

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Minimal important differences in the International Prostate Symptom Score and Overactive Bladder Questionnaire in a primary care population

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-032795
Article Type:	Research
Date Submitted by the Author:	06-Jul-2019
Complete List of Authors:	Blanker, Marco; University of Groningen, University Medical Center Groningen, department of General Practice and Elderly Care Medicine Alma, Harma; University of Groningen, University Medical Center Groningen, Department of General Practice and Elderly Care Medicine; University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC) Devji, Tahira; McMaster University Faculty of Health Sciences, Clinical Epidemiology and Biostatistics Roelofs, Marjan; University of Groningen, University Medical Center Groningen, department of General Practice and Elderly Care Medicine Steffens, Martijn ; Isala Hospitals, urology van der Worp, Henk; University of Groningen, University medical center groningen, Department General Practice and Elderly Care Medicine
Keywords:	Lower Urinary Tract Symptoms, Adrenergic alpha-Antagonists, Primary Health Care, Minimal Clinically Important Difference, Patient outcome assessment

SCHOLARONE™
Manuscripts

Minimal important differences in the International Prostate Symptom Score and Overactive Bladder Questionnaire in a primary care population

M.H. (Marco) Blanker¹, H.J. (Harma) Alma¹, T. (Tahira) Devji², M. (Marjan) Roelofs¹, M.G. (Martijn) Steffens³, H. (Henk) van der Worp¹

1. Department of General Practice and Elderly Care Medicine, University Medical Centre Groningen, University of Groningen
2. Department of Health Research Methods, Evidence and Impact, McMaster University, Canada
3. Department of Urology, Isala Clinics, Zwolle, The Netherlands

Corresponding author: Marco H. Blanker, MD PhD, general practitioner and epidemiologist, Department of General Practice and Elderly Care Medicine, University Medical Centre Groningen, University of Groningen, P.O. box 196, 9700 AD Groningen, The Netherlands, Phone: +31 6 14961542, Fax: +31 50 3632964

E-mail addresses: m.h.blanker@umcg.nl, h.j.alma@umcg.nl, devjits@mcmaster.ca, marjanroelofs@hotmail.com, m.steffens@isala.nl, h.van.der.worp@umcg.nl

Wordcount: 2,939

Keywords: Lower Urinary Tract Symptoms; Adrenergic alpha-Antagonists; Primary Health Care; Minimal Clinically Important Difference; Patient outcome assessment

ABSTRACT

Objectives: To determine the minimal important difference (MID) of the International Prostate Symptom Score (IPSS) and the Overactive Bladder Questionnaire short form (OAB-q SF) assessed in primary care among patients treated for LUTS.

Design: Single-arm, open-label observational cohort study with a 6-week follow-up.

Setting: Twenty-two pharmacies in the Netherlands.

Participants: We enrolled Dutch men with uncomplicated LUTS who received a new alpha-blocker prescription from their general practitioner or urologist.

Primary and secondary outcomes: The IPSS and OAB-q SF were completed before and after 6 weeks of therapy. At 6 weeks, men also completed the Perceived Global Impression of Improvement (PGI-I). The mean change scores of the IPSS and OAB-q SF were calculated for each PGI-I outcome category, with the category 'a little better' used to determine the MID. The standard error of measurement (SEM) was calculated for each questionnaire.

Results: In total, 165 men completed follow-up. The MID was 5.2 points (95%CI, 3.9–6.4; SEM 3.6) for the IPSS and 11.0 points (95%CI, 7.1–14.9; SEM 9.7) for the OAB-q SF. However, the MID for the IPSS was higher in men with severe baseline symptoms (7.1; 95%CI, 5.3–9.0) than in men with moderate baseline symptoms (3.2; 95%CI, 1.7–4.8).

Conclusion: In this study, the MID for the IPSS was considerably higher than the MID of 3.1 reported in the only other study on this topic, but may be due to methodological differences. Interpretation of the MID for the OAB-q SF is hampered by the overlap with the SEM. Future studies are needed to confirm our results because correlations between the PGI-I and symptom questionnaires were suboptimal.

ARTICLE SUMMARY

Strengths and limitations of this study:

- We assessed the minimal important difference (MID) of two frequently used questionnaires on lower urinary tract symptoms.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
- Given that many men are treated in primary care, MID values for this setting are particularly important to inform evidence-based decision-making and to facilitate interpretation of the IPSS and OAB-q SF.
 - Notably, the sample size of this study was small, which resulted in very low numbers of men being included in the PGI-I category 'very much better' or 'worsening of symptoms'.

16
17
18
19
20
21
22
23
24

Funding statement: This work was supported by The Hein Hogerzeil Foundation with an unrestricted grant. The Foundation was not involved in the design of this study, nor in the data collection, data analyses, interpretation of the outcomes, or the writing of this manuscript.

25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Conflict of interest statement: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

INTRODUCTION

Symptom severity is a key outcome for patients with LUTS and is most often evaluated by direct patient inquiry, using patient-reported outcome measures (PROMs). Although the IPSS is most often used for this purpose in both clinical trials and practice,¹⁻⁵ it fails to capture problematic symptoms such as urinary incontinence and urgency. Therefore, the Overactive Bladder Questionnaire (OAB-q) is increasingly being used to evaluate the treatment of overactive bladder,^{6,7} with the short form (i.e., OAB-q SF) having the advantage of being less time consuming.⁸ Together, both of the IPSS and OAB-q SF capture the spectrum of outcomes that are important to patients, but it is difficult to interpret the effects of an intervention expressed as mean scores or change scores over time.

The minimal important difference (MID) has proven invaluable when interpreting PROMs and could be of great value for both the IPSS and OAB-q SF.^{9,10} To date, the MID has only been reported for the IPSS in a study conducted in secondary care among participants of clinical trials.^{1,2} There has been no report of the MID for the OAB-q SF in any care setting. In countries like the Netherlands and the United Kingdom, most men with LUTS first visit their GP to seek treatment. Given that setting may affect the MID, possibly because of differences in baseline symptom severity,⁹⁻¹¹ we feel that it is important to assess the MID in a primary care setting. Knowledge about the MID in primary care will then provide invaluable data for interpreting treatment outcomes that may differ between primary and secondary care. In addition, evidence must be obtained from multiple studies to ensure that MID determinations are accurate.

In the current study, we aimed to determine the MIDs for both the IPSS and OAB-q SF in a patient cohort originating mainly from primary care.

METHODS

Study design

We conducted a prospective cohort study between January 2016 and April 2018. Baseline data for the IPSS and OAB-q SF were compared with follow-up data after 6 weeks of treatment. At

1
2
3 follow-up, participants also completed the Perceived Global Impression of Improvement (PGI-I)
4
5 and we calculated the MID.

7 **Participants**

9 Adult men who visited a participating pharmacy in the north of the Netherlands were included
10
11 if they received a new alpha-blocker prescription for uncomplicated LUTS from a GP or
12
13 urologist. A prescription was defined as new if no alpha-blocker prescription had been given
14
15 within the past year. The pharmacists checked if the alpha-blocker was indicated for LUTS and
16
17 excluded men prescribed alpha-blockers for urinary tract stones or indwelling catheters. All
18
19 participants provided written informed consent. The medical ethics committee of the University
20
21 Medical Centre Groningen, the Netherlands, approved the study (number 2016.122).
22
23

24 **Data collection**

25
26 At baseline, before starting alpha-blocker therapy, all participants provided relevant descriptive
27
28 data (e.g., age, duration of LUTS in months or years, and history of surgery for LUTS) and
29
30 completed the Dutch versions of the IPSS and OAB-q SF. After 6 weeks, men who consented
31
32 repeated the IPSS and OAB-q SF by postal invitation. At this time, we asked participants to
33
34 complete the Perceived Global Impression of Improvement (PGI-I) questionnaire.¹²
35
36

37 **Questionnaires**

38
39 The IPSS contains seven questions and produces a total score that may range from 0 (no
40
41 symptoms) to 35 points (maximum score).¹ IPSS scores are often categorized as no/mild
42
43 symptoms (0–7 points), moderate symptoms (8–19 points), or severe symptoms (≥ 20 points).
44
45 The MID for the IPSS is currently considered to be 3.1 points.² The OAB-q SF contains six
46
47 questions on 6-point Likert-type scales, with the outcomes transformed to a 0–100 point scale
48
49 in which higher scores indicate more severe symptoms.⁸ The PGI-I is a validated generic tool for
50
51 assessing overall improvement after treatment and is answered on a 7-point Likert-type scale,
52
53 with the following options: 'very much better', 'much better', 'a little better', 'no change', 'a little
54
55 worse', 'much worse' or 'very much worse'. We sent a reminder after two weeks to patients who
56
57 did not respond to follow-up requests.
58
59
60

Data analyses

Baseline characteristics are reported as continuous variables and summarized as mean and sd or as median and interquartile range, depending on the distribution checked by the Shapiro–Wilk test. These characteristics were also compared between men with and without completed follow-up data to test for selective nonresponse. Next, the change scores of the IPSS and OAB-q SF were calculated by comparing the data between the baseline and follow-up questionnaires. Change scores were inverted to facilitate intuitive interpretation, with positive scores reflecting symptom improvement.

Various methods exist to determine the MID of questionnaires and are typically either anchor-based or distribution-based.^{9, 13-15} The latter involve evaluating change in the PROM with the probability that the change occurred by chance, sample variation or measurement precision; however, they do not reflect patient perspectives.^{14, 16} Thus, we used an anchor-based method,⁹ in which we compared changes in the IPSS or OAB-q SF (PROM) with the PGI-I (the anchor). For each PGI-I category, we then present the mean change in scores from baseline to follow-up with the associated CIs. We defined the MID as the mean change in IPSS or OAB-q SF for the PGI-I category ‘a little better’, but still present the mean change scores for the other PGI-I categories.

The usefulness of anchor-based approaches depends on the relationship between the PROM and the anchor.¹⁷⁻¹⁹ The anchor and PROM should be measuring the same or similar underlying constructs and should therefore be appreciably correlated. We therefore examined the Spearman correlation coefficients between the PGI-I and the IPSS and OAB-q SF for the baseline, follow-up and change data to ensure the anchor’s validity. A correlation coefficient between the symptom change scores and the PGI-I of ≥ 0.50 , and an equal and opposite correlation of the PGI-I with the baseline score and the follow-up score, were considered ideal and likely to yield trustworthy MID estimates.¹⁷⁻¹⁹

To test the impact of baseline symptom severity on the distribution of results, a stratified analysis was performed for the IPSS categories ‘moderate symptoms’ and ‘severe

1
2
3 symptoms' because previous research has shown that such stratification has a large impact.² No
4 such categories have been defined for the OAB-q SF, so we did not perform a similar analysis for
5 this questionnaire. Subgroup analyses were also performed with participants who received
6 their prescription from their GP, allowing us to provide data that focused on the primary care
7 setting. Finally, we checked if the MID exceeded the measurement error.^{15,17} For this, we
8 calculated the standard error of measurement (SEM) as follows: $[sd \times (1 - reliability)^{1/2}]$.
9 Cronbach's alpha was used as the reliability measure.²⁰

10
11 The complete data set was used without imputing missing data. All analyses were
12 performed using IBM SPSS version 25.0 (IBM Corp., Armonk, NY, USA), and we considered a p-
13 value <0.05 to be statistically significant.

14 **Patient and public involvement**

15
16 This study was performed without patient involvement. We did not invite patients to comment
17 on the study design nor did we consult them to interpret the results. Patients were not invited
18 to contribute to the writing or editing of this document for readability or accuracy.

19 **RESULTS**

20
21 A total of 251 men completed the baseline questionnaires, of which 165 also completed the
22 follow-up questionnaires. The baseline characteristics of men with and without follow-up data
23 are shown in Table 1, with no statistically significant differences found between these groups.
24 Notably, 86.3% of men received their prescription from a GP and the remainder received it from
25 a urologist.

26
27 There were mean improvements in the IPSS and OAB-q SF scores during the study of 5.8
28 (sd 6.7) and 11.8 points (sd 17.4), respectively. Between baseline and follow-up at 6 weeks, the
29 mean IPSS score changed from 19.1 (sd 6.8) to 13.3 (sd 6.5) and the mean OAB-q SF score
30 changed from 39.7 (sd 19.2) to 27.9 (sd 16.9). The PGI-I outcomes are shown in Table 2 and
31 indicate that most men reported that they were 'a little better' or 'much better' (74.7%), whilst
32 only 23.5% perceived no change. Only three men (1.8%) reported 'worsened' symptoms, and
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 none of the participants reported 'much worsened' or 'very much worsened' symptoms.
4

5 Table 2 also shows the distribution of changes in the IPSS and OAB-q SF for each PGI-I
6 category. The MID for the IPSS was 5.2 points (95% CI, 3.9–6.4) and the PGI-I outcomes 'no
7 change' and 'much better' corresponded to IPSS symptom changes of 3.1 points (95% CI, 1.1–
8 5.1) and 8.7 points (95% CI, 6.8–10.7), respectively. The MID for the OAB-q SF was 11.0 points
9 (95% CI, 7.1 to 14.9) and the PGI-I outcomes 'no change' and 'much better' corresponded with
10 mean improvements of 3.0 points (95% CI, -2.3 to 8.4) and 19.1 points (95% CI, 14.3–24.0),
11 respectively. The Spearman correlation coefficients were then calculated between the PGI-I and
12 both the IPSS and the OAB-q SF. The correlation was -0.51 for the PGI-I and baseline IPSS, 0.43
13 for the PGI-I and follow-up IPSS and 0.38 for the PGI-I and change in IPSS. The corresponding
14 correlations for the OAB-q SF were -0.09 at baseline, 0.36 at follow-up and 0.42 for the change.
15
16
17
18
19
20
21
22
23
24
25
26

27 Subgroup analyses of data for men with a prescription from a GP found no relevant
28 differences, with MID values of 5.4 for the IPSS and 11.2 for the OAB-q SF (Table 3). Stratified
29 analysis of baseline data revealed that men with severe symptoms had higher MID values for
30 the IPSS, reaching 7.1 (95% CI, 5.3–9.0), compared with the MID value of 3.2 (95% CI, 1.7–4.8)
31 for men with moderate symptoms (Table 4).
32
33
34
35
36

37 Finally, the SEM was 3.6 for the IPSS and 9.7 for the OAB-q SF.
38
39
40
41

42 **DISCUSSION**

43
44 We estimated the MID for two questionnaires that are often used to assess male LUTS in
45 primary care. However, whereas the SEM of the IPSS was less than the 95% CI of the MID (5.2
46 points; 95% CI, 3.9–6.4; SEM 3.6), the SEM of the OAB-q SF fell within the 95% CI of the MID (11
47 points; 95% CI, 7.1–14.9; SEM 9.7). Thus, we can only conclude that the outcomes for the IPSS
48 were unlikely to have occurred because of chance or measurement imprecision. Given that
49 many questionnaires have used multiple MID values, we were surprised to find only one
50 previous estimate of the MID for the IPSS in the literature.^{1,2} Our study therefore adds relevant
51 information in the primary care setting for clinicians and guideline developers.
52
53
54
55
56
57
58
59
60

1
2
3 Our results for the IPSS were different to those of the seminal study on this topic
4 performed by Barry et al. in secondary care.^{1,2} In that study, the MID of 3.1 points (sd 0.27) fell
5 within the 95% CI of the 'no change' group, but outside the CI of the 'a little better' group,
6 suggesting a likely underestimation of the real value. In the current study, there was also some
7 overlap between the CIs of the 'no change' and the 'a little better' group, though this was within
8 a change of only 3.9 to 5.1 points. Given that treatment is typically in primary care, we have
9 therefore provided additional data that is applicable to most men with LUTS. Nevertheless, the
10 differences in outcomes compared with the study by Barry et al. need to be explained. It is our
11 contention that three methodological differences account for these differences.
12
13
14
15
16
17
18
19
20
21

22 First, Barry et al. compared patients between baseline and follow-up after 13 weeks. By
23 contrast, the follow-up period in the current study was only 6 weeks. Although this difference of
24 7 weeks may have affected the ability of patients to recall their prior health state accurately, the
25 true impact of this remains unclear. Change scores may also have been influenced by the natural
26 variation that occurs in symptom severity over time.
27
28
29
30
31
32

33 Second, in the research by Barry et al., a global assessment of patient improvement was
34 used at baseline. This included a 5-point scale with the options 'marked improvement',
35 'moderate improvement', 'slight improvement', 'no improvement', and 'worse' for which the
36 exact question was not reported. In our study, we used a 7-point Likert-type scale that ranged
37 from 'very much better' to 'very much worsened'. We considered that this difference probably
38 had no more than a marginal impact given that the positive outcome categories were
39 comparable in both studies. Notably, none of the participants in our study reported that the
40 symptoms had 'very' or 'very much' worsened.
41
42
43
44
45
46
47
48
49

50 Third, we mainly included men from primary care, rather than men solely from
51 secondary care. Although it is generally thought that men in primary care have fewer symptoms,
52 our men tended to have more severe symptoms (IPSS score >19) than in the study by Barry et
53 al. (45% versus 25%). Barry et al. also reported that baseline severity had a major impact on the
54 MID, but when we compare their stratified analysis with ours, we had higher mean change
55
56
57
58
59
60

1
2
3 scores for each PGI-I category. This might be explained by the fact that we only included men
4 who actually used an alpha-blocker. In contrast to this focused approach, Barry et al. used data
5 for all participants in a large, randomized, double-blind trial of four treatment strategies for
6 male LUTS. In their study, a lower MID could therefore have resulted from the inclusion of
7 patients receiving placebo, finasteride, terazosin or combination therapy given that the efficacy
8 of alpha-blockers exceeds that of both placebo and finasteride. The use of blinding meant that
9 men who used placebo or finasteride may have overestimated their subjective improvement,
10 whilst alpha-blocker users may have underestimated their subjective improvement. Given that
11 the IPSS objectively counts symptoms, the placebo and finasteride users would experience a
12 smaller change in the IPSS whereas the active drug users would experience a larger effect.
13 Although the actual impact of each intervention is unknown, researchers in other fields have
14 made similar observations.^{10, 11, 21}

15
16 We were unable to find any prior estimates of the MID for the OAB-q SF in the literature.
17 Our finding that 11.0 points (95% CI, 7.1–14.9) indicates a clinically relevant change is therefore
18 a novel and important finding, but one for which the reliability will need to be assessed in other
19 studies. We recognize that alpha-blockers are not specifically indicated for the treatment of
20 overactive bladder, but we contend that there is a considerable overlap with LUTS unrelated to
21 overactive bladder. Indeed, guidelines suggest prescribing alpha-blockers for most men with
22 LUTS who request active treatment. This is because these agents have a rapid onset of action,
23 good efficacy, and low rate and severity of adverse events.³⁻⁵ We recommend further study to
24 determine the MID in men with specific symptoms of overactive bladder treated with
25 anticholinergics or beta-3 agonists.

26
27 Some limitations need to be considered when assessing our results. Notably, the sample
28 size of this study was small, which resulted in very low numbers of men being included in the
29 PGI-I category 'very much better'. For that category, the mean change scores for both
30 questionnaires showed very wide CIs. The same holds for the categories linked to symptom
31 worsening. In those categories, a trial discontinuation may be more suitable for reliable

1
2
3 estimates. The sample size might also explain why the CI of the MID estimate for the OAB-q SF
4 included the SEM.
5
6

7 Another limitation is reflected by difficulties we encountered with some of the
8 associations between the PGI-I and the two PROM questionnaires. For the IPSS, the follow-up
9 IPSS and PGI-I scores correlated better than with IPSS change and PGI-I scores, suggesting that
10 this rating only reflected the current status, which in turn, decreases confidence in the MID
11 estimate. For the OAB-q SF, the correlation coefficient between the baseline OAB-q SF and PGI-I
12 scores was opposite in magnitude to that for the follow-up OAB-q SF and PGI-I scores. With both
13 questionnaires, the correlation coefficients for the change scores were lower than the threshold
14 of 0.5 that we set a priori.¹⁸ High correlation coefficients are preferred between the anchor and
15 the change in PROM, though some researchers have suggested applying lower thresholds.²²
16 Still, even the high correlation coefficients are insufficient to confirm that the transition rating is
17 in fact measuring change as opposed to current health status.¹⁸ Unfortunately, Barry et al. did
18 not report the correlation coefficients between the IPSS and the anchor,² which is consistent
19 with most other research for PROMs.²³ Given the suboptimal relationship between the PROM
20 and the anchor, we must stress that the estimates obtained for the MID should be interpreted
21 with caution and should be confirmed in future investigations with larger samples.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39

40 In conclusion, this study is the first to define MID values for two important PROMs used
41 to evaluate the effectiveness of treatment for male LUTS in primary care. Given that many men
42 are treated in primary care, MID values for this setting are particularly important to inform
43 evidence-based decision-making and to facilitate interpretation of the IPSS and OAB-q SF.
44 Moreover, we consider that this study emphasizes the importance of the MID to individual
45 patients in daily practice. We defined the MID based on the PGI-I outcome 'a little better' in the
46 present study, but patients may expect 'much better' as an outcome when starting therapy. To
47 date, most outcomes of alpha-blocker and other drug treatments for male LUTS have been
48 expressed as the mean IPSS change scores. In the vast majority of studies³, difference in IPSS
49 changes between active treatment and placebo have approached, but not exceeded, the
50
51
52
53
54
55
56
57
58
59
60

1
2
3 previously reported MID of 3.1.^{1,2} Applying a threshold for improvement of 5.2 points, as
4 described in our study, may change the interpretation of those studies.
5
6
7
8

9 **Acknowledgements**

10 We are grateful for the assistance of Tom Vermist in collecting the data. The authors thank all
11 patients and collaborating pharmacies for their participation, as well as Dr Robert Sykes
12 (www.doctored.org.uk) for providing editorial services.
13
14
15
16
17

18 **Author contributions**

19 The corresponding author attests that all listed authors meet authorship criteria and that no
20 others meeting the criteria have been omitted.
21
22
23

24 MHB and MGS initiated the study. Acquisition of the data was done by MHB and MR. Analysis
25 and interpretation of the data was done by MHB, HJA, TD and HW. MHB wrote the manuscript
26 with extensive support from HJA and TD.
27
28
29

30 All authors critically reviewed the manuscript.
31
32

33 **Data sharing statement**

34 Data collected for this study will be available from the corresponding author upon request.
35
36
37
38

39 **REFERENCES**

40
41
42 1 Barry MJ, Fowler FJ,Jr, O'Leary MP, et al. The American Urological Association symptom index
43 for benign prostatic hyperplasia. The Measurement Committee of the American Urological
44 Association. *J Urol* 1992;148:1549,57; discussion 1564.
45
46
47

48 2 Barry MJ, Williford WO, Chang Y, et al. Benign prostatic hyperplasia specific health status
49 measures in clinical research: how much change in the American Urological Association
50 symptom index and the benign prostatic hyperplasia impact index is perceptible to patients?. *J*
51 *Urol* 1995;154:1770-4.
52
53
54

55 3 National Clinical Guideline Centre (NICE). Lower urinary tract symptoms in men: management
56 (Clinical guideline CG97) 2015;CG97
57 doi:<http://www.nice.org.uk/nicemedia/live/12984/48554/48554.pdf>.
58
59
60

- 1
2
3 4 Gratzke C, Bachmann A, Descazeaud A, et al. EAU Guidelines on the Assessment of Non-
4 neurogenic Male Lower Urinary Tract Symptoms including Benign Prostatic Obstruction. *Eur*
5 *Urol* 2015;67:1099-109 doi:10.1016/j.eururo.2014.12.038 [doi].
6
7
8
9 5 Blanker MH, Klomp MA, van den Donk M, et al. Summary of the NHG practice guideline 'Lower
10 urinary tract symptoms in men'. *Ned Tijdschr Geneeskd* 2013;157:A6178.
11
12
13 6 Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom
14 and health-related quality of life questionnaire: the OAB-q. *Qual Life Res* 2002;11:563-74.
15
16
17 7 Coyne KS, Matza LS, Thompson CL. The responsiveness of the Overactive Bladder
18 Questionnaire (OAB-q). *Qual Life Res* 2005;14:849-55.
19
20
21
22 8 Coyne KS, Thompson CL, Lai JS, et al. An overactive bladder symptom and health-related
23 quality of life short-form: validation of the OAB-q SF. *Neurol Urodyn* 2015;34:255-63
24 doi:10.1002/nau.22559 [doi].
25
26
27
28 9 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal
29 clinically important difference. *Control Clin Trials* 1989;10:407-15.
30
31
32
33 10 Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and
34 methods. *J Clin Epidemiol* 2010;63:524-34 doi:10.1016/j.jclinepi.2009.08.010 [doi].
35
36
37 11 Olsen MF, Bjerre E, Hansen MD, et al. Minimum clinically important differences in chronic
38 pain vary considerably by baseline pain and methodological factors: systematic review of
39 empirical studies. *J Clin Epidemiol* 2018;101:87,106.e2 doi:S0895-4356(18)30124-0 [pii].
40
41
42
43 12 Viktrup L, Hayes RP, Wang P, et al. Construct validation of patient global impression of
44 severity (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower
45 urinary tract symptoms secondary to benign prostatic hyperplasia. *BMC Urol* 2012;12:30,2490-
46 12-30 doi:10.1186/1471-2490-12-30 [doi].
47
48
49
50
51 13 Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the
52 significance of outcome effects above the statistical level, with methodological implications for
53 future studies. *J Clin Epidemiol* 2017;82:128-36 doi:S0895-4356(16)30776-4 [pii].
54
55
56
57 14 Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related
58 quality of life. *J Clin Epidemiol* 2003;56:395-407 doi:S0895435603000441 [pii].
59
60

1
2
3 15 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life:
4 the remarkable universality of half a standard deviation. *Med Care* 2003;41:582-92
5 doi:10.1097/01.MLR.0000062554.74615.4C [doi].
6
7

8
9 16 de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal
10 important difference. *J Clin Epidemiol* 2010;63:804,5; author reply 806
11 doi:10.1016/j.jclinepi.2009.12.015 [doi].
12
13

14
15 17 Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores:
16 differences between improvement and worsening. *Qual Life Res* 2002;11:207-21.
17
18

19 18 Guyatt GH, Norman GR, Juniper EF, et al. A critical look at transition ratings. *J Clin Epidemiol*
20 2002;55:900-8 doi:S0895435602004353 [pii].
21
22

23
24 19 Guyatt GH, Jaeschke RJ. Reassessing quality-of-life instruments in the evaluation of new
25 drugs. *Pharmacoeconomics* 1997;12:621-6 doi:10.2165/00019053-199712060-00002 [doi].
26
27

28 20 Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to
29 identify important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res*
30 2002;11:1-7.
31
32

33
34 21 Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in
35 degenerative knee disease outcomes: a systematic review and case study to inform BMJ Rapid
36 Recommendations. *BMJ Open* 2017;7:e015587,2016-015587 doi:10.1136/bmjopen-2016-
37 015587 [doi].
38
39

40
41 22 Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness
42 and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*
43 2008;61:102-9 doi:10.1016/j.jclinepi.2007.03.012 [doi].
44
45
46

47
48 23 Alma H, de Jong C, Tsiligianni I, et al. Clinically relevant differences in COPD health status:
49 systematic review and triangulation. *Eur Respir J* 2018;52:1800412
50 doi:10.1183/13993003.00412.
51
52
53
54
55
56
57
58
59
60

Table 1. Baseline characteristics of all participants and participants who dropped out after the baseline questionnaire

	Full participants (n = 165)	Drop out after baseline (n = 86)	p-value
Age (mean ± sd)	66.7±9.7	65.4±12.1	0.42*
Prescription from GP (%)	86.3	83.2	0.55#
IPSS score (mean ± sd)	19.1±6.8	17.6±6.5	0.11*
IPSS categories (%)			0.12#
- none/mild	3.7	6.4	
- moderate	50.9	61.5	
- severe	45.3	32.1	
IPSS Quality of Life (median IQR)	4.0 2.0	4.0 2.0	0.52§
OAB-q SF (mean ± sd)	39.8±19.2	40.7±18.1	0.70*
Duration of LUTS in months (median IQR)	24.0 37	12.0 33	0.11§
History of surgery for LUTS (%)	1.2	3.8	0.19#

IQR: interquartile range. P-values refer to: * student-t-test; # chi-squared test; § Mann-Whitney U test.

Abbreviations: OAB-q SF, Overactive Bladder Questionnaire short form.

Table 2. Change scores for the IPSS and OAB-q SF by PGI-I outcomes

PGI-I outcome	N (%)	IPSS	Missing	OAB-q SF	Missing
Very much better	6 (3.6)	13.4 (2.9;23.9)	1	23.8 (2.3;45.3)	0
Much better	50 (30.3)	8.7 (6.8;10.7)	2	19.1 (14.3;24.0)	3
A little better	68 (41.2)	5.2 (3.9;6.4)	3	11.0 (7.1;14.9)	4
No change	38 (23.0)	3.1 (1.1;5.1)	0	3.0 (-2.3;8.4)	4
A little worse	3 (1.8)	-5.0 (-30.9;20.9)	0	-9.7 (-81.7;62.4)	0

Change in IPSS and OAB-q SF scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better' reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Perceived Global Impression of Improvement.

Table 3. Change scores for the IPSS and OAB-q SF by PGI-I outcomes: subgroup analysis for GP prescriptions

PGI-I outcome	N (%)	IPSS	Missing	OAB-q SF	Missing
Very much better	4 (3.1)	18.0 (1.7;34.3)	1	30.8 (-0.9;62.4)	0
Much better	39 (30.2)	9.2 (7.0;11.5)	2	19.9 (14.8;24.9)	2
A little better	57 (44.2)	5.4 (4.0;6.7)	3	11.2 (7.0;15.4)	3
No change	27 (20.9)	3.1 (0.5;5.6)	0	3.3 (-3.4;9.9)	3
A little worse	2 (1.6)	-8.5 (-11.6.5; 99.5)	0	-16.5 (-353.2;320.2)	0

Change in IPSS and OAB-q SF scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better' reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Perceived Global Impression of Improvement.

Table 4. Change scores for the IPSS by PGI-I outcomes

PGI-I outcome	Moderate symptoms (n = 88)			Severe symptoms (n = 73)	
	N (%)	Change	Missing	N (%)	Change
Very much better	3 (3.4)	6.5 (-50.7;63.7)	1	3 (4.1)	18 (1.7;34.3)
Much better	30 (34.1)	5.6 (3.7;7.5)	1	19 (26.0)	13.5 (10.5;16.5)
A little better	33 (37.5)	3.2 (1.7;4.8)	0	32 (43.8)	7.1 (5.3;9.0)
No change	19 (21.6)	1.3 (-1.7;4.3)	0	19 (26.0)	4.9 (2.3;7.6)
A little worse	3 (3.4)	-5.0 (-31.0;20.9)	0	0 (0.0)	-

These results are stratified by baseline symptom severity on the IPSS: moderate symptoms are scores of 8–19 and severe symptoms are scores of ≥ 20 . Change in IPSS scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category ‘a little better’ reflects the MID for both questionnaires. None of the participants scored ‘much worsened’ or ‘very much worsened’ on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Perceived Global Impression of Improvement.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation	Page No
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	2 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) If applicable, explain how loss to follow-up was addressed (e) Describe any sensitivity analyses	6-7 6 7
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	7
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	7 7 7
Outcome data	15*	Report numbers of outcome events or summary measures over time	-

1	Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-8
2			(b) Report category boundaries when continuous variables were categorized	7-8
3			(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
4				
5				
6				
7				
8				
9	Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-8
10				
11	Discussion			
12				
13	Key results	18	Summarise key results with reference to study objectives	8
14	Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10-11
15				
16	Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	8-11
17				
18				
19	Generalisability	21	Discuss the generalisability (external validity) of the study results	11
20				
21	Other information			
22	Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3
23				
24				

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at <http://www.strobe-statement.org>.

BMJ Open

Determining the minimal important differences in the International Prostate Symptom Score and Overactive Bladder Questionnaire: results from an observational cohort study in primary care

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-032795.R1
Article Type:	Original research
Date Submitted by the Author:	14-Oct-2019
Complete List of Authors:	Blanker, Marco; University of Groningen, University Medical Center Groningen, department of General Practice and Elderly Care Medicine Alma, Harma; University of Groningen, University Medical Center Groningen, Department of General Practice and Elderly Care Medicine; University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC) Devji, Tahira; McMaster University Faculty of Health Sciences, Clinical Epidemiology and Biostatistics Roelofs, Marjan; University of Groningen, University Medical Center Groningen, department of General Practice and Elderly Care Medicine Steffens, Martijn ; Isala Hospitals, urology van der Worp, Henk; University of Groningen, University medical center groningen, Department General Practice and Elderly Care Medicine
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Urology, Epidemiology
Keywords:	Lower Urinary Tract Symptoms, Adrenergic alpha-Antagonists, Primary Health Care, Minimal Clinically Important Difference, Patient outcome assessment

SCHOLARONE™
Manuscripts

1
2
3 **Determining the minimal important differences in the International Prostate**
4
5 **Symptom Score and Overactive Bladder Questionnaire: results from an**
6
7 **observational cohort study in primary care**
8
9

10
11
12 M.H. (Marco) Blanker¹, H.J. (Harma) Alma¹, T. (Tahira) Devji², M. (Marjan) Roelofs¹, M.G.
13
14 (Martijn) Steffens³, H. (Henk) van der Worp¹
15
16

- 17
18
19 1. Department of General Practice and Elderly Care Medicine, University Medical Centre
20
21 Groningen, University of Groningen
22
23 2. Department of Health Research Methods, Evidence and Impact, McMaster University,
24
25 Canada
26
27 3. Department of Urology, Isala Clinics, Zwolle, The Netherlands
28
29
30

31
32 Corresponding author: Marco H. Blanker, MD PhD, general practitioner and epidemiologist,
33
34 Department of General Practice and Elderly Care Medicine, University Medical Centre
35
36 Groningen, University of Groningen, P.O. box 196, 9700 AD Groningen, The Netherlands, Phone:
37
38 +31 6 14961542, Fax: +31 50 3632964
39
40

41
42 E-mail addresses: m.h.blanker@umcg.nl, h.j.alma@umcg.nl, devjits@mcmaster.ca,
43
44 marjanroelofs@hotmail.com, m.steffens@isala.nl, h.van.der.worp@umcg.nl
45
46

47
48 **Wordcount:** 3,116

49
50 **Keywords:** Lower Urinary Tract Symptoms; Adrenergic alpha-Antagonists; Primary
51
52 Health Care; Minimal Clinically Important Difference; Patient outcome assessment
53
54
55
56
57
58
59
60

ABSTRACT

Objectives: To determine the minimal important difference (MID) of the International Prostate Symptom Score (IPSS) and the Overactive Bladder Questionnaire short form (OAB-q SF) assessed in primary care among patients treated for lower urinary tract symptoms (LUTS).

Design: Single-arm, open-label observational cohort study with a 6-week follow-up.

Setting: Twenty-two pharmacies in the Netherlands.

Participants: We enrolled Dutch men with uncomplicated LUTS who received a new alpha-blocker prescription from their general practitioner or urologist.

Primary and secondary outcomes: The IPSS and OAB-q SF were completed before and after 6 weeks of therapy. At 6 weeks, men also completed the Patient Global Impression of Improvement (PGI-I). The mean change scores of the IPSS and OAB-q SF were calculated for each PGI-I outcome category, with the category 'a little better' used to determine the MID. The standard error of measurement (SEM) was calculated for each questionnaire.

Results: In total, 165 men completed follow-up. The MID was 5.2 points (95%CI, 3.9–6.4; SEM 3.6) for the IPSS and 11.0 points (95%CI, 7.1–14.9; SEM 9.7) for the OAB-q SF. For both questionnaires, confidence intervals showed an overlap with the no-change categories. However, the MID for the IPSS was higher in men with severe baseline symptoms (7.1; 95%CI, 5.3–9.0) than in men with moderate baseline symptoms (3.2; 95%CI, 1.7–4.8).

Conclusion: In this study, the MID for the IPSS was considerably higher than the MID of 3.1 reported in the only other study on this topic, but may be due to methodological differences. Interpretation of the MID for the OAB-q SF is hampered by the overlap with the SEM. Future studies are needed to confirm our results because correlations between the PGI-I and symptom questionnaires were suboptimal.

ARTICLE SUMMARY

Strengths and limitations of this study:

- We assessed the minimal important difference (MID) of two frequently used questionnaires

1
2
3 on lower urinary tract symptoms.
4

- 5 • Given that many men are treated in primary care, MID values for this setting are
6 particularly important to inform evidence-based decision-making and to facilitate
7 interpretation of the IPSS and OAB-q SF.
8
9
- 10 • Notably, the sample size of this study was small, which resulted in very low numbers of men
11 being included in the PGI-I category 'very much better' or 'worsening of symptoms', and may
12 clarify the small overlap of the confidence intervals with the no-change category.
13
14
15
16
17
18
19

20 **Funding statement:** This work was supported by The Hein Hogerzeil Foundation with an
21 unrestricted grant. The Foundation was not involved in the design of this study, nor in the data
22 collection, data analyses, interpretation of the outcomes, or the writing of this manuscript.
23
24
25

26
27
28 **Conflict of interest statement:** All authors have completed the ICMJE uniform disclosure form
29 at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the
30 submitted work; no financial relationships with any organisations that might have an interest in
31 the submitted work in the previous three years; no other relationships or activities that could
32 appear to have influenced the submitted work.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Symptom severity is a key outcome for patients with LUTS and is most often evaluated by direct patient inquiry, using patient-reported outcome measures (PROMs). Although the IPSS is most often used for this purpose in both clinical trials and practice,¹⁻⁵ it fails to capture problematic symptoms such as urinary incontinence and urgency. Therefore, the Overactive Bladder Questionnaire (OAB-q) is increasingly being used to evaluate the treatment of overactive bladder,^{6,7} with the short form (i.e., OAB-q SF) having the advantage of being less time consuming.⁸ Together, both of the IPSS and OAB-q SF capture the spectrum of outcomes that are important to patients, but it is difficult to interpret the effects of an intervention expressed as mean scores or change scores over time.

The minimal important difference (MID) has proven invaluable when interpreting PROMs and could be of great value for both the IPSS and OAB-q SF.^{9,10} To date, the MID has only been reported for the IPSS in a study conducted in secondary care among participants of clinical trials.^{1,2} There has been no report of the MID for the OAB-q SF in any care setting. In countries like the Netherlands and the United Kingdom, most men with LUTS first visit their GP to seek treatment. Given that setting may affect the MID, possibly because of differences in baseline symptom severity,⁹⁻¹¹ we feel that it is important to assess the MID in a primary care setting. Knowledge about the MID in primary care will then provide invaluable data for interpreting treatment outcomes that may differ between primary and secondary care. In addition, evidence must be obtained from multiple studies to ensure that MID determinations are accurate.

In the current study, we aimed to determine the MIDs for both the IPSS and OAB-q SF in a patient cohort originating mainly from primary care.

METHODS

Study design

We conducted a prospective cohort study between January 2016 and April 2018.¹² Baseline data for the IPSS and OAB-q SF were compared with follow-up data after 6 weeks of treatment.

1
2
3 At follow-up, participants also completed the Patient Global Impression of Improvement (PGI-I)
4
5 and we calculated the MID.
6

7 **Participants**

8
9 Adult men who visited a participating pharmacy in the north of the Netherlands were included
10
11 if they received a new alpha-blocker prescription for uncomplicated LUTS from a GP or
12
13 urologist. A prescription was defined as new if no alpha-blocker prescription had been given
14
15 within the past year. The pharmacists checked if the alpha-blocker was indicated for LUTS and
16
17 excluded men prescribed alpha-blockers for urinary tract stones or indwelling catheters. All
18
19 participants provided written informed consent. The medical ethics committee of the University
20
21 Medical Centre Groningen, the Netherlands, approved the study (number 2016.122).
22
23

24 **Data collection**

25
26 At baseline, before starting alpha-blocker therapy, all participants provided relevant descriptive
27
28 data (e.g., age, duration of LUTS in months or years, and history of surgery for LUTS) and
29
30 completed the Dutch versions of the IPSS and OAB-q SF. After 6 weeks, men who consented
31
32 repeated the IPSS and OAB-q SF by postal invitation. At this time, we asked participants to
33
34 complete the Patient Global Impression of Improvement (PGI-I) questionnaire.¹³ The period of
35
36 six weeks was chosen as clinical effects of alpha-blockers take a few weeks to develop fully, but
37
38 significant efficacy over placebo can occur within hours to days.^{4,14}
39

40 **Questionnaires**

41
42 The IPSS contains seven questions and produces a total score that may range from 0 (no
43
44 symptoms) to 35 points (maximum score).¹ Each question has response options ranging from 0
45
46 to five, with higher scores reflecting more severe symptoms. IPSS scores are often categorized
47
48 as no/mild symptoms (0–7 points), moderate symptoms (8–19 points), or severe symptoms
49
50 (≥ 20 points). The IPSS was internally consistent (Cronbach's alpha = 0.86) and has excellent
51
52 test-retest reliability ($r = 0,92$).¹ The MID for the IPSS is currently considered to be 3.1 points.²
53
54 The OAB-q SF contains six questions on 6-point Likert-type scales, with the outcomes
55
56 transformed to a 0–100 point scale in which higher scores indicate more severe symptoms.⁸
57
58
59
60

1
2
3 This scale demonstrated good convergent validity, discriminant validity, internal reliability,
4 reproducibility, and responsiveness to change.⁸ The PGI-I is a validated generic tool for
5 assessing overall improvement after treatment and is answered on a 7-point Likert-type scale,
6 with the following options: 'very much better', 'much better', 'a little better', 'no change', 'a little
7 worse', 'much worse' or 'very much worse'.^{13,15} Full versions of these questionnaires are
8 presented as supplementary file 1.

9
10 We sent a reminder after two weeks to patients who did not respond to follow-up requests.

11 **Data analyses**

12 Baseline characteristics are reported as continuous variables and summarized as mean and SD
13 or as median and interquartile range, depending on the distribution checked by the Shapiro-
14 Wilk test. These characteristics were also compared between men with and without completed
15 follow-up data to test for selective nonresponse. Next, the change scores of the IPSS and OAB-q
16 SF were calculated by comparing the data between the baseline and follow-up questionnaires.
17 Change scores were inverted to facilitate intuitive interpretation, with positive scores reflecting
18 symptom improvement.

19
20 Various methods exist to determine the MID of questionnaires and are typically either
21 anchor-based or distribution-based.^{9,16-18} The latter involve evaluating change in the PROM
22 with the probability that the change occurred by chance, sample variation or measurement
23 precision; however, they do not reflect patient perspectives.^{17,19} Thus, we used an anchor-based
24 method,⁹ in which we compared changes in the IPSS or OAB-q SF (PROM) with the PGI-I (the
25 anchor). For each PGI-I category, we then present the mean change in scores from baseline to
26 follow-up with the associated CIs. We defined the MID as the mean change in IPSS or OAB-q SF
27 for the PGI-I category 'a little better', but still present the mean change scores for the other PGI-I
28 categories.

29
30 The usefulness of anchor-based approaches depends on the relationship between the
31 PROM and the anchor.²⁰⁻²² The anchor and PROM should be measuring the same or similar
32 underlying constructs and should therefore be appreciably correlated. Correlations between
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 questionnaire change scores and the anchor PGI-I should be obviously strong, as else these
4
5 measure different concepts. Correlations between the anchor PGI-I and the baseline and follow-
6
7 up questionnaire scores are performed to check for a possible response shift. Mostly anchor
8
9 PGI-I scores seem correlated with follow-up scores (due to response shift). We therefore
10
11 examined the Spearman correlation coefficients between the PGI-I and the IPSS and OAB-q SF
12
13 for the baseline, follow-up and change data to ensure the anchor's validity. A correlation
14
15 coefficient between the symptom change scores and the PGI-I of ≥ 0.50 , and an equal and
16
17 opposite correlation of the PGI-I with the baseline score and the follow-up score, were
18
19 considered ideal and likely to yield trustworthy MID estimates.²⁰⁻²²

22
23 To test the impact of baseline symptom severity on the distribution of results, a
24
25 stratified analysis was performed for the IPSS categories 'moderate symptoms' and 'severe
26
27 symptoms' because previous research has shown that such stratification has a large impact.² No
28
29 such categories have been defined for the OAB-q SF, so we did not perform a similar analysis for
30
31 this questionnaire. Subgroup analyses were also performed with participants who received
32
33 their prescription from their GP, allowing us to provide data that focused on the primary care
34
35 setting. Finally, we checked if the MID exceeded the measurement error.^{15,17} For this, we
36
37 calculated the standard error of measurement (SEM) as follows: $[SD \times (1 - \text{reliability})^{1/2}]$.
38
39 Cronbach's alpha was used as the reliability measure.²³

41
42 The complete data set was used without imputing missing data. All analyses were
43
44 performed using IBM SPSS version 25.0 (IBM Corp., Armonk, NY, USA), and we considered a p-
45
46 value < 0.05 to be statistically significant.

48 **Patient and public involvement**

49
50 This study was performed without patient involvement. We did not invite patients to comment
51
52 on the study design nor did we consult them to interpret the results. Patients were not invited
53
54 to contribute to the writing or editing of this document for readability or accuracy.

58 **RESULTS**

A total of 251 men completed the baseline questionnaires, of which 165 also completed the follow-up questionnaires. The baseline characteristics of men with and without follow-up data are shown in Table 1, with no statistically significant differences found between these groups. Notably, 86.3% of the participants received their prescription from a GP and the remainder received it from a urologist.

Table 1. Baseline characteristics of all participants and participants who dropped out after the baseline questionnaire

	Participants with completed follow-up (n = 165)	Drop out after baseline (n = 86)	p-value
Age (mean ± SD)	66.7±9.7	65.4±12.1	0.42*
Prescription from GP (%)	86.3	83.2	0.55#
IPSS score (mean ± SD)	19.1±6.8	17.6±6.5	0.11*
IPSS categories (%)			0.12#
- none/mild	3.7	6.4	
- moderate	50.9	61.5	
- severe	45.3	32.1	
IPSS Quality of Life (median IQR)	4.0 3.0-5.0	4.0 3.0-5.0	0.52§
OAB-q SF (mean ± SD)	39.8±19.2	40.7±18.1	0.70*
Duration of LUTS in months (median IQR)	24.0 5.0-42.0	12.0 3.0-36.0	0.11§
History of surgery for LUTS (%)	1.2	3.8	0.19#

IQR: interquartile range. P-values refer to: * student-t-test; # chi-squared test; § Mann-Whitney U test.

Abbreviations: OAB-q SF, Overactive Bladder Questionnaire short form.

There were mean improvements in the IPSS and OAB-q SF scores during the study of 5.8 (SD 6.7) and 11.8 points (SD 17.4), respectively. Between baseline and follow-up at 6 weeks, the mean IPSS score changed from 19.1 (SD 6.8) to 13.3 (SD 6.5) and the mean OAB-q SF score

changed from 39.7 (SD 19.2) to 27.9 (SD 16.9). The PGI-I outcomes are shown in Table 2 and indicate that most men reported that they were 'a little better' or 'much better' (74.7%), whilst only 23.5% perceived no change. Only three men (1.8%) reported 'worsened' symptoms, and none of the participants reported 'much worsened' or 'very much worsened' symptoms.

Table 2. Change scores for the IPSS and OAB-q SF by PGI-I outcomes

PGI-I outcome	N (%)	IPSS	Missing	OAB-q SF	Missing
Very much better	6 (3.6)	13.4 (2.9;23.9)	1	23.8 (2.3;45.3)	0
Much better	50 (30.3)	8.7 (6.8;10.7)	2	19.1 (14.3;24.0)	3
A little better	68 (41.2)	5.2 (3.9;6.4)	3	11.0 (7.1;14.9)	4
No change	38 (23.0)	3.1 (1.1;5.1)	0	3.0 (-2.3;8.4)	4
A little worse	3 (1.8)	-5.0 (-30.9;20.9)	0	-9.7 (-81.7;62.4)	0

Change in IPSS and OAB-q SF scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better' reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Patient Global Impression of Improvement.

Table 2 also shows the distribution of changes in the IPSS and OAB-q SF for each PGI-I category. The MID for the IPSS was 5.2 points (95% CI, 3.9–6.4) and the PGI-I outcomes 'no change' and 'much better' corresponded to IPSS symptom changes of 3.1 points (95% CI, 1.1–5.1) and 8.7 points (95% CI, 6.8–10.7), respectively. The MID for the OAB-q SF was 11.0 points (95% CI, 7.1 to 14.9) and the PGI-I outcomes 'no change' and 'much better' corresponded with mean improvements of 3.0 points (95% CI, -2.3 to 8.4) and 19.1 points (95% CI, 14.3–24.0), respectively. For both questionnaires, the confidence intervals of the MID-categories showed an overlap with the 'no change' categories.

The Spearman correlation coefficients were then calculated between the PGI-I and both

the IPSS and the OAB-q SF. The correlation was -0.51 for the PGI-I and baseline IPSS, 0.43 for the PGI-I and follow-up IPSS and 0.38 for the PGI-I and change in IPSS. The corresponding correlations for the OAB-q SF were -0.09 at baseline, 0.36 at follow-up and 0.42 for the change.

Table 3. Change scores for the IPSS and OAB-q SF by PGI-I outcomes: subgroup analysis for GP prescriptions

PGI-I outcome	N (%)	IPSS	Missing	OAB-q SF	Missing
Very much better	4 (3.1)	18.0 (1.7;34.3)	1	30.8 (-0.9;62.4)	0
Much better	39 (30.2)	9.2 (7.0;11.5)	2	19.9 (14.8;24.9)	2
A little better	57 (44.2)	5.4 (4.0;6.7)	3	11.2 (7.0;15.4)	3
No change	27 (20.9)	3.1 (0.5;5.6)	0	3.3 (-3.4;9.9)	3
A little worse	2 (1.6)	-8.5 (-11.6.5; 99.5)	0	-16.5 (-353.2;320.2)	0

Change in IPSS and OAB-q SF scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better' reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Patient Global Impression of Improvement.

Subgroup analyses of data for men with a prescription from a GP found no relevant differences, with MID values of 5.4 for the IPSS and 11.2 for the OAB-q SF (Table 3). Stratified analysis of baseline data revealed that men with severe symptoms had higher MID values for the IPSS, reaching 7.1 (95% CI, 5.3–9.0), compared with the MID value of 3.2 (95% CI, 1.7–4.8) for men with moderate symptoms (Table 4).

Table 4. Change scores for the IPSS by PGI-I outcomes

PGI-I outcome	Moderate symptoms (n = 88)			Severe symptoms (n = 73)	
	N (%)	Change	Missing	N (%)	Change
Very much better	3 (3.4)	6.5 (-50.7;63.7)	1	3 (4.1)	18 (1.7;34.3)
Much better	30 (34.1)	5.6 (3.7;7.5)	1	19 (26.0)	13.5 (10.5;16.5)
A little better	33 (37.5)	3.2 (1.7;4.8)	0	32 (43.8)	7.1 (5.3;9.0)
No change	19 (21.6)	1.3 (-1.7;4.3)	0	19 (26.0)	4.9 (2.3;7.6)
A little worse	3 (3.4)	-5.0 (-31.0;20.9)	0	0 (0.0)	-

These results are stratified by baseline symptom severity on the IPSS: moderate symptoms are scores of 8–19 and severe symptoms are scores of ≥ 20 . Change in IPSS scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category ‘a little better’ reflects the MID for both questionnaires. None of the participants scored ‘much worsened’ or ‘very much worsened’ on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Patient Global Impression of Improvement.

Finally, the SEM was 3.6 for the IPSS and 9.7 for the OAB-q SF.

DISCUSSION

We estimated the MID for two questionnaires that are often used to assess male LUTS in primary care. However, whereas the SEM of the IPSS was less than the 95% CI of the MID (5.2 points; 95% CI, 3.9–6.4; SEM 3.6), the SEM of the OAB-q SF fell within the 95% CI of the MID (11 points; 95% CI, 7.1–14.9; SEM 9.7). Thus, we can only conclude that the outcomes for the IPSS were unlikely to have occurred because of chance or measurement imprecision. Given that many questionnaires have used multiple MID values, we were surprised to find only one previous estimate of the MID for the IPSS in the literature.^{1,2} Our study therefore adds relevant information in the primary care setting for clinicians and guideline developers.

Our results for the IPSS were different to those of the seminal study on this topic performed by Barry et al. in secondary care.^{1,2} In that study, the MID of 3.1 points (SD 0.27) fell

1
2
3 within the 95% CI of the 'no change' group (consisting of men who expressed that they hadn't
4 experienced any change in symptoms), but outside the CI of the 'a little better' group, suggesting
5 a likely underestimation of the real value. In the current study, there was also some overlap
6 between the CIs of the 'no change' and the 'a little better' group, though this was within a change
7 of only 3.9 to 5.1 points. Given that treatment is typically in primary care, we have therefore
8 provided additional data that is applicable to most men with LUTS. Nevertheless, the differences
9 in outcomes compared with the study by Barry et al. need to be explained. It is our contention
10 that three methodological differences account for these differences.
11
12
13
14
15
16
17
18
19

20 First, Barry et al. compared patients between baseline and follow-up after 13 weeks. By
21 contrast, the follow-up period in the current study was only 6 weeks. Although this difference of
22 7 weeks may have affected the ability of patients to recall their prior health state accurately, the
23 true impact of this remains unclear. Change scores may also have been influenced by the natural
24 variation that occurs in symptom severity over time.
25
26
27
28
29

30 Second, in the research by Barry et al., a different global assessment of patient
31 improvement was used. This included a 5-point scale with the options 'marked improvement',
32 'moderate improvement', 'slight improvement', 'no improvement', and 'worse' for which the
33 exact question was not reported. In our study, we used a 7-point Likert-type scale that ranged
34 from 'very much better' to 'very much worsened'. We considered that this difference probably
35 had no more than a marginal impact given that the positive outcome categories were
36 comparable in both studies. Notably, none of the participants in our study reported that the
37 symptoms had 'very' or 'very much' worsened.
38
39
40
41
42
43
44
45
46
47

48 Third, we mainly included men from primary care, rather than men solely from
49 secondary care. Although it is generally thought that men in primary care have fewer symptoms,
50 our men tended to have more severe symptoms (IPSS score >19) than in the study by Barry et
51 al. (45% versus 25%). Barry et al. also reported that baseline severity had a major impact on the
52 MID, but when we compare their stratified analysis with ours, we had higher mean change
53 scores for each PGI-I category. This might be explained by the fact that we only included men
54
55
56
57
58
59
60

1
2
3 who actually used an alpha-blocker. In contrast to this focused approach, Barry et al. used data
4
5 for all participants in a large, randomized, double-blind trial of four treatment strategies for
6
7 male LUTS. In their study, a lower MID could therefore have resulted from the inclusion of
8
9 patients receiving placebo, finasteride, terazosin or combination therapy given that the efficacy
10
11 of alpha-blockers exceeds that of both placebo and finasteride. The use of blinding meant that
12
13 men who used placebo or finasteride may have overestimated their subjective improvement,
14
15 whilst alpha-blocker users may have underestimated their subjective improvement. Given that
16
17 the IPSS objectively counts symptoms, the placebo and finasteride users would experience a
18
19 smaller change in the IPSS whereas the active drug users would experience a larger effect.
20
21 Although the actual impact of each intervention is unknown, researchers in other fields have
22
23 made similar observations.^{10, 11, 24}
24
25

26
27 We were unable to find any prior estimates of the MID for the OAB-q SF in the literature.
28
29 Our finding that 11.0 points (95% CI, 7.1–14.9) indicates a clinically relevant change is therefore
30
31 a novel and important finding, but one for which the reliability will need to be assessed in other
32
33 studies. We recognize that alpha-blockers are not specifically indicated for the treatment of
34
35 overactive bladder, but we contend that there is a considerable overlap with LUTS unrelated to
36
37 overactive bladder. Indeed, guidelines suggest prescribing alpha-blockers for most men with
38
39 LUTS who request active treatment. This is because these agents have a rapid onset of action,
40
41 good efficacy, and low rate and severity of adverse events.³⁻⁵ We recommend further study to
42
43 determine the MID in men with specific symptoms of overactive bladder treated with
44
45 anticholinergics or beta-3 agonists.
46
47

48
49 Some limitations need to be considered when assessing our results. Notably, the sample
50
51 size of this study was small, which resulted in very low numbers of men being included in the
52
53 PGI-I category 'very much better'. For that category, the mean change scores for both
54
55 questionnaires showed very wide CIs. The same holds for the categories linked to symptom
56
57 worsening. In those categories, a discontinuation trial, in which men stop their treatment, may
58
59 be more suitable for reliable estimates. The sample size might also explain why the CI of the
60

1
2
3 MID estimate for the OAB-q SF included the SEM.
4

5 Another limitation is reflected by difficulties we encountered with some of the
6 associations between the PGI-I and the two PROM questionnaires. For the IPSS, the follow-up
7 IPSS and PGI-I scores correlated better than with IPSS change and PGI-I scores, suggesting that
8 this rating only reflected the current status, which in turn, decreases confidence in the MID
9 estimate. For the OAB-q SF, the correlation coefficient between the baseline OAB-q SF and PGI-I
10 scores was opposite in magnitude to that for the follow-up OAB-q SF and PGI-I scores. With both
11 questionnaires, the correlation coefficients for the change scores were lower than the threshold
12 of 0.5 that we set a priori.²¹ High correlation coefficients are preferred between the anchor and
13 the change in PROM, though some researchers have suggested applying lower thresholds.²⁵
14 Still, even the high correlation coefficients are insufficient to confirm that the transition rating is
15 in fact measuring change as opposed to current health status.²¹ Unfortunately, Barry et al. did
16 not report the correlation coefficients between the IPSS and the anchor,² which is consistent
17 with most other research for PROMs.²⁶ Given the suboptimal relationship between the PROM
18 and the anchor, we must stress that the estimates obtained for the MID should be interpreted
19 with caution and should be confirmed in future investigations with larger samples.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

37 In conclusion, this study is the first to define MID values for two important PROMs used
38 to evaluate the effectiveness of treatment for male LUTS in primary care. Given that many men
39 are treated in primary care, MID values for this setting are particularly important to inform
40 evidence-based decision-making and to facilitate interpretation of the IPSS and OAB-q SF.
41 Moreover, we consider that this study emphasizes the importance of the MID to individual
42 patients in daily practice. We defined the MID based on the PGI-I outcome 'a little better' in the
43 present study, but patients may expect 'much better' as an outcome when starting therapy. To
44 date, most outcomes of alpha-blocker and other drug treatments for male LUTS have been
45 expressed as the mean IPSS change scores. In the vast majority of studies³, difference in IPSS
46 changes between active treatment and placebo have approached, but not exceeded, the
47 previously reported MID of 3.1.^{1,2} Applying a threshold for improvement of 5.2 points, as
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 described in our study, may change the interpretation of those studies.
4
5
6

7 **Acknowledgements**

8
9 We are grateful for the assistance of Tom Vermist in collecting the data. The authors thank all
10 patients and collaborating pharmacies for their participation, as well as Dr Robert Sykes
11 (www.doctored.org.uk) for providing editorial services.
12
13
14

15 **Author contributions**

16
17 The corresponding author attests that all listed authors meet authorship criteria and that no
18 others meeting the criteria have been omitted.
19
20

21 MHB and MGS initiated the study. Acquisition of the data was done by MHB and MR. Analysis
22 and interpretation of the data was done by MHB, HJA, TD and HW. MHB wrote the manuscript
23 with extensive support from HJA and TD.
24
25

26 All authors critically reviewed the manuscript.
27
28

29 **Data sharing statement**

30
31 Data collected for this study will be available from the corresponding author upon request.
32
33
34
35
36

37 **REFERENCES**

38
39
40 1 Barry MJ, Fowler FJ, Jr, O'Leary MP, et al. The American Urological Association symptom index
41 for benign prostatic hyperplasia. The Measurement Committee of the American Urological
42 Association. *J Urol* 1992;148:1549,57; discussion 1564.
43

44 2 Barry MJ, Williford WO, Chang Y, et al. Benign prostatic hyperplasia specific health status
45 measures in clinical research: how much change in the American Urological Association symptom
46 index and the benign prostatic hyperplasia impact index is perceptible to patients?. *J Urol*
47 1995;154:1770-4.
48

49 3 National Clinical Guideline Centre (NICE). Lower urinary tract symptoms in men: management
50 (Clinical guideline CG97) 2015;CG97
51 doi:<http://www.nice.org.uk/nicemedia/live/12984/48554/48554.pdf>.
52

53 4 Gratzke C, Bachmann A, Descazeaud A, et al. EAU Guidelines on the Assessment of Non-
54 neurogenic Male Lower Urinary Tract Symptoms including Benign Prostatic Obstruction. *Eur Urol*
55 2015;67:1099-109 doi:10.1016/j.eururo.2014.12.038 [doi].
56
57

58 5 Blanker MH, Klomp MA, van den Donk M, et al. Summary of the NHG practice guideline 'Lower
59 urinary tract symptoms in men'. *Ned Tijdschr Geneeskd* 2013;157:A6178.
60

- 1
2
3 6 Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and
4 health-related quality of life questionnaire: the OAB-q. *Qual Life Res* 2002;11:563-74.
5
6 7 Coyne KS, Matza LS, Thompson CL. The responsiveness of the Overactive Bladder Questionnaire
7 (OAB-q). *Qual Life Res* 2005;14:849-55.
8
9 8 Coyne KS, Thompson CL, Lai JS, et al. An overactive bladder symptom and health-related quality
10 of life short-form: validation of the OAB-q SF. *Neurourol Urodyn* 2015;34:255-63
11 doi:10.1002/nau.22559 [doi].
12
13 9 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically
14 important difference. *Control Clin Trials* 1989;10:407-15.
15
16 10 Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and
17 methods. *J Clin Epidemiol* 2010;63:524-34 doi:10.1016/j.jclinepi.2009.08.010 [doi].
18
19 11 Olsen MF, Bjerre E, Hansen MD, et al. Minimum clinically important differences in chronic pain
20 vary considerably by baseline pain and methodological factors: systematic review of empirical
21 studies. *J Clin Epidemiol* 2018;101:87,106.e2 doi:S0895-4356(18)30124-0 [pii].
22
23 12 van der Worp H, Kollen BJ, Vermist T, et al. Symptom improvement and predictors associated
24 with improvement after 6 weeks of alpha-blocker therapy: An exploratory, single-arm, open-label
25 cohort study. *PLoS One* 2019;14:e0220417 doi:10.1371/journal.pone.0220417 [doi].
26
27 13 Viktrup L, Hayes RP, Wang P, et al. Construct validation of patient global impression of severity
28 (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract
29 symptoms secondary to benign prostatic hyperplasia. *BMC Urol* 2012;12:30,2490-12-30
30 doi:10.1186/1471-2490-12-30 [doi].
31
32 14 Barendrecht MM, Abrams P, Schumacher H, et al. Do alpha1-adrenoceptor antagonists improve
33 lower urinary tract symptoms by reducing bladder outlet resistance?. *Neurourol Urodyn* 2008;27:226-
34 30 doi:10.1002/nau.20481 [doi].
35
36 15 Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J*
37 *Obstet Gynecol* 2003;189:98-101.
38
39 16 Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the
40 significance of outcome effects above the statistical level, with methodological implications for future
41 studies. *J Clin Epidemiol* 2017;82:128-36 doi:S0895-4356(16)30776-4 [pii].
42
43 17 Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related
44 quality of life. *J Clin Epidemiol* 2003;56:395-407 doi:S0895435603000441 [pii].
45
46 18 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life:
47 the remarkable universality of half a standard deviation. *Med Care* 2003;41:582-92
48 doi:10.1097/01.MLR.0000062554.74615.4C [doi].
49
50 19 de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important
51 difference. *J Clin Epidemiol* 2010;63:804,5; author reply 806 doi:10.1016/j.jclinepi.2009.12.015
52 [doi].
53
54 20 Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores:
55 differences between improvement and worsening. *Qual Life Res* 2002;11:207-21.
56
57
58
59
60

1
2
3 21 Guyatt GH, Norman GR, Juniper EF, et al. A critical look at transition ratings. *J Clin Epidemiol*
4 2002;55:900-8 doi:S0895435602004353 [pii].
5

6 22 Guyatt GH, Jaeschke RJ. Reassessing quality-of-life instruments in the evaluation of new drugs.
7 *Pharmacoeconomics* 1997;12:621-6 doi:10.2165/00019053-199712060-00002 [doi].
8

9 23 Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify
10 important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res* 2002;11:1-7.
11

12 24 Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative
13 knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations.
14 *BMJ Open* 2017;7:e015587,2016-015587 doi:10.1136/bmjopen-2016-015587 [doi].
15
16

17 25 Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and
18 minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9
19 doi:10.1016/j.jclinepi.2007.03.012 [doi].
20

21 26 Alma H, de Jong C, Tsiligianni I, et al. Clinically relevant differences in COPD health status:
22 systematic review and triangulation. *Eur Respir J* 2018;52:1800412 doi:10.1183/13993003.00412.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTERNATIONAL-PROSTATE SYMPTOM SCORE (IPSS)

	Not at all	Less than 1 time in 5	Less than half the time	About half the time	More than half the time	Almost always
1. Over the past 4 weeks, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0	1	2	3	4	5
2. Over the past 4 weeks, how often have you had to urinate again less than two hours after you finished urinating?	0	1	2	3	4	5
3. Over the past 4 weeks, how often have you found you stopped and started again several times when you urinated?	0	1	2	3	4	5
4. Over the past 4 weeks, how often have you found it difficult to postpone urination?	0	1	2	3	4	5
5. Over the past 4 weeks, how often has your urinary stream been weaker than usual?	0	1	2	3	4	5
6. Over the past 4 weeks, how often have you had to push or strain to begin urination?	0	1	2	3	4	5
	None	1 time	2 times	3 times	4 times	5 or more times
7. Over the past 4 weeks, how many times, in general, did you get up to urinate from the time you went to bed at night until the time you got up in the morning?	0	1	2	3	4	5

IPSS Quality of life question

	Delighted	Pleased	Mostly satisfied	Mixed - neither satisfied nor dissatisfied	Mostly dissatisfied	Unhappy	Terrible
If you were to spend the rest of your life with your urinary condition just the way it is now, how would you feel about that?	0	1	2	3	4	5	6

OAB-q-SF

This questionnaire asks about how much you have been bothered by selected bladder symptoms during the past 4 weeks. Please place a ✓ or x in the box that best describes the extent to which you were bothered by each symptom during the past 4 weeks. There are no right or wrong answers. Please be sure to answer every question.

During the past 4 weeks, how bothered were you by...

	Not at all	A little bit	Some- what	Quite a bit	A great deal	A very great deal
1. An uncomfortable urge to urinate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. A sudden urge to urinate with little or no warning?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Accidental loss of small amounts of urine?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Nighttime urination?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Waking up at night because you had to urinate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Urine loss associated with a strong desire to urinate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Patient Global Impression of Improvement, PGI-I

Have your symptoms changed since the start of the medication (the moment you completed the previous questionnaire)?

Very much better	Much better	A little better	No change	A little worse	Much worse	Very much worse
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation	Page No
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract	2
		(b) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) If applicable, explain how loss to follow-up was addressed (e) Describe any sensitivity analyses	6-7 6 7
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	7
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	8 8 8
Outcome data	15*	Report numbers of outcome events or summary measures over time	-

1	Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-11
2			(b) Report category boundaries when continuous variables were categorized	7-11
3			(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
4				
5				
6				
7				
8				
9	Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-11
10				
11	Discussion			
12				
13	Key results	18	Summarise key results with reference to study objectives	11
14	Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
15				
16	Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-14
17				
18				
19	Generalisability	21	Discuss the generalisability (external validity) of the study results	11
20				
21	Other information			
22	Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3
23				
24				

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at <http://www.strobe-statement.org>.

BMJ Open

Determining the minimal important differences in the International Prostate Symptom Score and Overactive Bladder Questionnaire: results from an observational cohort study in Dutch primary care

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2019-032795.R2
Article Type:	Original research
Date Submitted by the Author:	25-Nov-2019
Complete List of Authors:	Blanker, Marco; University of Groningen, University Medical Center Groningen, department of General Practice and Elderly Care Medicine Alma, Harma; University of Groningen, University Medical Center Groningen, Department of General Practice and Elderly Care Medicine; University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC) Devji, Tahira; McMaster University Faculty of Health Sciences, Clinical Epidemiology and Biostatistics Roelofs, Marjan; University of Groningen, University Medical Center Groningen, department of General Practice and Elderly Care Medicine Steffens, Martijn ; Isala Hospitals, urology van der Worp, Henk; University of Groningen, University medical center groningen, Department General Practice and Elderly Care Medicine
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Urology, Epidemiology
Keywords:	Lower Urinary Tract Symptoms, Adrenergic alpha-Antagonists, Primary Health Care, Minimal Clinically Important Difference, Patient outcome assessment

SCHOLARONE™
Manuscripts

1
2
3 **Determining the minimal important differences in the International Prostate**
4
5 **Symptom Score and Overactive Bladder Questionnaire: results from an**
6
7 **observational cohort study in Dutch primary care**
8
9

10
11
12 M.H. (Marco) Blanker¹, H.J. (Harma) Alma¹, T. (Tahira) Devji², M. (Marjan) Roelofs¹, M.G.
13
14 (Martijn) Steffens³, H. (Henk) van der Worp¹
15
16

- 17
18
19 1. Department of General Practice and Elderly Care Medicine, University Medical Centre
20
21 Groningen, University of Groningen
22
23 2. Department of Health Research Methods, Evidence and Impact, McMaster University,
24
25 Canada
26
27 3. Department of Urology, Isala Clinics, Zwolle, The Netherlands
28
29
30

31
32 Corresponding author: Marco H. Blanker, MD PhD, general practitioner and epidemiologist,
33
34 Department of General Practice and Elderly Care Medicine, University Medical Centre
35
36 Groningen, University of Groningen, P.O. box 196, 9700 AD Groningen, The Netherlands, Phone:
37
38 +31 6 14961542, Fax: +31 50 3632964
39
40

41
42 E-mail addresses: m.h.blanker@umcg.nl, h.j.alma@umcg.nl, devjits@mcmaster.ca,
43
44 marjanroelofs@hotmail.com, m.steffens@isala.nl, h.van.der.worp@umcg.nl
45
46

47
48 **Wordcount:** 3,116

49
50 **Keywords:** Lower Urinary Tract Symptoms; Adrenergic alpha-Antagonists; Primary
51
52 Health Care; Minimal Clinically Important Difference; Patient outcome assessment
53
54
55
56
57
58
59
60

ABSTRACT

Objectives: To determine the minimal important difference (MID) of the International Prostate Symptom Score (IPSS) and the Overactive Bladder Questionnaire short form (OAB-q SF) assessed in primary care among patients treated for lower urinary tract symptoms (LUTS).

Design: Single-arm, open-label observational cohort study with a 6-week follow-up.

Setting: Twenty-two pharmacies in the Netherlands.

Participants: We enrolled Dutch men with uncomplicated LUTS who received a new alpha-blocker prescription from their general practitioner or urologist.

Primary and secondary outcomes: The IPSS and OAB-q SF were completed before and after 6 weeks of therapy. At 6 weeks, men also completed the Patient Global Impression of Improvement (PGI-I). The mean change scores of the IPSS and OAB-q SF were calculated for each PGI-I outcome category, with the category 'a little better' used to determine the MID. The standard error of measurement (SEM) was calculated for each questionnaire.

Results: In total, 165 men completed follow-up. The MID was 5.2 points (95%CI, 3.9–6.4; SEM 3.6) for the IPSS and 11.0 points (95%CI, 7.1–14.9; SEM 9.7) for the OAB-q SF. For both questionnaires, confidence intervals showed an overlap with the no-change categories. However, the MID for the IPSS was higher in men with severe baseline symptoms (7.1; 95%CI, 5.3–9.0) than in men with moderate baseline symptoms (3.2; 95%CI, 1.7–4.8).

Conclusion: In this study, the MID for the IPSS was considerably higher than the MID of 3.1 reported in the only other study on this topic, but may be due to methodological differences. Interpretation of the MID for the OAB-q SF is hampered by the overlap with the SEM. Future studies are needed to confirm our results because correlations between the PGI-I and symptom questionnaires were suboptimal.

ARTICLE SUMMARY

Strengths and limitations of this study:

- We assessed the minimal important difference (MID) of two frequently used questionnaires

1
2
3 on lower urinary tract symptoms.
4

- 5 • Given that many men are treated in primary care, MID values for this setting are
6 particularly important to inform evidence-based decision-making and to facilitate
7 interpretation of the IPSS and OAB-q SF.
8
9
- 10 • Notably, the sample size of this study was small, which resulted in very low numbers of
11 men being included in the PGI-I category 'very much better' or 'worsening of symptoms',
12 and may clarify the small overlap of the confidence intervals with the no-change category.
13
14
15
16
17
18
19

20 **Funding statement:** The Hein Hogerzeil Foundation supported this work with an unrestricted
21 grant. The Foundation was not involved in the design of this study, nor in the data collection,
22 data analyses, interpretation of the outcomes, or the writing of this manuscript.
23
24
25
26
27
28

29 **Conflict of interest statement:** All authors have completed the ICMJE uniform disclosure form
30 at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the
31 submitted work; no financial relationships with any organisations that might have an interest in
32 the submitted work in the previous three years; no other relationships or activities that could
33 appear to have influenced the submitted work.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Symptom severity is a key outcome for patients with lower urinary tract symptoms (LUTS) and is most often evaluated by direct patient inquiry, using patient-reported outcome measures (PROMs). Although the IPSS is most often used for this purpose in both clinical trials and practice,¹⁻⁵ it fails to capture problematic symptoms such as urinary incontinence and urgency. Therefore, the Overactive Bladder Questionnaire (OAB-q) is increasingly being used to evaluate the treatment of overactive bladder,^{6,7} with the short form (i.e., OAB-q SF) having the advantage of being less time consuming.⁸ Together, both of the IPSS and OAB-q SF capture the spectrum of outcomes that are important to patients, but it is difficult to interpret the effects of an intervention expressed as mean scores or change scores over time.

The minimal important difference (MID) has proven invaluable when interpreting PROMs and could be of great value for both the IPSS and OAB-q SF.^{9,10} To date, the MID has only been reported for the IPSS in a study conducted in secondary care among participants of clinical trials.^{1,2} There has been no report of the MID for the OAB-q SF in any care setting. In countries like the Netherlands and the United Kingdom, most men with LUTS first visit their GP to seek treatment. Given that setting may affect the MID, possibly because of differences in baseline symptom severity,⁹⁻¹¹ we feel that it is important to assess the MID in a primary care setting. To date the MID for secondary care settings has been applied in guidelines for primary care.^{3,5} It is unclear if applying the threshold for a clinically relevant outcome is appropriate. Men who receive treatment need to be aware of what can be expected. Knowledge about the MID in primary care will then provide invaluable data for interpreting treatment outcomes that may differ between primary and secondary care. In addition, evidence must be obtained from multiple studies to ensure that MID determinations are accurate.

In the current study, we aimed to determine the MIDs for both the IPSS and OAB-q SF in a patient cohort originating mainly from primary care.

METHODS

Study design

We conducted a prospective cohort study between January 2016 and April 2018.¹² Baseline data for the IPSS and OAB-q SF were compared with follow-up data after 6 weeks of treatment.

At follow-up, participants also completed the Patient Global Impression of Improvement (PGI-I) and we calculated the MID.

Participants

Adult men who visited a participating pharmacy in the north of the Netherlands were included if they received a new alpha-blocker prescription for uncomplicated LUTS from a GP or urologist. A prescription was defined as new if no alpha-blocker prescription had been given within the past year. The pharmacists checked if the alpha-blocker was indicated for LUTS and excluded men prescribed alpha-blockers for urinary tract stones or indwelling catheters. All participants provided written informed consent. The medical ethics committee of the University Medical Centre Groningen, the Netherlands, approved the study (number 2016.122).

Data collection

At baseline, before starting alpha-blocker therapy, all participants provided relevant descriptive data (e.g., age, duration of LUTS in months or years, and history of surgery for LUTS) and completed the Dutch versions of the IPSS and OAB-q SF. After 6 weeks, men who consented repeated the IPSS and OAB-q SF by postal invitation. At this time, we asked participants to complete the Patient Global Impression of Improvement (PGI-I) questionnaire.¹³ The period of six weeks was chosen as clinical effects of alpha-blockers take a few weeks to develop fully, but significant efficacy over placebo can occur within hours to days.^{4,14}

Questionnaires

The IPSS questionnaire was originally validated as the *American Urological Association Symptom Index for benign prostatic hyperplasia*.¹ It includes 7 questions covering frequency, nocturia, weak urinary stream, hesitancy, intermittence, incomplete emptying and urgency. Each question has response options ranging from 0 to five, with higher scores reflecting more severe symptoms. Total scores that may range from 0 (no symptoms) to 35 points (maximum score),

1
2
3 and scores are often categorized as no/mild symptoms (0–7 points), moderate symptoms (8–19
4 points), or severe symptoms (≥ 20 points). The questionnaire was internally consistent
5
6 (Cronbach's alpha = 0.86) and has excellent test-retest reliability ($r = 0.92$).¹ The MID for the
7
8 IPSS is currently considered to be 3.1 points.² The AUA-SI has been internationally adopted and
9
10 implemented worldwide under the name IPSS.
11

12
13 IPSS focuses on the concept of “benign prostatic hyperplasia” as cause of male LUTS, which
14
15 appeared to have a multifactorial origin. Overactive bladder (OAB) is one of the alternative
16
17 explanations of LUTS. Although urgency (included in the IPSS) relates to OAB, OAB includes
18
19 other symptoms as well, which are not included in the IPSS questionnaire. Therefor, Coyne et al
20
21 developed a condition specific questionnaire, the OAB-q.^{6,7} The OAB-q was developed from
22
23 focus groups of men and women, clinician opinion, and a thorough literature review. More
24
25 recently, this OAB-q has been shortened to benefit patients, researchers and clinicians.⁸ The
26
27 OAB-q SF contains six questions on 6-point Likert-type scales, with the outcomes transformed
28
29 to a 0–100 point scale in which higher scores indicate more severe symptoms.⁸ This scale
30
31 demonstrated good convergent validity, discriminant validity, internal reliability,
32
33 reproducibility, and responsiveness to change.⁸
34
35

36
37 Both IPSS and OAB-q-SF capture symptoms that are not by definition patient important, but
38
39 rather reflect the conditions under study. To study if changes on a questionnaire over time are
40
41 relevant for patients, the PGI-I has been developed using a quantitative approach.^{13,15}
42
43

44 The PGI-I is a validated generic tool for assessing overall improvement after treatment and is
45
46 answered on a 7-point Likert-type scale, with the following options: ‘very much better’, ‘much
47
48 better’, ‘a little better’, ‘no change’, ‘a little worse’, ‘much worse’ or ‘very much worse’.^{13,15} Full
49
50 versions of these questionnaires are presented as supplementary file 1.
51

52 We sent a reminder after two weeks to patients who did not respond to follow-up requests.
53

54 **Data analyses**

55
56 Baseline characteristics are reported as continuous variables and summarized as mean and SD
57
58 or as median and interquartile range, depending on the distribution checked by the Shapiro–
59
60

1
2
3 Wilk test. These characteristics were also compared between men with and without completed
4 follow-up data to test for selective nonresponse. Next, the change scores of the IPSS and OAB-q
5 SF were calculated by comparing the data between the baseline and follow-up questionnaires.
6
7 Change scores were inverted to facilitate intuitive interpretation, with positive scores reflecting
8 symptom improvement.
9
10
11
12

13
14 Various methods exist to determine the MID of questionnaires and are typically either
15 anchor-based or distribution-based.^{9, 16-18} The latter involve evaluating change in the PROM
16 with the probability that the change occurred by chance, sample variation or measurement
17 precision; however, they do not reflect patient perspectives.^{17, 19} Thus, we used an anchor-
18 based method,⁹ in which we compared changes in the IPSS or OAB-q SF (PROM) with the PGI-I
19 (the anchor). For each PGI-I category, we then present the mean change in scores from baseline
20 to follow-up with the associated CIs. We defined the MID as the mean change in IPSS or OAB-q
21 SF for the PGI-I category 'a little better', as the M in MID reflects the *minimal* change that is
22 considered relevant. We also present the mean change scores for the other PGI-I categories.
23
24
25
26
27
28
29
30
31
32

33 The usefulness of anchor-based approaches depends on the relationship between the
34 PROM and the anchor.²⁰⁻²² The anchor and PROM should be measuring the same or similar
35 underlying constructs and should therefore be appreciably correlated. Correlations between
36 questionnaire change scores and the anchor PGI-I should be obviously strong, as else these
37 measure different concepts. Correlations between the anchor PGI-I and the baseline and follow-
38 up questionnaire scores are performed to check for a possible response shift. Mostly anchor
39 PGI-I scores seem correlated with follow-up scores (due to response shift). We therefore
40 examined the Spearman correlation coefficients between the PGI-I and the IPSS and OAB-q SF
41 for the baseline, follow-up and change data to ensure the anchor's validity. A correlation
42 coefficient between the symptom change scores and the PGI-I of ≥ 0.50 , and an equal and
43 opposite correlation of the PGI-I with the baseline score and the follow-up score, were
44 considered ideal and likely to yield trustworthy MID estimates.²⁰⁻²²
45
46
47
48
49
50
51
52
53
54
55
56
57
58

59 To test the impact of baseline symptom severity on the distribution of results, a
60

1
2
3 stratified analysis was performed for the IPSS categories ‘moderate symptoms’ and ‘severe
4 symptoms’ because previous research has shown that such stratification has a large impact.²
5
6 No such categories have been defined for the OAB-q SF, so we did not perform a similar analysis
7
8 for this questionnaire. Subgroup analyses were also performed with participants who received
9
10 their prescription from their GP, allowing us to provide data that focused on the primary care
11
12 setting. Finally, we checked if the MID exceeded the measurement error.^{15,17} For this, we
13
14 calculated the standard error of measurement (SEM) as follows: $[SD \times (1 - \text{reliability})^{1/2}]$.
15
16 Cronbach’s alpha was used as the reliability measure.²³
17
18
19

20 The complete data set was used without imputing missing data. All analyses were
21
22 performed using IBM SPSS version 25.0 (IBM Corp., Armonk, NY, USA), and we considered a p-
23
24 value <0.05 to be statistically significant.
25

26 **Patient and public involvement**

27 This study was performed without patient involvement. We did not invite patients to comment
28
29 on the study design nor did we consult them to interpret the results. Patients were not invited
30
31 to contribute to the writing or editing of this document for readability or accuracy.
32
33
34
35

36 **RESULTS**

37 A total of 251 men completed the baseline questionnaires, of which 165 also completed the
38
39 follow-up questionnaires. The baseline characteristics of men with and without follow-up data
40
41 are shown in Table 1, with no statistically significant differences found between these groups.
42
43 Notably, 86.3% of the participants received their prescription from a GP and the remainder
44
45 received it from a urologist.
46
47
48
49

50 **Table 1.** Baseline characteristics of all participants and participants who dropped out after the
51
52 baseline questionnaire

	Participants with	Drop out after baseline	p-value
	completed	(n = 86)	
	follow-up		

	(n = 165)		
Age (mean ± SD)	66.7±9.7	65.4±12.1	0.42*
Prescription from GP (%)	86.3	83.2	0.55#
IPSS score (mean ± SD)	19.1±6.8	17.6±6.5	0.11*
IPSS categories (%)			0.12#
- none/mild	3.7	6.4	
- moderate	50.9	61.5	
- severe	45.3	32.1	
IPSS Quality of Life (median IQR)	4.0 3.0-5.0	4.0 3.0-5.0	0.52§
OAB-q SF (mean ± SD)	39.8±19.2	40.7±18.1	0.70*
Duration of LUTS in months (median IQR)	24.0 5.0-42.0	12.0 3.0-36.0	0.11§
History of surgery for LUTS (%)	1.2	3.8	0.19#

IQR: interquartile range. P-values refer to: * student-t-test; # chi-squared test; § Mann-Whitney U test.

Abbreviations: OAB-q SF, Overactive Bladder Questionnaire short form.

There were mean improvements in the IPSS and OAB-q SF scores during the study of 5.8 (SD 6.7) and 11.8 points (SD 17.4), respectively. Between baseline and follow-up at 6 weeks, the mean IPSS score changed from 19.1 (SD 6.8) to 13.3 (SD 6.5) and the mean OAB-q SF score changed from 39.7 (SD 19.2) to 27.9 (SD 16.9). The PGI-I outcomes are shown in Table 2 and indicate that most men reported that they were 'a little better' or 'much better' (74.7%), whilst only 23.5% perceived no change. Only three men (1.8%) reported 'worsened' symptoms, and none of the participants reported 'much worsened' or 'very much worsened' symptoms.

Table 2. Change scores for the IPSS and OAB-q SF by PGI-I outcomes

PGI-I outcome	N (%)	IPSS	Missing	OAB-q SF	Missing
Very much better	6 (3.6)	13.4 (2.9;23.9)	1	23.8 (2.3;45.3)	0
Much better	50 (30.3)	8.7 (6.8;10.7)	2	19.1 (14.3;24.0)	3

A little better	68 (41.2)	5.2 (3.9;6.4)	3	11.0 (7.1;14.9)	4
No change	38 (23.0)	3.1 (1.1;5.1)	0	3.0 (-2.3;8.4)	4
A little worse	3 (1.8)	-5.0 (-30.9;20.9)	0	-9.7 (-81.7;62.4)	0

Change in IPSS and OAB-q SF scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better' reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Patient Global Impression of Improvement.

Table 2 also shows the distribution of changes in the IPSS and OAB-q SF for each PGI-I category. The MID for the IPSS was 5.2 points (95% CI, 3.9–6.4) and the PGI-I outcomes 'no change' and 'much better' corresponded to IPSS symptom changes of 3.1 points (95% CI, 1.1–5.1) and 8.7 points (95% CI, 6.8–10.7), respectively. The MID for the OAB-q SF was 11.0 points (95% CI, 7.1 to 14.9) and the PGI-I outcomes 'no change' and 'much better' corresponded with mean improvements of 3.0 points (95% CI, -2.3 to 8.4) and 19.1 points (95% CI, 14.3–24.0), respectively. For both questionnaires, the confidence intervals of the MID-categories showed an overlap with the 'no change' categories.

The Spearman correlation coefficients were then calculated between the PGI-I and both the IPSS and the OAB-q SF. The correlation was -0.51 for the PGI-I and baseline IPSS, 0.43 for the PGI-I and follow-up IPSS and 0.38 for the PGI-I and change in IPSS. The corresponding correlations for the OAB-q SF were -0.09 at baseline, 0.36 at follow-up and 0.42 for the change.

Table 3. Change scores for the IPSS and OAB-q SF by PGI-I outcomes: subgroup analysis for GP prescriptions

PGI-I outcome	N (%)	IPSS	Missing	OAB-q SF	Missing
----------------------	--------------	-------------	----------------	-----------------	----------------

Very much better	4 (3.1)	18.0 (1.7;34.3)	1	30.8 (-0.9;62.4)	0
Much better	39 (30.2)	9.2 (7.0;11.5)	2	19.9 (14.8;24.9)	2
A little better	57 (44.2)	5.4 (4.0;6.7)	3	11.2 (7.0;15.4)	3
No change	27 (20.9)	3.1 (0.5;5.6)	0	3.3 (-3.4;9.9)	3
A little worse	2 (1.6)	-8.5 (-11.6.5; 99.5)	0	-16.5 (-353.2;320.2)	0

Change in IPSS and OAB-q SF scores were estimated by comparing symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better' reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder Questionnaire short form; PGI-I, Patient Global Impression of Improvement.

Subgroup analyses of data for men with a prescription from a GP found no relevant differences, with MID values of 5.4 for the IPSS and 11.2 for the OAB-q SF (Table 3). Stratified analysis of baseline data revealed that men with severe symptoms had higher MID values for the IPSS, reaching 7.1 (95% CI, 5.3–9.0), compared with the MID value of 3.2 (95% CI, 1.7–4.8) for men with moderate symptoms (Table 4).

Table 4. Change scores for the IPSS by PGI-I outcomes

PGI-I outcome	Moderate symptoms (n = 88)			Severe symptoms (n = 73)	
	N (%)	Change	Missing	N (%)	Change
Very much better	3 (3.4)	6.5 (-50.7;63.7)	1	3 (4.1)	18 (1.7;34.3)
Much better	30 (34.1)	5.6 (3.7;7.5)	1	19 (26.0)	13.5 (10.5;16.5)
A little better	33 (37.5)	3.2 (1.7;4.8)	0	32 (43.8)	7.1 (5.3;9.0)
No change	19 (21.6)	1.3 (-1.7;4.3)	0	19 (26.0)	4.9 (2.3;7.6)
A little worse	3 (3.4)	-5.0 (-31.0;20.9)	0	0 (0.0)	-

These results are stratified by baseline symptom severity on the IPSS: moderate symptoms are scores of 8–19 and severe symptoms are scores of ≥ 20 . Change in IPSS scores were estimated by comparing

1
2
3 symptom scores between baseline and 6 weeks. Mean change and 95% CIs are presented. Outcomes are
4 inverted so that positive changes reflect symptom improvement. The PGI-I category 'a little better'
5 reflects the MID for both questionnaires. None of the participants scored 'much worsened' or 'very much
6 worsened' on the PGI-I. Abbreviations: MID, Minimal important difference; OAB-q SF, Overactive Bladder
7 Questionnaire short form; PGI-I, Patient Global Impression of Improvement.
8
9
10
11
12

13
14
15 Finally, the SEM was 3.6 for the IPSS and 9.7 for the OAB-q SF.
16
17

18 19 **DISCUSSION**

20
21 We estimated the MID for two questionnaires that are often used to assess male LUTS in
22 primary care. However, whereas the SEM of the IPSS was less than the 95% CI of the MID (5.2
23 points; 95% CI, 3.9–6.4; SEM 3.6), the SEM of the OAB-q SF fell within the 95% CI of the MID (11
24 points; 95% CI, 7.1–14.9; SEM 9.7). Thus, we can only conclude that the outcomes for the IPSS
25 were unlikely to have occurred because of chance or measurement imprecision. Given that
26 many questionnaires have used multiple MID values, we were surprised to find only one
27 previous estimate of the MID for the IPSS in the literature.^{1,2} Our study therefore adds relevant
28 information in the primary care setting for clinicians and guideline developers.
29
30
31
32
33
34
35
36
37

38
39 Our results for the IPSS were different to those of the seminal study on this topic
40 performed by Barry et al. in secondary care.^{1,2} In that study, the MID of 3.1 points (SD 0.27) fell
41 within the 95% CI of the 'no change' group (consisting of men who expressed that they hadn't
42 experienced any change in symptoms), but outside the CI of the 'a little better' group, suggesting
43 a likely underestimation of the real value. In the current study, there was also some overlap
44 between the CIs of the 'no change' and the 'a little better' group, though this was within a change
45 of only 3.9 to 5.1 points. This could be explained by the relative small samples in the subgroup
46 analyses. Given that treatment is typically in primary care, we have therefore provided
47 additional data that is applicable to most men with LUTS. Nevertheless, the differences in
48 outcomes compared with the study by Barry et al. need to be explained. It is our contention that
49 three methodological differences account for these differences.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 First, Barry et al. compared patients between baseline and follow-up after 13 weeks. By
4 contrast, the follow-up period in the current study was only 6 weeks. Although this difference of
5 7 weeks may have affected the ability of patients to recall their prior health state accurately, the
6 true impact of this remains unclear. Change scores may also have been influenced by the natural
7 variation that occurs in symptom severity over time.
8
9

10
11
12
13
14 Second, in the research by Barry et al., a different global assessment of patient
15 improvement was used. This included a 5-point scale with the options 'marked improvement',
16 'moderate improvement', 'slight improvement', 'no improvement', and 'worse' for which the
17 exact question was not reported. In our study, we used a 7-point Likert-type scale that ranged
18 from 'very much better' to 'very much worsened'. We considered that this difference probably
19 had no more than a marginal impact given that the positive outcome categories were
20 comparable in both studies. Notably, none of the participants in our study reported that the
21 symptoms had 'very' or 'very much' worsened.
22
23
24
25
26
27
28
29

30
31 Third, we mainly included men from primary care, rather than men solely from
32 secondary care. Although it is generally thought that men in primary care have fewer symptoms,
33 our men tended to have more severe symptoms (IPSS score >19) than in the study by Barry et
34 al. (45% versus 25%). Barry et al. also reported that baseline severity had a major impact on the
35 MID, but when we compare their stratified analysis with ours, we had higher mean change
36 scores for each PGI-I category. This might be explained by the fact that we only included men
37 who actually used an alpha-blocker. In contrast to this focused approach, Barry et al. used data
38 for all participants in a large, randomized, double-blind trial of four treatment strategies for
39 male LUTS. In their study, a lower MID could therefore have resulted from the inclusion of
40 patients receiving placebo, finasteride, terazosin or combination therapy given that the efficacy
41 of alpha-blockers exceeds that of both placebo and finasteride. The use of blinding meant that
42 men who used placebo or finasteride may have overestimated their subjective improvement,
43 whilst alpha-blocker users may have underestimated their subjective improvement. Given that
44 the IPSS objectively counts symptoms, the placebo and finasteride users would experience a
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 smaller change in the IPSS whereas the active drug users would experience a larger effect.

4
5 Although the actual impact of each intervention is unknown, researchers in other fields have
6
7 made similar observations.^{10, 11, 24}

8
9 We were unable to find any prior estimates of the MID for the OAB-q SF in the literature.
10
11 Our finding that 11.0 points (95% CI, 7.1–14.9) indicates a clinically relevant change is therefore
12
13 a novel and important finding, but one for which the reliability will need to be assessed in other
14
15 studies. We recognize that alpha-blockers are not specifically indicated for the treatment of
16
17 overactive bladder, but we contend that there is a considerable overlap with LUTS unrelated to
18
19 overactive bladder. Indeed, guidelines suggest prescribing alpha-blockers for most men with
20
21 LUTS who request active treatment. This is because these agents have a rapid onset of action,
22
23 good efficacy, and low rate and severity of adverse events.³⁻⁵ We recommend further study to
24
25 determine the MID in men with specific symptoms of overactive bladder treated with
26
27 anticholinergics or beta-3 agonists.
28
29

30
31 Some limitations need to be considered when assessing our results. Notably, the sample
32
33 size of this study was small, which resulted in very low numbers of men being included in the
34
35 PGI-I category 'very much better'. For that category, the mean change scores for both
36
37 questionnaires showed very wide CIs. The same holds for the categories linked to symptom
38
39 worsening. In those categories, a discontinuation trial, in which men stop their treatment, may
40
41 be more suitable for reliable estimates. The sample size might also explain why the CI of the
42
43 MID estimate for the OAB-q SF included the SEM.
44
45

46
47 Another limitation is reflected by difficulties we encountered with some of the
48
49 associations between the PGI-I and the two PROM questionnaires. For the IPSS, the follow-up
50
51 IPSS and PGI-I scores correlated better than with IPSS change and PGI-I scores, suggesting that
52
53 this rating only reflected the current status, which in turn, decreases confidence in the MID
54
55 estimate. For the OAB-q SF, the correlation coefficient between the baseline OAB-q SF and PGI-I
56
57 scores was opposite in magnitude to that for the follow-up OAB-q SF and PGI-I scores. With both
58
59 questionnaires, the correlation coefficients for the change scores were lower than the threshold
60

1
2
3 of 0.5 that we set a priori.²¹ High correlation coefficients are preferred between the anchor and
4 the change in PROM, though some researchers have suggested applying lower thresholds.²⁵
5
6 Still, even the high correlation coefficients are insufficient to confirm that the transition rating is
7 in fact measuring change as opposed to current health status.²¹ Unfortunately, Barry et al. did
8 not report the correlation coefficients between the IPSS and the anchor,² which is consistent
9 with most other research for PROMs.²⁶ Given the suboptimal relationship between the PROM
10 and the anchor, we must stress that the estimates obtained for the MID should be interpreted
11 with caution and should be confirmed in future investigations with larger samples.
12
13
14
15
16
17
18
19

20 In conclusion, this study is the first to define MID values for two important PROMs used
21 to evaluate the effectiveness of treatment for male LUTS in primary care. Given that many men
22 are treated in primary care, MID values for this setting are particularly important to inform
23 evidence-based decision-making and to facilitate interpretation of the IPSS and OAB-q SF.
24 Moreover, we consider that this study emphasizes the importance of the MID to individual
25 patients in daily practice. We defined the MID based on the PGI-I outcome 'a little better' in the
26 present study, but patients may expect 'much better' as an outcome when starting therapy. To
27 date, most outcomes of alpha-blocker and other drug treatments for male LUTS have been
28 expressed as the mean IPSS change scores. In the vast majority of studies³, difference in IPSS
29 changes between active treatment and placebo have approached, but not exceeded, the
30 previously reported MID of 3.1.^{1,2} Applying a threshold for improvement of 5.2 points, as
31 described in our study, may change the interpretation of those studies.
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 **Acknowledgements**

49 We are grateful for the assistance of Tom Vermist in collecting the data. The authors thank all
50 patients and collaborating pharmacies for their participation, as well as Dr Robert Sykes
51 (www.doctored.org.uk) for providing editorial services.
52
53
54
55

56 **Author contributions**

57 The corresponding author attests that all listed authors meet authorship criteria and that no
58
59
60

1
2
3 others meeting the criteria have been omitted.

4
5 MHB and MGS initiated the study. Acquisition of the data was done by MHB and MR. Analysis
6
7 and interpretation of the data was done by MHB, HJA, TD and HW. MHB wrote the manuscript
8
9 with extensive support from HJA and TD.

10
11 All authors critically reviewed the manuscript.

12 13 **Data sharing statement**

14
15
16 Data collected for this study will be available from the corresponding author upon request.

17 18 19 **REFERENCES**

20
21
22
23 1 Barry MJ, Fowler FJ, Jr, O'Leary MP, et al. The American Urological Association symptom index
24 for benign prostatic hyperplasia. The Measurement Committee of the American Urological
25 Association. *J Urol* 1992;148:1549,57; discussion 1564.

26
27 2 Barry MJ, Williford WO, Chang Y, et al. Benign prostatic hyperplasia specific health status
28 measures in clinical research: how much change in the American Urological Association symptom
29 index and the benign prostatic hyperplasia impact index is perceptible to patients?. *J Urol*
30 1995;154:1770-4.

31
32 3 National Clinical Guideline Centre (NICE). Lower urinary tract symptoms in men: management
33 (Clinical guideline CG97) 2015;CG97
34 doi:<http://www.nice.org.uk/nicemedia/live/12984/48554/48554.pdf>.

35
36 4 Gratzke C, Bachmann A, Descazeaud A, et al. EAU Guidelines on the Assessment of Non-
37 neurogenic Male Lower Urinary Tract Symptoms including Benign Prostatic Obstruction. *Eur Urol*
38 2015;67:1099-109 doi:10.1016/j.eururo.2014.12.038 [doi].

39
40 5 Blanker MH, Klomp MA, van den Donk M, et al. Summary of the NHG practice guideline 'Lower
41 urinary tract symptoms in men'. *Ned Tijdschr Geneesk* 2013;157:A6178.

42
43 6 Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and
44 health-related quality of life questionnaire: the OAB-q. *Qual Life Res* 2002;11:563-74.

45
46 7 Coyne KS, Matza LS, Thompson CL. The responsiveness of the Overactive Bladder Questionnaire
47 (OAB-q). *Qual Life Res* 2005;14:849-55.

48
49 8 Coyne KS, Thompson CL, Lai JS, et al. An overactive bladder symptom and health-related quality
50 of life short-form: validation of the OAB-q SF. *Neurourol Urodyn* 2015;34:255-63
51 doi:10.1002/nau.22559 [doi].

52
53 9 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically
54 important difference. *Control Clin Trials* 1989;10:407-15.

55
56 10 Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and
57 methods. *J Clin Epidemiol* 2010;63:524-34 doi:10.1016/j.jclinepi.2009.08.010 [doi].
58
59
60

- 1
2
3 11 Olsen MF, Bjerre E, Hansen MD, et al. Minimum clinically important differences in chronic pain
4 vary considerably by baseline pain and methodological factors: systematic review of empirical
5 studies. *J Clin Epidemiol* 2018;101:87,106.e2 doi:S0895-4356(18)30124-0 [pii].
6
7
8 12 van der Worp H, Kollen BJ, Vermist T, et al. Symptom improvement and predictors associated
9 with improvement after 6 weeks of alpha-blocker therapy: An exploratory, single-arm, open-label
10 cohort study. *PLoS One* 2019;14:e0220417 doi:10.1371/journal.pone.0220417 [doi].
11
12 13 Viktrup L, Hayes RP, Wang P, et al. Construct validation of patient global impression of severity
13 (PGI-S) and improvement (PGI-I) questionnaires in the treatment of men with lower urinary tract
14 symptoms secondary to benign prostatic hyperplasia. *BMC Urol* 2012;12:30,2490-12-30
15 doi:10.1186/1471-2490-12-30 [doi].
16
17 14 Barendrecht MM, Abrams P, Schumacher H, et al. Do alpha1-adrenoceptor antagonists improve
18 lower urinary tract symptoms by reducing bladder outlet resistance?. *Neurourol Urodyn* 2008;27:226-
19 30 doi:10.1002/nau.20481 [doi].
20
21 15 Yalcin I, Bump RC. Validation of two global impression questionnaires for incontinence. *Am J*
22 *Obstet Gynecol* 2003;189:98-101.
23
24 16 Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the
25 significance of outcome effects above the statistical level, with methodological implications for future
26 studies. *J Clin Epidemiol* 2017;82:128-36 doi:S0895-4356(16)30776-4 [pii].
27
28 17 Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related
29 quality of life. *J Clin Epidemiol* 2003;56:395-407 doi:S0895435603000441 [pii].
30
31 18 Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life:
32 the remarkable universality of half a standard deviation. *Med Care* 2003;41:582-92
33 doi:10.1097/01.MLR.0000062554.74615.4C [doi].
34
35 19 de Vet HC, Terwee CB. The minimal detectable change should not replace the minimal important
36 difference. *J Clin Epidemiol* 2010;63:804,5; author reply 806 doi:10.1016/j.jclinepi.2009.12.015
37 [doi].
38
39 20 Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores:
40 differences between improvement and worsening. *Qual Life Res* 2002;11:207-21.
41
42 21 Guyatt GH, Norman GR, Juniper EF, et al. A critical look at transition ratings. *J Clin Epidemiol*
43 2002;55:900-8 doi:S0895435602004353 [pii].
44
45 22 Guyatt GH, Jaeschke RJ. Reassessing quality-of-life instruments in the evaluation of new drugs.
46 *Pharmacoeconomics* 1997;12:621-6 doi:10.2165/00019053-199712060-00002 [doi].
47
48 23 Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify
49 important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res* 2002;11:1-7.
50
51 24 Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative
52 knee disease outcomes: a systematic review and case study to inform BMJ Rapid Recommendations.
53 *BMJ Open* 2017;7:e015587,2016-015587 doi:10.1136/bmjopen-2016-015587 [doi].
54
55
56
57
58
59
60

1
2
3 25 Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and
4 minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008;61:102-9
5 doi:10.1016/j.jclinepi.2007.03.012 [doi].
6

7 26 Alma H, de Jong C, Tsiligianni I, et al. Clinically relevant differences in COPD health status:
8 systematic review and triangulation. *Eur Respir J* 2018;52:1800412 doi:10.1183/13993003.00412.
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

INTERNATIONAL-PROSTATE SYMPTOM SCORE (IPSS)

	Not at all	Less than 1 time in 5	Less than half the time	About half the time	More than half the time	Almost always
1. Over the past 4 weeks, how often have you had a sensation of not emptying your bladder completely after you finished urinating?	0	1	2	3	4	5
2. Over the past 4 weeks, how often have you had to urinate again less than two hours after you finished urinating?	0	1	2	3	4	5
3. Over the past 4 weeks, how often have you found you stopped and started again several times when you urinated?	0	1	2	3	4	5
4. Over the past 4 weeks, how often have you found it difficult to postpone urination?	0	1	2	3	4	5
5. Over the past 4 weeks, how often has your urinary stream been weaker than usual?	0	1	2	3	4	5
6. Over the past 4 weeks, how often have you had to push or strain to begin urination?	0	1	2	3	4	5
	None	1 time	2 times	3 times	4 times	5 or more times
7. Over the past 4 weeks, how many times, in general, did you get up to urinate from the time you went to bed at night until the time you got up in the morning?	0	1	2	3	4	5

IPSS Quality of life question

	Delighted	Pleased	Mostly satisfied	Mixed - neither satisfied nor dissatisfied	Mostly dissatisfied	Unhappy	Terrible
If you were to spend the rest of your life with your urinary condition just the way it is now, how would you feel about that?	0	1	2	3	4	5	6

OAB-q-SF

This questionnaire asks about how much you have been bothered by selected bladder symptoms during the past 4 weeks. Please place a or X in the box that best describes the extent to which you were bothered by each symptom during the past 4 weeks. There are no right or wrong answers. Please be sure to answer every question.

During the past 4 weeks, how bothered were you by...

	Not at all	A little bit	Some- what	Quite a bit	A great deal	A very great deal
1. An uncomfortable urge to urinate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. A sudden urge to urinate with little or no warning?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Accidental loss of small amounts of urine?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Nighttime urination?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Waking up at night because you had to urinate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Urine loss associated with a strong desire to urinate?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Patient Global Impression of Improvement, PGI-I

Have your symptoms changed since the start of the medication (the moment you completed the previous questionnaire)?

Very much better	Much better	A little better	No change	A little worse	Much worse	Very much worse
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation	Page No
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	2 2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	4-5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-6
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-6
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	-
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	6
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) If applicable, explain how loss to follow-up was addressed (e) Describe any sensitivity analyses	6-7 6 7
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	7
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	8 8 8
Outcome data	15*	Report numbers of outcome events or summary measures over time	-

1	Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	7-11
2			(b) Report category boundaries when continuous variables were categorized	7-11
3			(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	
4				
5				
6				
7				
8				
9	Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	7-11
10				
11	Discussion			
12				
13	Key results	18	Summarise key results with reference to study objectives	11
14	Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
15				
16	Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	11-14
17				
18				
19	Generalisability	21	Discuss the generalisability (external validity) of the study results	11
20				
21	Other information			
22	Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	3
23				
24				

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at <http://www.strobe-statement.org>.