Pelvic Organ Prolapse
Panman, Chantal; Wiegersma, Marian

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 10
MINIMAL IMPORTANT CHANGE IN THE
PELVIC FLOOR DISTRESS INVENTORY-20
AMONG WOMEN OPTING FOR CONSERVATIVE
PROLAPSE TREATMENT

Marian Wiegersma, Chantal M.C.R. Panman, Marjolein Y. Berger,
Henrica C.W. de Vet, Boudewijn J. Kollen, Janny H. Dekker

ABSTRACT

Background
The PFDI-20 is used to evaluate symptoms and treatment effects in women with pelvic floor disorders. To interpret changes in the scores of this inventory, information is needed on what patients and clinicians perceive as the minimal important (meaningful) change. Although this change in the inventory score has previously been investigated in women undergoing pelvic floor surgery, the results could not be generalized to women with milder symptoms (i.e., lower scores), who often require only conservative treatment.

Objective
We aimed to estimate the minimal important change in the PFDI-20 that was needed to demonstrate clinical improvement in women qualifying for conservative pelvic floor treatment.

Study design
The data of 214 women aged ≥55 years were used. All participants were from two randomized controlled trials comparing conservative prolapse treatments in primary care in the Netherlands. The degree of prolapse was assessed using the POP-Q system and participants completed the PFDI-20 at baseline and 12 months, as well as a global perception of improvement question at 12 months. To assess both the patient perspective and the clinical perspective, two anchors were assessed: 1) the global perception of improvement was considered the anchor for the patients’ perspective, and 2) the difference in the degree of prolapse was considered the anchor for the clinical perspective. Provided that the anchors were correlated by at least 0.3 to the PFDI-20 change scores, we estimated the minimal important change as follows: 1) the optimal cutoff-point of the ROC curve discriminating between women with and without improvement in the global perception of improvement scale, and 2) the mean PFDI-20 change score of participants who improved one assessment stage. We then calculated the smallest detectable change to check whether the minimal important change was larger than the measurement error of the questionnaire.

Results
Using the global perception of improvement as the anchor, we found a minimal important change for improvement of 13.5 points (95% CI 6.2 to 20.9). The POP-Q change scores correlated poorly to the PFDI-20 change scores and could therefore not be used as an anchor. The smallest detectable change at the group level was 5.5
points. Thus, the minimal important change was larger than the smallest detectable change at the group level.

**Conclusion**

In women with relatively mild pelvic floor symptoms, an improvement of 13.5 points (or a 23% reduction) in the PFDI-20 score can be considered clinically relevant. This minimal important change can be used for clinical trial planning and evaluation of treatment effects in women considered suitable for conservative treatment.
INTRODUCTION

Patient-reported outcomes are commonly used to evaluate symptoms and treatment effects in research and clinical practice. The PFDI-20 is a recommended questionnaire for use when evaluating the degree to which pelvic floor symptoms cause distress. Although it has been shown to have good validity, reliability, and responsiveness for this purpose, information is needed on what patients (and/or clinicians) perceive as a meaningful difference when interpreting the changes in PFDI-20 scores.

To determine whether a statistically significant change is also clinically relevant, Jaeschke et al. introduced the concept of the minimal clinically important difference, also termed the minimal important change (MIC). They defined the MIC as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive costs, a change in the patient’s management”. Any change larger than the MIC can then be considered clinically relevant. The MIC can be determined from many different perspectives, including that of the patient, the clinician, the researcher, the consumer, or even of society.

Various methods have been proposed to determine the MIC. Anchor-based methods compare changes in patient-reported outcome scores with other clinical changes or results, while distribution-based approaches rely on statistical distributions of the results. Although there is no gold standard for determining the MIC, most authorities recommend using anchor-based methods when applying it to the various relevant perspectives (e.g. patient-rated, clinician-rated, and disease-specific measures).

Only a few studies have attempted to estimate the MIC for the PFDI-20, but these were mostly among women undergoing prolapse or incontinence surgery who had relatively high baseline scores (94–121 points), and were all performed in tertiary urogynecology units. These resulted in anchor-based MICs for improvement ranging from 23 to 45 points (24%–47%), but were in a highly selected patient group. By contrast, women opting for conservative treatment generally experience less severe symptoms than women preferring surgical treatment. For example, studies evaluating conservative prolapse treatments have reported average PFDI-20 baseline scores of approximately 60 points, to further confound matters, other research has shown that the MIC depends on the baseline score, with evidence that patients with more severe baseline symptoms seem to require greater improvements to consider them clinically relevant (i.e., a larger MIC) than patients with less severe initial symptoms. Consequently, the MICs established in tertiary care populations probably are unsuitable for use in evaluating conservative treatments in women with milder symptoms of pelvic floor dysfunction.
The aim of his study was to estimate the MIC for the PFDI-20 among women qualifying for conservative treatment of pelvic floor disorders.

METHODS

Participants and setting
This analysis was conducted with data of the "Pelvic Organ Prolapse in Primary Care: Effects of Pelvic Floor Muscle Training and Pessary Treatment Study" (POPPS), for which the design and primary outcomes have been published. The POPPS comprised two randomized controlled trials, with participants (women aged ≥55 years with symptomatic prolapse) recruited from 20 primary care practices. In the first trial, PFMT was compared with watchful waiting in women with a prolapse above the hymen. In the second trial, PFMT was compared with pessary treatment in women with a prolapse at or beyond the hymen. The trials were approved by the Medical Ethics Committee of the University Medical Center Groningen (METc2009.215) and were registered in the Dutch Trial Register (www.trialregister.nl, identifier NTR 2047). All participants provided written informed consent. The present study included participants of both trials and used data collected at baseline and at the 12-month follow-up assessment.

Measures
Participants completed the PFDI-20 questionnaire at baseline and after 12-months’ follow-up. The PFDI-20 score ranges from 0 to 300, with higher scores indicating higher symptom distress. PFDI-20 change scores were calculated by subtracting the follow-up score from the baseline score, such that a negative change score represented symptoms getting worse and a positive change score represented symptoms getting better. The degree of prolapse was measured according to the POP-Q system at baseline and after 12 months. The POP-Q stage (0–4) was assessed for each compartment (anterior vaginal wall, posterior vaginal wall, and uterus or vaginal vault), with the overall POP-Q stage being equal to the POP-Q stage of the most severely prolapsed compartment. A higher POP-Q stage represented more severe prolapse. The change score for the POP-Q stage was calculated by subtracting the overall POP-Q stage at follow-up from the overall POP-Q stage at baseline. This led to a POP-Q change score ranging from -3 (all participants had at least stage 1 prolapse at baseline) to +4. We also assessed the change of prolapse using a continuous measure of anatomic support. We calculated change scores for the degree of prolapse of the anterior vaginal wall (POP-Q point Ba), the posterior vaginal wall (POP-Q point Bp), and the uterus or
vaginal vault (POP-Q point C) by subtracting the follow-up value from the baseline value. This led to changes scores ranging from -2 to +3 centimetres for Ba, -3 to +2 centimetres for Bp and -7 to +4 centimetres for C. In each of these anatomical change scores, a negative score represented a deteriorating prolapse, zero represented no change, and a positive score represented an improving prolapse.

After 12 months, participants were also asked to rate their global perception of improvement (GPI) since baseline, according to the following question: “Overall, do you feel that your symptoms are: much worse (-2), worse (-1), about the same (0), better (+1), or much better (+2).”

### Statistical methods

The MIC was determined by ROC analysis and visualized using an anchor-based distribution plot. In an attempt to assess both the patient and the clinical perspectives, two anchors were assessed for eligibility. The GPI was considered the anchor for the patients’ perspective, and the difference in the degree of prolapse on physical examination was considered the anchor for the clinical perspective. The anchors were only considered suitable for further analysis if they correlated with the PFDI-20 change score by at least 0.3 (Spearman’s ρ). Statistical analyses were performed using IBM SPSS for Windows, Version 23.0 (IBM Corp., Armonk, NY, USA). Stata/SE 14 was used to estimate a confidence interval for the MIC based on 1000 bootstrap replicates.

#### Patients’ perspective

Provided that the GPI and the PFDI-20 change score were sufficiently correlated, the participants were divided into two groups: those who reported symptom improvement (GPI categories “better” and “much better”) and those who did not (GPI categories “about the same,” “worse,” and “much worse”). The MIC, or the optimal ROC cut-off point, was defined as the value for which the sum of the false positive and false negative ([1-sensitivity] + [1-specificity]) percentages was the smallest. This value was taken to represent the PFDI-20 change score that best discriminated between participants with and without clinical improvement according to the GPI.

#### Clinical perspective

Provided that the POP-Q change score and the PFDI-20 change score were sufficiently correlated, participants were divided into two groups: those who improved one or more POP-Q stage(s), and those who did not (remained stable or deteriorated), with the MIC defined as the optimal cut-off point on the ROC curve. This value was taken as the PFDI-20 change score that allowed for the best discrimination between participants with and without improvement of one or more POP-Q stage.
For the continuous measure of anatomic support, provided that the Ba, Bp, and C change scores and the PFDI-20 change score were sufficiently correlated, participants were divided into two groups: those who improved two or more centimetres in Ba, Bp, or C, and those who did not (remained stable or deteriorated). The MIC was defined as the optimal cut-off point on the ROC curve. This value was taken as the PFDI-20 change score that allowed for the best discrimination between participants with and without improvement of two or more centimetres in the degree of prolapse.

Smallest detectable change

Not every change in a questionnaire score can be considered real. Small changes in questionnaire scores could reflect measurement error. In this context, the smallest detectable change (SDC) is the smallest measurement change that can be detected by an instrument beyond measurement error. Measurement error can be reduced by measuring in groups of patients and calculating average scores. This allows for the SDC on group level to be smaller than the SDC on the individual level (i.e., greater change needs to be detected beyond measurement error in an individual patient compared with a group of patients).

We compared the estimated MIC to the SDC at both the individual and group levels using the standard error of measurement (SEM). The SDC at the individual level was calculated as \([1.96 \times SEM \times \sqrt{2}]\) while the SDC at the group level was calculated as \([(1.96 \times SEM \times \sqrt{2}) / \sqrt{n}]\). The SEM was derived from the intraclass correlation coefficient (ICC) for agreement of the relationship between the PFDI-20 scores at baseline and follow-up. The ICC was used as a reliability parameter and calculated as a ratio of variances. The variance components were obtained by analysis of variance and represented different sources of within and between variances, as well as residual variance. This latter variance was used to calculate the SEM, as \([\sqrt{(\sigma^2_{\text{observation}} + \sigma^2_{\text{residual}})}]\), in the subgroup of participants who reported their symptoms to be “about the same” on the GPI.

RESULTS

We obtained PFDI-20 scores at baseline and 12 months, plus the GPI scores at 12 months, for 214 women. Of these, 74 were randomized to watchful waiting, 110 were randomized to PFMT, and 39 were randomized to pessary treatment. The characteristics of the study population are shown in Table 1.

The PFDI-20 change scores and GPI scores were moderately but sufficiently correlated, with a Spearman’s \(\rho\) of 0.35. The mean PFDI-20 change scores per GPI category are shown in Table 2, and the distribution of the PFDI-20 change scores
are shown within each GPI category in Figure 1. None of the participants reported their symptoms to be “much worse”. The ROC curve analysis, which used the PFDI-20 change score cut-off points to discriminate between women with and without improvement, is shown in Figure 2. The area under the ROC curve was 0.67, with a significant 95% CI of 0.59 to 0.74 (p < 0.001). The MIC for improvement, defined as the optimal ROC cut-off point, was 13.5 points (95% CI 6.2 to 20.9). This cut-off point had a sensitivity of 56% and a specificity of 75%, meaning that 56% of the participants who reported symptom improvement had a PFDI-20 change score of 13.5 points or higher, while 75% of the participants who did not report symptom improvement had a PFDI-20 change score smaller than 13.5 points. The SDC at the individual level was 58.1 points, while the SDC at the group level was 5.5 points. Figure 3 shows the anchor-based MIC distribution plot, and gives a visual representation of the distributions of the PFDI-20 change scores for participants who reported symptom improvement (GPI categories “better” and “much better”) and for those who did not (GPI categories “the same” and “worse”), combined with the MIC estimate.

POP-Q measurements were available for 198 participants (93%). Of these, 31 (16%)

### TABLE 1 CHARACTERISTICS OF THE STUDY POPULATION

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), median (IQR)</td>
<td>63.1 (58.7–68.4)</td>
</tr>
<tr>
<td>BMI (kg/m²), mean ± SD</td>
<td>26.7 ± 4.2</td>
</tr>
<tr>
<td>Postmenopausal, n (%)</td>
<td>210 (98)</td>
</tr>
<tr>
<td>Parity, n (%)</td>
<td></td>
</tr>
<tr>
<td>No children</td>
<td>5 (2)</td>
</tr>
<tr>
<td>1 child</td>
<td>12 (6)</td>
</tr>
<tr>
<td>2 children</td>
<td>109 (51)</td>
</tr>
<tr>
<td>≥3 children</td>
<td>88 (41)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
</tr>
<tr>
<td>Primary education</td>
<td>10 (5)</td>
</tr>
<tr>
<td>Lower education</td>
<td>74 (35)</td>
</tr>
<tr>
<td>Intermediate education</td>
<td>62 (29)</td>
</tr>
<tr>
<td>Higher education</td>
<td>68 (32)</td>
</tr>
<tr>
<td>Hysterectomy, n (%)</td>
<td>42 (20)</td>
</tr>
<tr>
<td>Other pelvic floor surgery, n (%)</td>
<td>6 (3)</td>
</tr>
<tr>
<td>PFDI-20 score, mean ± SD</td>
<td>59.5 ± 36.6</td>
</tr>
<tr>
<td>POP-Q stage at baseline, n (%)</td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>77 (36)</td>
</tr>
<tr>
<td>Stage 2</td>
<td>122 (57)</td>
</tr>
<tr>
<td>Stage 3</td>
<td>15 (7)</td>
</tr>
<tr>
<td>Stage 4</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>
TABLE 2: PFDI-20 CHANGE SCORES FOR EACH GPI CATEGORY

<table>
<thead>
<tr>
<th>GPI</th>
<th>n</th>
<th>PFDI-20 change score, mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Much better</td>
<td>13</td>
<td>40.5 ± 35.3</td>
</tr>
<tr>
<td>Better</td>
<td>65</td>
<td>14.8 ± 29.8</td>
</tr>
<tr>
<td>About the same</td>
<td>110</td>
<td>6.0 ± 22.2</td>
</tr>
<tr>
<td>Worse</td>
<td>26</td>
<td>-9.3 ± 18.5</td>
</tr>
<tr>
<td>Much worse</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

FIGURE 1: DISTRIBUTION OF THE PFDI-20 CHANGE SCORE WITHIN EACH GPI CATEGORY

Positive PFDI-20 change scores represent symptoms getting better, while negative PFDI-20 change scores represent symptoms getting worse. Boxes enclose medians and the first and third quartiles. Circles are outliers.
showed an improvement of one POP-Q stage, 27 (14%) showed deterioration of one POP-Q stage, and 140 (71%) remained stable. The POP-Q change score was very poorly correlated with the PFDI-20 change score (Spearman’s ρ = 0.001) and could therefore not be used as an anchor to determine the MIC. In total, 46 participants (23%) showed an improvement of two or more centimetres in Ba, Bp, and/or C. The change scores for Ba, Bp, and C were also poorly correlated with the PFDI-20 change score (Spearman’s ρ = 0.079 for Ba, 0.033 for Bp, and -0.031 for C) and could therefore not be used as an anchor to determine the MIC either.

**COMMENT**

**Principal findings**
This study used an anchor-based approach to estimate the MIC for the PFDI-20 in women qualifying for conservative prolapse treatment. This resulted in a MIC for improvement of 13.5 points (95% CI 6.2 to 20.9), which equates to a 23% reduction in the PFDI-20 score from baseline.
FIGURE 3 ANCHOR-BASED MIC DISTRIBUTION PLOT OF THE PFDI-20 CHANGE SCORES FOR PARTICIPANTS WHO DID AND DID NOT REPORT SYMPTOM IMPROVEMENT

A: participants correctly classified as “improved” by the MIC (56%); B: participants misclassified as “improved” by the MIC (25%); C: participants misclassified as “not improved” by the MIC (44%); D: participants correctly classified as “not improved” by the MIC (75%).

$\Delta$PFDI-20 = Pelvic Floor Distress Inventory-20 change score, with a positive change score representing symptom improvement, and a negative change score representing symptom deterioration.
Strengths and limitations
To the best of our knowledge, this is the first study to estimate the MIC for the PFDI-20 in women with relatively mild pelvic floor symptoms, and should be useful when evaluating response to conservative prolapse treatments. Nevertheless, there are some methodological issues that should be considered. First, we were only able to determine the MIC from the patient’s perspective, despite intending to include the clinical perspective, because the anatomical change scores were only poorly correlated to the PFDI-20 change score, making these unsuitable for use as an anchor. However, because the PFDI-20 is a patient-reported outcome measure, which is intended to provide details of the patients’ perspective on the impact of pelvic floor symptoms and its treatment, we contend that the patients’ perspective is more important.
Second, all women in our study were aged ≥ 55 years and most women were of Caucasian origin, therefore the results of this study should be interpreted with caution when applied to younger women, or to women of other ethnic origins. Third, although the anchor-based method we used is the generally accepted and most widely used method, there are some limitations to the approach. Such methods rely on the GPI, which is susceptible to recall-bias, with some patients finding it difficult to remember the degree of bother they experienced at the start of the study. Also, recall has been shown to be influenced by recent events and by the patients’ current mood state.安娜
Another problem is the definition of what constitutes an important improvement. There is presently no consensus as to what the best cut-off is on the GPI when determining the MIC, and the choice between the GPI categories “a little better,” “better,” and “much better” is rather arbitrary. The GPI asks patients to rate how much change they have experienced, but it might be more useful to know if they feel that the change is sufficient and if the outcome is satisfactory. A final limitation of anchor-based approaches is that they do not take into account the measurement precision of an instrument.安娜 However, we accounted for this by comparing the MIC with the SDC to ensure that the MIC did not lie within the measurement error of the PFDI-20. The MIC can also be determined by distribution-based methods, for example using half a standard deviation of the baseline score. In our study, this would have resulted in a MIC of 18.3 points (Table 1), which is slightly larger than the MIC we found using an anchor-based method.

Interpretation
As expected, the MIC in this study was lower than that reported in previous studies on this subject.安娜 Several studies have shown that patients with more severe symptoms at baseline may require a greater change in their symptom score to consider that change clinically relevant (i.e., have a larger MIO).安娜 The MIC found by Utomo et
al. was the closest to ours, showing that a MIC of 23 points (24%) was needed to show improvement in a group of 67 women with an average PFDI-20 baseline score of 94 points who were treated conservatively, pharmaceutically, and/or surgically. Two other studies have estimated the MIC in women undergoing prolapse surgery, but based on higher average PFDI-20 baseline scores (97–121 points), they found higher MICs, with levels varying between 37 and 45 points (32%–47%).

MIC values can be used to compare treatments between groups of patients, such as in randomized controlled trials. For a treatment effect to be considered clinically relevant in a randomized controlled trial, a statistically significant improvement in the PFDI-20 score should exceed the MIC. However, this may not apply to each individual patient. Even groups with clinically meaningless mean changes in questionnaire scores likely contain individual patients whose improvement is noteworthy for that individual. Also, relatively modest improvement at the individual level may be clinically important when considered at the group level. In clinical practice, one might want to use the MIC to interpret the change score of an individual patient. However, when interpreting MIC values at the individual level, three types of uncertainty should be considered.

First, one has to take the sampling variation into account, which is depicted by the 95% CI around the MIC value. When the 95% CI is large, the uncertainty about whether the estimated MIC is the “true” MIC increases; in our study, the 95% CI around the MIC was 6.2 to 20.9.

Second, one must assess the extent to which the MIC value determined in a group of patients applies to each individual in that group. This can be expressed by the sensitivity and the specificity of the MIC value. When these are both high, the probability of misclassification of an individual patient is low. In our study, the MIC had a low sensitivity of 56%, which meant that patients who were classified as showing improvement according to the GPI actually had a 44% probability of being misclassified as not showing improvement by the MIC. By contrast, the MIC had a much higher specificity of 75%, meaning that patients who were classified as not showing improvement according to the GPI had a 25% probability of being misclassified as showing improvement by the MIC. This was also illustrated by the considerable overlap of the curves for the groups that were and were not improved in the anchor-based MIC distribution plot (Figure 3).

Finally, one has to ensure that the MIC value is larger than the measurement error, as depicted by the SDC. In our study population, the MIC was considerably smaller than the SDC at the individual level, suggesting a high chance of measurement error if the MIC is applied to individual patients. However, if the PFDI-20 is used to study groups of patients, the scores are averaged, which should reduce the measurement error and
size of the SDC. In our population of 214 patients, the MIC did exceed the SDC at group level, implying that the MIC should be suitable for comparing multiple treatments at the group level, such as is necessary in clinical research or when guiding decisions regarding health policy.

CONCLUSION

In women with relatively mild pelvic floor symptoms, we estimated the MIC for improvement in the PFDI-20 to be 13.5 points (95% CI 6.2 to 20.9), representing a 23% reduction in the PFDI-20 score from baseline. This MIC may be useful for planning clinical trials and evaluating treatment effects in groups of women suitable for conservative treatment. However, due to the considerable probability of misclassification and the high chance of measurement error at the individual level, this MIC must be used with caution when interpreting the change scores of individual patients.
REFERENCES


10. Utomo E, Blok BF, Steensma AB, Korfage IJ. Validation of the pelvic floor distress inventory (PFDI-20) and pelvic floor impact questionnaire (PFIO-7) in a Dutch population. Int Urogynecol J. 2014;25:531-544.


15. de Vet HC, Foumani M, Scholten MA, et al. Minimally important change values of a measurement
instrument depend more on baseline values than on the type of intervention. J Clin Epidemiol. 2015;68:518-524.


