFDI-20	13.5	Women with relatively mild PF symptoms	Wiegersma 2017
I-QOL (within treatment)	6.3	Predominant SUI	Yalcin 2006
I-QOL (between treatment)	2.5		
incontinence episode frequency	50% reduction	SUI in women	Yalcin 2010
frequency of faecal incontinence	50% reduction	Urge predominant FI in women	Noelting 2016
Incontinence Modular Questionnaire–Urinary Incontinence Short Form (ICIQ-UI SF)	2.52 (SD 2.56)	Women with SUI	Nystrom 2015
Lower Urinary Tract Symptoms Quality of Life (ICIQ-LUTSqol)	3.71 (SD 4.95)		
UDI-6	11	SUI	Roman 2016
IIQ-7	16		
Incontinence Quality of Life (I-QOL)	4 to 11	Urinary incontinence due to neurogenic detrusor over activity	Schurch 2007
Incontinence Questionnaire-Urinary Incontinence Short Form (ICIQ-UI SF)	-5(at 12 months) -4 (at 24 months)	Women with predominant SUI	Sirls 2015
Michigan Incontinence Symptom Index (M-ISI)	4	Men and women with UI	Suskind 2014
Fecal Incontinence Quality of Life scale (FIQL)	0.4	Men and women with FI	't Hoen 2017
Fecal Incontinence Severity Index (FISI)	11.5		

APFQ - Australian Pelvic Floor Questionnaire; CRADI: ; CRAIQ ; EQ-5D: EuroQol 5 dimension quality of life measure; FI: faecal incontinence; FISI: fecal incontinence severity index; FIQL: fecal incontinence quality of life scale; FSAD: female sexual arousal disorder; FSDS-DAO: Female Sexual Distress Scale Desire/ Arousal/ Orgasm FSFI: Female Sexual Function Index; HRQoL: health related quality of life; HSDD: hypoactive sexual desire disorder; I-QOL: incontinence Quality of Life questionnaire; ICIQ-LUTSqol: Incontinence Questionnaire-Lower Urinary Tract Symptoms Quality of Life; ICIQ-UI SF: Incontinence Questionnaire-Urinary Incontinence Short Form; IIQ-7: incontinence impact questionnaire v7; M-ISI: Michigan Incontinence Symptom Index; MMHQ: Modified Manchester Health Questionnaire; OAB: overactive bladder; OAB-Q: overactive bladder questionnaire; OBSS: Overactive Bladder Symptom Score; ODS: obstructed defecation syndrome; PFD: pelvic floor dysfunction; PISQ: ;POP: pelvic organ prolapse; POPDI ; POPIQ ; PPBC: Patient Perception of Bladder Condition; PRAFAB-Q: protection, amount, frequency, adjustment & body questionnaire; SF-6D: short form 6 dimension general health measure; SUI: stress incontinence; UDI: urinary distress inventory; UI: urinary incontinence; UIQ: urinary impact questionnaire

Although there were a number of published MIDs, they could not always be used due to differences in the study populations or in the reporting of the data. In the absence

of usable published or accepted MIDs, the committee agreed to use the GRADE MIDs to assess imprecision. For dichotomous outcomes minimally important thresholds for a RR of 0.8 and 1.25 respectively were used as default MIDs in the guideline. The committee also chose to use 0.8 and 1.25 as the MIDs for ORs & HRs in the absence of published or accepted MIDs. ORs were predominantly used in the guideline for prognostic reviews and when Peto OR were indicated due to low event rates, at low event rates OR are mathematically similar to RR making the extrapolation appropriate. While no default MIDs exist for HR, the committee agreed for consistency to continue to use 0.8 and 1.25 for these outcomes.

If risk difference was used for meta-analysis, for example if the majority of studies had zero events in either arm, imprecision was assessed based on sample size using 200 and 400 as cut-offs for very serious and serious imprecision respectively. The committee used these numbers based on commonly used optimal information size thresholds.

The same thresholds were used as MIDs in the guideline for all dichotomous outcomes considered in intervention evidence reviews. For continuous outcomes MIDs are equal to half the median SD of the control groups at baseline (or at follow-up if the SD is not available a baseline).

Assessing publication bias in intervention reviews

Where 10 or more studies were included as part of a single meta-analysis, a funnel plot was produced to graphically assess the potential for publication bias. Where fewer than 10 studies were included for an outcome, the committee subjectively assessed the likelihood of publication bias based on factors such as the proportion of trials funded by industry and the propensity for publication bias in the topic area.

Prognostic studies

Adapted GRADE methodology for prognostic reviews

For prognostic reviews with evidence from comparative studies an adapted GRADE approach was used. As noted above, GRADE methodology is designed for intervention reviews but the quality assessment elements were adapted for prognostic reviews.

The evidence for each outcome in the prognostic reviews was examined separately for the quality elements listed and defined in Table 6. The criteria considered in the rating of these elements are discussed below. Each element was graded using the quality levels summarised in Table 3. Footnotes to GRADE tables were used to record reasons for grading a particular quality element as having ‘serious’ or ‘very serious’ quality issues. The ratings for each component were combined to obtain an overall assessment of quality for each outcome as described in Table 4.

Table 6: Adaptation of GRADE quality elements for prognostic reviews

Quality element	Description
Risk of bias (‘Study limitations’)	Limitations in study design and implementation may bias estimates and interpretation of the effect of the prognostic/risk factor. High risk of bias for the majority of the evidence reduces confidence in the estimated effect. Prognostic studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)

Quality element	Description
Inconsistency	This refers to unexplained heterogeneity between studies looking at the same prognostic/risk factor, resulting in wide variability in estimates of association (such as RRs or ORs), with little or no overlap in confidence intervals
Indirectness	This refers to any departure from inclusion criteria listed in the review protocol (such as differences in study populations or prognostic/risk factors), that may affect the generalisability of results
Imprecision	This occurs when a study has relatively few participants and also when the number of participants is too small for a multivariable analysis (as a rule of thumb, 10 participants are needed per variable). This was assessed by considering the confidence interval in relation to the point estimate for each outcome reported in the included studies

RR, relative risk; OR, odds ratio

Assessing risk of bias in prognostic reviews

The Quality in Prognosis Studies (QUIPS) tool developed by Hayden 2013 was used to assess risk of bias in studies included in prognostic reviews (see Appendix H in the Developing NICE guidelines: the manual). The risk of bias in each study was determined by assessing the following domains:

- selection bias
- attrition bias
- prognostic factor bias
- outcome measurement bias
- control for confounders
- appropriate statistical analysis.

For cross-sectional studies of pelvic floor dysfunction in women with long term co-existing conditions the Joanna Briggs Institute Appraisal Checklist for Cross Sectional Studies (see Appendix H in the Developing NICE guidelines: the manual).

Assessing inconsistency in prognostic reviews

Where multiple results were deemed appropriate to meta-analyse (that is, there was sufficient similarity between risk factor and outcome under investigation) inconsistency was assessed by visually inspecting forest plots and observing whether there was considerable heterogeneity in the results of the meta-analysis. This was assessed by calculating the I-squared statistic for the meta-analysis with an I-squared value of more than 50% indicating serious heterogeneity, and more than 80% indicating very serious heterogeneity. When serious or very serious heterogeneity was observed, possible reasons were explored and subgroup analyses were performed as pre-specified in the review protocol where possible.

When no plausible explanation for the heterogeneity could be found, the quality of the evidence was downgraded in GRADE for inconsistency.

Assessing indirectness in prognostic reviews

Indirectness in prognostic reviews was assessed by comparing the populations, prognostic factors and outcomes in the evidence to those defined in the review protocol.

Assessing imprecision and importance in prognostic reviews

Prognostic studies may have a variety of purposes, for example, establishing typical prognosis in a broad population, establishing the effect of patient characteristics on prognosis, and developing a prognostic model. While by convention MIDs relate to intervention effects, the committee agreed to use GRADE default MIDs for risk ratios as a starting point from which to assess whether the size of an outcome effect in a prognostic study would be large enough to be meaningful in practice.

Qualitative studies

GRADE-CERQual methodology for qualitative reviews

For qualitative reviews an adapted GRADE Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach (Lewin 2015) was used. In this approach the quality of evidence is considered according to themes in the evidence. The themes may have been identified in the primary studies or they may have been identified by considering the reports of a number of studies. Quality elements assessed using GRADE-CERQual are listed and defined in Table 7. Each element was graded using the levels of concern summarised in Table 8.

The ratings for each component were combined (as with other types of evidence) to obtain an overall assessment of quality for each theme as described in Table 9. 'Confidence' in this context refers to the extent to which the review finding is a reasonable representation of the phenomenon of interest set out in the protocol. Similar to other types of evidence all review findings start off with 'high confidence' and are rated down by one or more levels if there are concerns about any of the individual CERQual components. In line with advice from the CERQual developers, the overall assessment does not involve numerical scoring for each component but in order to ensure consistency across and between guidelines, the NGA established some guiding principles for overall ratings. For example, a review finding would not be downgraded (and therefore would be assessed with 'high' confidence) if all 4 components had 'no or very minor' concerns or 3 'no or very minor' and 1 'minor'. At the other extreme, a review finding would be downgraded 3 times (to 'very low') if at least 2 components had serious concerns or at least 3 had moderate concerns. A basic principle was that if any components had serious concerns then overall confidence in the review finding would be downgraded at least once (potentially more depending on the other ratings). Transparency about overall judgements is provided in the CERQual tables, including a brief reference to components for which there were concerns in the 'overall confidence' cell.

Table 7: Adaptation of GRADE quality elements for qualitative reviews

Quality element	Description
Risk of bias ('Methodological limitations')	Limitations in study design and implementation may bias interpretation of qualitative themes identified. High risk of bias for the majority of the evidence reduces confidence in review findings. Qualitative studies are not usually randomised and therefore would not be downgraded for study design from the outset (they start as high quality)

Quality element	Description
Relevance (or applicability) of evidence	This refers to the extent to which the evidence supporting the review findings is applicable to the context specified in the review question
Coherence of findings	This refers to the extent to which review findings are well grounded in data from the contributing primary studies and provide a credible explanation for patterns identified in the evidence
Adequacy of data (theme saturation or sufficiency)	This corresponds to a similar concept in primary qualitative research, that is, whether a theoretical point of theme saturation was achieved, at which point no further citations or observations would provide more insight or suggest a different interpretation of the particular theme. Individual studies that may have contributed to a theme or sub-theme may have been conducted in a manner that by design would have not reached theoretical saturation at an individual study level

Table 8: CERQual levels of concern (by quality element)

Level of concern	Definition
None or very minor concerns	Unlikely to reduce confidence in the review finding
Minor concerns	May reduce confidence in the review finding
Moderate concerns	Will probably reduce confidence in the review finding
Serious concerns	Very likely to reduce confidence in the review finding

Table 9: Overall confidence in the evidence in CERQual (by review finding)

Overall confidence level	Definition
High	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest
Moderate	It is likely that the review finding is a reasonable representation of the phenomenon of interest
Low	It is possible that the review finding is a reasonable representation of the phenomenon of interest
Very low	It is unclear whether the review finding is a reasonable representation of the phenomenon of interest

Assessing methodological limitations in qualitative reviews

Methodological limitations in qualitative studies were assessed using the Critical Appraisal Skills Programme (CASP) checklist for qualitative studies (see appendix H in Developing NICE guidelines: the manual). Overall methodological limitations were derived by assessing the methodological limitations across the 6 domains summarised in Table 10.

Table 10: Methodological limitations in qualitative studies

Aim and appropriateness of qualitative evidence	This domain assesses whether the aims and relevance of the study were described clearly and whether qualitative research methods were appropriate for investigating the research question
Rigour in study design or validity of theoretical approach	This domain assesses whether the study approach was documented clearly and whether it was based on a theoretical framework (such as ethnography or grounded theory). This does not necessarily mean that the framework has to be stated explicitly, but a detailed description ensuring transparency and reproducibility should be provided
Sample selection	This domain assesses the background, the procedure and reasons for the method of selecting participants. The assessment should include consideration of any relationship between the researcher and the participants, and how this might have influenced the findings
Data collection	This domain assesses the documentation of the method of data collection (in-depth interviews, semi-structured interviews, focus groups or observations). It also assesses who conducted any interviews, how long they lasted and where they took place
Data analysis	This domain assesses whether sufficient detail was documented for the analytical process and whether it was in accordance with the theoretical approach. For example, if a thematic analysis was used, the assessment would focus on the description of the approach used to generate themes. Consideration of data saturation would also form part of this assessment (it could be reported directly or it might be inferred from the citations documented that more themes could be found)
Results	This domain assesses any reasoning accompanying reporting of results (for example, whether a theoretical proposal or framework is provided)

Assessing relevance of evidence in qualitative reviews

Relevance (applicability) of findings in qualitative research is the equivalent of indirectness for quantitative outcomes, and refers to how closely the aims and context of studies contributing to a theme reflect the objectives outlined in the guideline review protocol.

Assessing coherence of findings in qualitative reviews

For qualitative research, a similar concept to inconsistency is coherence, which refers to the way findings within themes are described and whether they make sense. This concept was used in the quality assessment across studies for individual themes. This does not mean that contradictory evidence was automatically downgraded, but that it was highlighted and presented, and that reasoning was provided. Provided the themes, or components of themes, from individual studies fit into a theoretical framework, they do not necessarily have to reflect the same perspective. It should, however, be possible to explain these by differences in context (for example, the views of healthcare professionals might not be the same as those of family members, but they could contribute to the same overarching themes).

Assessing adequacy of data in qualitative reviews

Adequacy of data corresponds to the depth of evidence and whether sufficient quotations or observations were provided to underpin the findings. The complexity of the themes is also taken into account when assessing their adequacy.

Reviewing economic evidence

Inclusion and exclusion of economic studies

A global economic literature search was undertaken for pelvic floor dysfunction. This covered all review questions, which were reported in 18 evidence reports in this guideline. Titles and abstracts of articles identified through the economic literature search were independently assessed for inclusion using the predefined eligibility criteria listed in **Error! Reference source not found.**.

Table 11: Inclusion and exclusion criteria for systematic reviews of economic evaluations

Inclusion criteria
Intervention or comparators in accordance with the guideline scope
Study population in accordance with the guideline scope
Full economic evaluations (cost-utility, cost effectiveness, cost-benefit or cost-consequence analyses) assessing both costs and outcomes associated with interventions of interest
Studies from Organisation for Economic Co-operation and Development (OECD) countries were included, as the aim of the review was to identify economic information transferable to the UK context
Exclusion criteria
Abstracts containing insufficient methodological details
Cost-of-illness type studies
Conference abstracts

Once the screening of titles and abstracts was completed, full-text copies of potentially relevant articles were requested for detailed assessment. Inclusion and exclusion criteria were applied to articles obtained as full-text copies.

Details of the economic evidence study selection for each question, list of excluded studies, economic evidence tables, the results of quality assessment of economic evidence (see below) and health economic evidence profiles are presented in

appendices G, K, H and I of the evidence report. Existing economic evidence considered in the guideline is provided in the respective evidence reviews.

Appraising the quality of economic evidence

The quality of economic evidence was assessed using the economic evaluations checklist specified in Developing NICE guidelines: the manual (NICE 2018).

Economic modelling

The aims of the economic input to the guideline were to inform the guideline committee of potential economic issues to ensure that recommendations represented a cost effective use of healthcare resources. Economic evaluations aim to integrate data on healthcare benefits (ideally in terms of quality-adjusted life-years; QALYs) with the costs of different options. In addition, the economic input aimed to identify areas of high resource impact; these are recommendations which (while cost effective) might have a large impact on NHS finances and so need special attention.

The guideline committee prioritised the following review questions for economic modelling where it was thought that economic considerations would be particularly important in formulating recommendations.

- What is the effectiveness of pelvic floor muscle training for preventing pelvic floor dysfunction?
- What is the effectiveness of physical devices (including support garments, pessaries and dilators) for improving symptoms of pelvic floor dysfunction?

The methods and results of the de novo economic analyses are reported in Appendix J of the relevant evidence reports. When new economic analysis was not prioritised, the committee made a qualitative judgement regarding cost effectiveness by considering expected differences in resource and cost use between options, alongside clinical effectiveness evidence identified from the clinical evidence review.

Cost effectiveness criteria

NICE's report Our principles sets out the principles that committees should consider when judging whether an intervention offers good value for money. In general, an intervention was considered to be cost effective if any of the following criteria applied (provided that the estimate was considered plausible):

- the intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more effective compared with all the other relevant alternative strategies)
- the intervention cost less than £20,000 per QALY gained compared with the next best strategy
- the intervention provided important benefits at an acceptable additional cost when compared with the next best strategy.

The committee's considerations of cost effectiveness are discussed explicitly under the heading 'Cost effectiveness and resource use' in the relevant evidence reviews.

Developing recommendations

Guideline recommendations

Recommendations were drafted on the basis of the committee's interpretation of the available evidence, taking account of the balance of benefits, harms and costs between different courses of action. When effectiveness and economic evidence was of poor quality, conflicting or absent, the committee drafted recommendations based on their expert opinion. The considerations for making consensus-based recommendations include the balance between potential benefits and harms, the economic costs or implications compared with the economic benefits, current practices, recommendations made in other relevant guidelines, person's preferences and equality issues.

The main considerations specific to each recommendation are outlined under the heading 'The committee's discussion of the evidence' within each evidence review.

For further details refer to Developing NICE guidelines: the manual.

Research recommendations

When areas were identified for which evidence was lacking, the committee considered making recommendations for future research. For further details refer to Developing NICE guidelines: the manual and NICE's Research recommendations process and methods guide.

Validation process

This guideline was subject to a 6-week public consultation and feedback process. All comments received from registered stakeholders were responded to in writing and posted on the NICE website at publication. For further details refer to Developing NICE guidelines: the manual.

Updating the guideline

Following publication, NICE will undertake a surveillance review to determine whether the evidence base has progressed sufficiently to consider altering the guideline recommendations and warrant an update. For further details refer to Developing NICE guidelines: the manual.

Funding

The NGA was commissioned by NICE to develop this guideline.

References

Abrams 2017

Abrams P, Kelleher C, Staskin D, Kay R, Martan A, Mincik I. Combination treatment with mirabegron and solifenacin in patients with overactive bladder: exploratory responder analyses of efficacy and evaluation of patient-reported outcomes from a randomized, double-blind, factorial, dose-ranging, Phase II study (SYMPHONY). World Journal of Urology, 35, 827-838, 2017.

Althof 2019

Althof S, Derogatis LR, Greenberg S, Clayton AH, Jordan R, Lucas J, et al. Responder Analyses from a Phase 2b Dose-Ranging Study of Bremelanotide. Journal of Sexual Medicine.16,1226-1235, 2019.

Baessler 2019

Baessler K, Mowat A, Maher CF. The minimal important difference of the Australian Pelvic Floor Questionnaire. International urogynecology journal and pelvic floor dysfunction. 30, 115-122, 2019.

Barber 2009

Barber MD, Spino C, Janz NK, Brubaker L, Nygaard I, Nager CW, et al. The minimum important differences for the urinary scales of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire. American Journal of Obstetrics and Gynecology.;200, 580.e1-7, 2009.

Bols 2010

Bols EM, Hendriks EJ, Deutekom M, Berghmans BC, Baeten CG, de Bie RA. Inconclusive psychometric properties of the Vaizey score in fecally incontinent patients: a prospective cohort study. Neurourolgy and Urodynamics. 29, 370-377, 2010.

Bradburn 2007

Bradburn, M. J., Deeks, J. J., Berlin, J. A., & Localio, A. R. Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events. Statistics in Medicine, 26, 53–77, 2007.

Caetano 2018

Caetano AC, Dias S, Santa-Cruz A, Rolanda C. Renzi Score for Obstructed Defecation Syndrome - Validation of the Portuguese Version According to the Cosmin Checklist. Archives of Gastroenterology. 55, 55-60, 2018.

Chan 2013

Chan SS, Cheung RY, Lai BP, Lee LL, Choy KW, Chung TK. Responsiveness of the Pelvic Floor Distress Inventory and Pelvic Floor Impact Questionnaire in women

undergoing treatment for pelvic floor disorders. International urogynecology journal and pelvic floor dysfunction. 24, 213-221, 2013.

Coyne 2006

Coyne KS, Matza LS, Thompson CL, Kopp ZS, Khullar V. Determining the importance of change in the overactive bladder questionnaire. Journal of Urology. 176, 627-632, 2006.

Dyer 2011

Dyer KY, Xu Y, Brubaker L, Nygaard I, Markland A, Rahn D. Minimum important difference for validated instruments in women with urge incontinence. Neurourolgy and Urodynamics. 30, 1319-1324, 2011.

Gotoh 2011

Gotoh M, Homma Y, Yokoyama O, Nishizawa O. Responsiveness and minimal clinically important change in overactive bladder symptom score. Urology. 78, 768-773, 2011.

Halme 2015

Halme AS, Fritel X, Benedetti A, Eng K, Tannenbaum C. Implications of the minimal clinically important difference for health-related quality-of-life outcomes: a comparison of sample size requirements for an incontinence treatment trial. Value Health. 18, 292-298, 2015.

Harvie 2019

Harvie HS, Honeycutt AA, Neuwahl SJ, Barber MD, Richter HE, Visco AG, et al. Responsiveness and minimally important difference of SF-6D and EQ-5D utility scores for the treatment of pelvic organ prolapse. American Journal of Obstetrics and Gynecology. 220, 265.e1-.e11, 2019.

Hayden 2013

Jill A. Hayden, Danielle A. van der Windt, Jennifer L. Cartwright, Pierre Côté, Claire Bombardier. Assessing Bias in Studies of Prognostic Factors. Annals of Internal Medicine. 158, 280–286, 2013.

Hendriks 2007

Hendriks EJ, Bernards AT, Berghmans BC, de Bie RA. The psychometric properties of the PRAFAB-questionnaire: a brief assessment questionnaire to evaluate severity of urinary incontinence in women. Neurourolgy and Urodynamics. 26, 998-1007, 2007.

Hendriks 2008

Hendriks EJ, Bernards AT, de Bie RA, de Vet HC. The minimal important change of the PRAFAB questionnaire in women with stress urinary incontinence: results from a prospective cohort study. Neurourolgy and Urodynamics. 27, 379-387, 2008.

Higgins 2011

Higgins JPT, Green S (editors) (2011) Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated 2019] The Cochrane Collaboration. Available from www.handbook.cochrane.org (accessed 15 April 2021)

Homma 2006

Homma Y, Koyama N. Minimal clinically important change in urinary incontinence detected by a quality of life assessment tool in overactive bladder syndrome with urge incontinence. Neurourolgy and Urodynamics. 25, 228-235, 2006.

Jelovsek 2014

Jelovsek JE, Chen Z, Markland AD, Brubaker L, Dyer KY, Meikle S, et al. Minimum important differences for scales assessing symptom severity and quality of life in patients with fecal incontinence. Female pelvic medicine. 20, 342-348, 2014.

Kelleher 2004

Kelleher CJ, Pleil AM, Reese PR, Burgess SM, Brodish PH. How much is enough and who says so? The case of the King's Health Questionnaire and overactive bladder. BJOG: An International Journal of Obstetrics and Gynaecology. 111, 605-612, 2004.

Lewin 2015

Lewin S, Glenton C, Munthe-Kaas H. (2015) Using qualitative evidence in decision making for health and social interventions: an approach to assess confidence in findings from qualitative evidence syntheses (GRADE-CERQual). PLoS Med 12(10), e1001895

Lim 2019

Lim R, Liong ML, Lim KK, Leong WS, Yuen KH. The Minimum Clinically Important Difference of the International Consultation on Incontinence Questionnaires (ICIQ-UI SF and ICIQ-LUTSqol). Urology. 10, 2019

Ma 2019

Ma Y, Xu T, Zhang Y, Mao M, Kang J, Zhu L. Validation of the Chinese version of the Pelvic Floor Distress Inventory-20 (PFDI-20) according to the COSMIN checklist. International urogynecology journal and pelvic floor dysfunction. 30, 1127-1139, 2019.

Mamik 2014

Mamik MM, Rogers RG, Qualls CR, Morrow JD. The minimum important difference for the Pelvic Organ Prolapse-Urinary Incontinence Sexual Function Questionnaire. International urogynecology journal and pelvic floor dysfunction. 25, 1321-1326, 2014.

McGowan 2016

McGowan J, Sampson M, Salzwedel DM et al. (2016) PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. Journal of Clinical Epidemiology 75: 40–6

NICE 2018

National Institute for Health and Care Excellence (NICE) (2014) NICE Policy on conflicts of interest (updated 2017). Available from https://www.nice.org.uk/Media/Default/About/Who-we-are/Policies-and-procedures/declaration-of-interests-policy.pdf (accessed 15 April 2021)

Noelting 2016

Noelting J, Zinsmeister AR, Bharucha AE. Validating endpoints for therapeutic trials in fecal incontinence. Neurogastroenterology and Motility. 28, 1148-1156, 2016.

Nystrom 2015

Nystrom E, Sjostrom M, Stenlund H, Samuelsson E. ICIQ symptom and quality of life instruments measure clinically relevant improvements in women with stress urinary incontinence. Neurourolgy and Urodynamics. 34, 747-751, 2015.

Patrick 1999

Patrick DL, Martin ML, Bushnell DM, Yalcin I, Wagner TH, Buesching DP. Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). Urology. 53, 71-76, 1999.

Rao 2016

Rao SS. Endpoints for therapeutic interventions in faecal incontinence: small step or game changer. Neurogastroenterology and Motility. 28,1123-1133, 2016.

Roman 2016

Roman JD. Subjective outcome of 166 tension-free vaginal tape procedures performed by a single surgeon: the Braemar experience. Australian and New Zealand Journal of Obstetrics and Gynaecology. 56, 503-507, 2016.

Santesso 2016

Santesso N, Carrasco-Labra A, Langendam M et al. (2016) Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. Journal of clinical epidemiology 74, 28-39

Schurch 2007

Schurch B, Denys P, Kozma CM, Reese PR, Slaton T, Barron R. Reliability and validity of the Incontinence Quality of Life questionnaire in patients with neurogenic urinary incontinence. Archives of Physical Medicine and Rehabilitation. 88, 646-652, 2007.

Sirls 2015

Sirls LT, Tennstedt S, Brubaker L, Kim HY, Nygaard I, Rahn DD, et al. The minimum important difference for the International Consultation on Incontinence Questionnaire-Urinary Incontinence Short Form in women with stress urinary incontinence. Neurourolgy and Urodynamics. 34, 183-187, 2015.

Suskind 2014

Suskind AM, Dunn RL, Morgan DM, DeLancey JO, McGuire EJ, Wei JT. The Michigan Incontinence Symptom Index (M-ISI): a clinical measure for type, severity, and bother related to urinary incontinence. Neurourolgy and Urodynamics. 33, 1128-1134, 2014.

Symonds 2007

Symonds T, Spino C, Sisson M, Soni P, Martin M, Gunter L, et al. Methods to determine the minimum important difference for a sexual event diary used by postmenopausal women with hypoactive sexual desire disorder. Journal of Sexual Medicine. 4, 1328-1335, 2007.

t Hoen 2017

t Hoen LA, Utomo E, Schouten WR, Blok BF, Korfage IJ. The fecal incontinence quality of life scale (FIQL) and fecal incontinence severity index (FISI): Validation of the Dutch versions. Neurourolgy and Urodynamics. 36, 710-715, 2017.

Teig 2017

Teig CJ, Grotle M, Bond MJ, Prinsen CAC, Engh MAE, Cvancarova MS, et al. Norwegian translation, and validation, of the Pelvic Floor Distress Inventory (PFDI-20) and the Pelvic Floor Impact Questionnaire (PFIQ-7). International urogynecology journal and pelvic floor dysfunction. 28,1005-1017, 2017.

Wiegersma 2017

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. American Journal of Obstetrics and Gynecology. 216, 397.e1-.e7, 2017.

Yalcin 2004

Yalcin I, Patrick D, Summers K, Kinchen K, Bump R. The minimum clinically important difference in incontinence quality of life questionnaire (I-QOL) total and subscale scores in women with stress urinary incontinence (SUI). Neurourology and urodynamics. 23, 568-570, 2004.

Yalcin 2006

Yalcin I, Patrick DL, Summers K, Kinchen K, Bump RC. Minimal clinically important differences in Incontinence Quality-of-Life scores in stress urinary incontinence. Urology. 67, 1304-1308, 2006.

Yalcin 2010

Yalcin I, Peng G, Viktrup L, Bump RC. Reductions in stress urinary incontinence episodes: what is clinically important for women? Neurourolgy and Urodynamics. 29, 344-347, 2010.