Published in final edited form as:

J Urol. 2019 October ; 202(4): 770–778. doi:10.1097/JU.0000000000000310.

Can 7- or 30-day recall questions capture self-reported lower urinary tract symptoms accurately?

Kathryn E. Flynn, PhD¹, Sarah A. Mansfield, MS², Abigail R. Smith, PhD², Brenda W. Gillespie, PhD³, Catherine S. Bradley, MD MCSE⁴, David Cella, PhD⁵, J. Quentin Clemens, MD FACS MSCI³, Margaret E. Helmut, MA², H. Henry Lai, MD⁶, Ziya Kirkali, MD⁷, Pooja Talaty, MS, MHA⁸, Kevin P. Weinfurt, PhD⁹, LURN Study Group

¹Medical College of Wisconsin, Milwaukee, WI

²Arbor Research Collaborative for Health, Ann Arbor, MI

³University of Michigan, Ann Arbor, MI

⁴University of Iowa Carver College of Medicine, Iowa City, IA

⁵Northwestern University, Chicago, IL

⁶Washington University in St Louis, St Louis, MO

⁷National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD

⁸NorthShore University Health System, Glenview, IL

⁹Duke University Medical Center, Durham, NC

Abstract

Purpose: Self-reported measurement tools often provide a recall period, e.g., “In the past 7 days...” For lower urinary tract symptoms (LUTS), concordance between end-of-day (daily) reports and 7- and 30-day recalled reports is unknown. We evaluated how accurately 7- or 30- day recall questions capture LUTS.

Materials and Methods: Participants (261 women, 254 men) were recruited from 6 U.S. tertiary care sites. We evaluated 18 items representing 7 symptoms covering storage, voiding, and post-micturition. Item responses on the daily forms were averaged over a 7- or 30-day period and compared to the corresponding 7-day or 30-day recall version of the item. Analyses were item- and gender-specific. Within-person concordance was assessed using Pearson’s correlation, and bias (systematic over- or under-reporting) was calculated as the difference between the recalled item and the averaged daily item score, reported as a percent of the item scale.

Results: All correlations exceeded 0.60. Correlations between averaged daily reports and recalled reports ranged from 0.72–0.89 (7-day) and 0.71–0.91 (30-day) among women, and from 0.68–0.90 (7-day) and 0.68–0.95 (30-day) among men. Most items did not show systematic bias; the median percent bias did not exceed 10% for any item; however, bias exceeding $\pm 10\%$ was observed in a subset of individuals for some items.

Conclusions: Recalled reports over both 7 and 30 days tracked well with averaged daily reports for men and women, and systematic bias was minimal, suggesting 7- and 30-day recall periods for self-reported LUTS are reasonable.

Keywords

lower urinary tract symptoms; humans; urination; mental recall; self-report; bias

Introduction

Self-reported measurement tools are used to characterize patients with lower urinary tract symptoms (LUTS) and to measure treatment outcomes.¹ Questionnaire items usually refer to a time period over which respondents are asked to recall their experience (e.g., “In the past 7 days...”). Questionnaires for LUTS use a variety of recall periods, from 7 days (e.g., LUTS Tool^{2,3}) to 4 weeks (e.g., American Urological Association Symptom Index [AUA-SI],⁴ International Consultation on Incontinence Modular Questionnaire [ICIQ-LUTS],^{5,6} and Overactive Bladder Symptom and Health-Related Quality of Life Questionnaire [OAB-q]⁷), while other measures ask patients to report on their experiences without reference to a recall period (e.g., “Do you experience any urinary incontinence?” in the Urogenital Distress Inventory⁸).

There is no gold standard for the recall period in a self-reported tool. An appropriate recall period balances the likelihood of accurate recall over a desired time period with patient burden if repeated assessments are needed.⁹ Voiding diaries—assessing voiding frequency, urgency, incontinence, intake and output volumes, etc.—have a very short (or no) recall period. Longitudinal studies with multiple assessments of daily recall periods may place undue burden on patients and increase study costs. On the other hand, longer recall periods may yield inaccurate reports due to recall biases. It is not known how well people can remember their LUTS experiences over different recall periods. Previous studies have suggested poor concordance between voiding diaries and recalled reports of particular symptoms, but in most of these studies, patients recalled experience was collected as part of their medical history, that is, *before* completing a voiding diary.^{10–14} No study has empirically tested the concordance of daily reports of several different symptoms with 7- and 30-day recall of those symptoms over the same reporting period in a large sample of men and women with LUTS.

The Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN) is a 6-site collaboration supported by the U.S. National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).¹⁵ The LURN Recall Study was designed to inform appropriate recall periods for LURN questionnaires, including the LURN Comprehensive Assessment of Self-reported Urinary Symptoms (CASUS) for women and men.¹⁶ For a subset of LURN CASUS items, we evaluated the correspondence between 1) averaged daily recall over 7 days and weekly recall of self-reported LUTS and 2) averaged daily recall over 30 days and monthly recall of self-reported LUTS. We also sought to understand the cognitive heuristics (mental shortcuts) that people might use to construct their weekly and monthly reports of

LUTS, e.g., reporting most severe or most recent experience instead of the average, as has been studied in recall of other experiences.^{17–20}

Methods

Study design and population

Women and men 18 years and older were recruited from the 6 LURN sites between May 2017 and April 2018. Participants were required to understand English and be able to complete self-reported questionnaires electronically. Exclusion criteria included cognitive impairment, planned changes in LUTS treatment (medication or surgery) during the 30-day study time frame, current treatment for any malignancy, general or spinal/epidural anesthesia in the past 3 months, and pregnancy or childbirth in the past 6 months. All participants provided informed consent. The study was approved by the Institutional Review Boards of each participating site and the Data Coordinating Center.

We screened potential participants for 7 symptoms from 3 symptom groups: storage (daytime frequency, nocturia, urgency, incontinence), voiding (slow/weak stream), and post-micturition (incomplete emptying, post-micturition dribble). We enrolled patients if they endorsed at least one symptom as at least moderately severe and bothersome (scores of 3, 4, or 5 on 5-point scales for both severity and bother) in the past 2 weeks as well as in the past 3 months. We monitored enrollment based on these screening criteria to ensure adequate representation for each symptom. Based on our prior experience and published data,^{21–23} we expected most participants to contribute data on multiple symptoms.

We randomly assigned participants to one of three groups (1, 2A, 2B) in a 2:1:1 ratio (Table 1). All groups completed a baseline assessment, 7-day recall assessments at the end of each 7-day period, and a 30-day recall assessment at the end of the 30-day study. In addition, Group 1 completed 30 24-hour recall assessments at the end of each day during the study. To test whether daily reporting of symptoms might influence the level of symptoms recalled during the 7-day recall period, participants in groups 2A and 2B each completed only one week of 24-hour recall assessments. Participants randomized into group 2A completed a 3-day voiding diary and a 3-day recall in week 1, followed by 7 24-hour recall assessments in week 2, while the order of these activities in weeks 1 and 2 were switched for participants in group 2B (diary data to be published separately). This study design allowed us to examine whether 7-day recall differed with or without prior daily assessments. We also assessed variability in individual responses to identify items that are consistent and those that may vary from day-to-day.

At baseline we collected the full set of LURN CASUS items using a 7-day recall period, sociodemographic (sex, race, ethnicity, marital status, education, employment) and self-reported health variables (current health status, medical history, functional limitations). We collected daily, weekly, and monthly recalled assessments for a subset of CASUS items covering the 7 selected symptoms using a 24-hour, 7-day, or 30-day recall period, respectively (Supplemental Table 1).

Statistical Analysis

For each item, analysis included all participants who endorsed at least mild symptom severity on at least one assessment during the study period. We averaged the 24-hour recall responses for each item over either 7-day or 30-day intervals and compared it to the corresponding item on the 7-day or 30-day recall form. All participants were included in the 7-day analyses; for group 1 participants, the first week of data was used. Participants with fewer than 5 completed 24-hour recall forms or an incomplete 7-day recall form were excluded ($n=21$, 4%). The 30-day analyses included group 1 participants who had at least three weeks with five or more 24-hour recall forms completed per week and a complete 30-day recall form (excluding $n=6$, 2%).

We assessed concordance of recall using correlation and bias. Pearson correlation coefficients were used to evaluate the within-person consistency between averaged 24-hour recall over 7 (or 30) days and 7-day (or 30-day) recall. Bias was calculated as the difference in scores between the recalled item and the averaged daily items. We defined positive bias as systematic over-reporting of 7- or 30-day recall compared to averaged daily recall, and negative bias as systematic under-reporting. To compare bias across items with different scales, we reported bias as a percent of the total question scale (i.e., a bias of 1-unit for an item on a 1–5 scale is reported as 20%). We investigated variability in group 1 participants' daily item reports over 30 days using the standard deviation and the range. We defined an item as 'highly variable' if 40–60% of the responses were 'Yes' for Yes/No items or had a range spanning 40% of the item scale for ordinal items. To assess whether recalled symptom levels differed when preceded by daily reporting of symptoms, we conducted paired t-tests between the 7-day recalled values that were and were not preceded by a week of daily reporting, separately for each sex and item for participants in groups 2A and 2B, using weeks 2 and 4 for group 2A, and weeks 1 and 3 for group 2B. To explore evidence of heuristics used to recall items, we calculated the correlations between the 7- and 30-day recall items and the most recent and the highest severity 24-hour recall item.

Our sample size calculations were based on precision of estimation (measured as length of the 95% confidence interval) for both bias (mean difference) and correlation between averaged daily and weekly (or monthly) scores in each participant. We pre-specified that we would interpret correlations higher than 0.70 as "good" and higher than 0.50 as "acceptable" when weighing other considerations for selecting a recall period. Likewise, we considered absolute bias less than 5% as "good" and less than 10% as "acceptable". A sample size of 100–150 for each symptom would produce confidence intervals narrow enough to rule out substantially undesirable values, such as absolute bias of more than 10% (assuming the true bias 5%) and correlations of less than 0.40 (assuming the true correlation is at least 0.50). We assumed separate analyses for men and women and dropout of 10% or less, based on previous experience in a similar study.²⁴ All analyses were conducted using SAS software, Version 9.4 (SAS Institute Inc., 2013, Cary, NC).

Results

Of 569 enrollees, 90% (261 women and 254 men) completed the study. The mean age was 63 years for men and 56 years for women (Supplemental Table 2). Most were white, non-

Hispanic, married, with some college education or more; less than half were employed. Symptom burden was high; all symptoms were reported by at least 50% of the sample at some point during the 30-day study, and more than 90% of participants reported nocturia and/or urgency during the study.

Correlation, bias, and item variability

Estimated correlations between averaged daily reports and 7- and 30-day recalled reports were all greater than 0.50 (Figure 1, Supplemental Tables 3–4). Among women, the correlations between averaged daily reports and recalled reports ranged from 0.72–0.89 (7-day) and 0.71–0.91 (30-day). Correlations for men ranged from 0.68–0.90 (7-day) and 0.68–0.95 (30-day). Only items on nocturia for men had correlations below 0.70.

Most items did not show systematic bias (Figure 2, Supplemental Tables 3–4), and the median percent bias did not exceed $\pm 10\%$ for any item. For two items the median percent bias exceeded $\pm 5\%$: urgency (feeling a sudden need to urinate) at 30 days and urgency incontinence (leaking after feeling a sudden need) at 7 days, both among women. For several items the distribution of bias exceeded $\pm 10\%$ (i.e., 25th percentile $< -10\%$ or 75th percentile $> +10\%$), meaning some men and women over-reported symptoms (nocturia, urgency, slow/weak stream, incomplete emptying, and post-micturition dribble) and some under-reported symptoms (urgency incontinence and nocturia) at recall compared to their averaged daily reports. Some women also under-reported stress incontinence (specifically leaking while laughing, sneezing, or coughing) and incontinence for unknown reasons at 7 days and over-reported stress incontinence (physical activity) at 30 days.

Overall day-to-day item variability was higher among women compared to men. Daytime frequency, urgency, and slow/weak stream symptom items had the most variability, while incontinence items had the least (Table 2).

Effect of daily recall on weekly recall

Across all the items there were no significant differences between 7-day recall from a week with prior daily reporting vs. a week without prior daily reporting in the same participant. The mean percent difference (week with daily reports vs. week with no daily reports) for each item ranged from -1% to 5% for women and -3% to 4% for men (Supplemental Table 5).

Evidence for Use of Cognitive Heuristics

A 7- or 30-day recalled report appeared to reflect the average of the preceding days more than the most recent or most severe report during the same time period (Figures 3 and 4). Correlations were the same or higher between 7- or 30-day recalled reports and averaged daily reports compared to either the most recent daily report or the most severe daily report. The sole exception was leaking urine while laughing, sneezing, or coughing for men, where the correlation was highest between the 7-day report and the most recent daily report; however, both correlations (averaged and most recent) were above 0.8 and the confidence intervals overlapped.

Discussion

The goal of this study was to assess the similarity between the average of daily reports of LUTS and recall of LUTS experiences over 7 and 30 days. This information can be used to choose a recall period with confidence that symptoms can be remembered over the chosen period. In a large sample of men and women who presented for specialty care with at least one symptomatic and bothersome LUTS, all of the items across both recall periods met our pre-specified criterion for “acceptable” correlation between averaged daily and either 7-day or 30-day scores (>0.5) and absolute bias ($<10\%$), and the vast majority met the higher criterion for “good” correlation (>0.7) and absolute bias ($<5\%$).

To further understand how participants might construct their memories of LUTS, we assessed whether the 7- or 30-day recalled values most closely reflected the daily averaged symptom experience, most recent experience, or worst (most severe) experience. We found that participants’ 7- and 30-day recalled reports best reflected the average of the preceding days. Thus, not only are the recalled reports reasonably accurate proxies for the average daily experience; they are not influenced by the most recent experience or the worst experiences. This finding should give both researchers and clinicians confidence that 7- and 30-day recalled reports of LUTS reflect the average across the entire reporting period.

One previous study suggested poor concordance between voiding diaries and recalled reports of urge incontinence among women²⁵ and other studies have suggested over-reporting of daytime frequency by women¹⁰ and of nocturia by men¹² during medical histories compared to voiding diaries. In our study, the averaged daily values reported were consistent with later 7- and 30-day recalled reports; however, we cannot rule out the possibility that for some items and/or for some participant subgroups, the recall may be under- or over-estimated even at a daily level, leading to consistent but inaccurate estimates. In our study, some symptoms, including nocturia and urgency incontinence, had higher proportions of individuals with $>10\%$ bias despite being relatively unbiased on average. This may be related to the variability of symptoms, as has been shown in pain,¹⁹ or to other characteristics such as symptom bother or patient anxiety; we will evaluate these relationships in future work.

Our study was large, diverse with respect to sex, age, and symptoms, with minimal drop-out over the study time frame. The innovative study design allowed us to answer multiple research questions. We conclude that recalled reports tracked well with averaged daily reports for men and women and systematic bias was minimal, suggesting that the use of either 7- or 30-day recall periods for self-reported LUTS is reasonable. For many contexts, this should reduce patient burden. For studies with targeted LUTS, tailored study design/ measurement may be useful, as a small degree of bias was noted in some symptoms. We found no evidence that daily reporting influenced the level of symptom recall. The main limitation of this work was that it was conducted in a selected sample of English-speaking, generally well-educated white, non-Hispanic participants from specialty care clinics and does not represent all people who experience LUTS.

Conclusions

For LUTS patient-reported questionnaires, both 7- and 30-day recall give results reflecting the average of the preceding 7- or 30-day reports among men and women. This supports the ability of people to recall LUTS symptoms up to a 30-day period, with minimal bias due to recent or severe events during the period.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This is publication number 21 of the Symptoms of Lower Urinary Tract Dysfunction Research Network (LURN).

Research reported in this publication was supported at Northwestern University, in part, by the National Institutes of Health's National Center for Advancing Translational Sciences, Grant Number UL1TR001422. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The following individuals were instrumental in the planning and conduct of this study at each of the participating institutions:

Duke University, Durham, North Carolina (DK097780): PI: Cindy Amundsen, MD, Kevin Weinfurt, PhD; Co-Is: Kathryn Flynn, PhD, Matthew O. Fraser, PhD, Todd Harshbarger, PhD, Eric Jelovsek, MD, Aaron Lentz, MD, Drew Peterson, MD, Nazema Siddiqui, MD, Alison Weidner, MD; Study Coordinators: Carrie Dombeck, MA, Robin Gilliam, MSW, Akira Hayes, Shantae McLean, MPH

University of Iowa, Iowa City, IA (DK097772): PI: Karl Kreder, MD, MBA, Catherine S Bradley, MD, MSCE, Co-Is: Bradley A. Erickson, MD, MS, Susan K. Lutgendorf, PhD, Vince Magnotta, PhD, Michael A. O'Donnell, MD, Vivian Sung, MD; Study Coordinator: Ahmad Alzubaidi

Northwestern University, Chicago, IL (DK097779): PIs: David Cella, Brian Helfand, MD, PhD; Co-Is: James W Griffith, PhD, Kimberly Kenton, MD, MS, Christina Lewicky-Gaupp, MD, Todd Parrish, PhD, Jennie Yufen Chen, PhD, Margaret Mueller, MD; Study Coordinators: Sarah Buono, Maria Corona, Beatriz Menendez, Alexis Siurek, Meera Tavathia, Veronica Venezuela, Azra Muftic, Pooja Talaty, Jasmine Nero. Dr. Helfand, Ms. Talaty, and Ms. Nero are at NorthShore University HealthSystem.

University of Michigan Health System, Ann Arbor, MI (DK099932): PI: J Quentin Clemens, MD, FACS, MSCI; Co-Is: Mitch Berger, MD, PhD, John DeLancey, MD, Dee Fenner, MD, Rick Harris, MD, Steve Harte, PhD, Anne P. Cameron, MD, John Wei, MD; Study Coordinators: Morgen Barroso, Linda Drnek, Greg Mowatt, Julie Tumbarello

University of Washington, Seattle Washington (DK100011): PI: Claire Yang, MD; Co-I: John L. Gore, MD, MS; Study Coordinators: Alice Liu, MPH, Brenda Vicars, RN

Washington University in St. Louis, St. Louis Missouri (DK100017): PI: Gerald L. Andriole, MD, H. Henry Lai; Co-I: Joshua Shimony, MD, PhD; Study Coordinators: Susan Mueller, RN, BSN, Heather Wilson, LPN, Deborah Ksiazek, BS, Aleksandra Klim, RN, MHS, CCRC

National Institute of Diabetes and Digestive and Kidney Diseases, Division of Kidney, Urology, and Hematology, Bethesda, MD: Project Scientist: Ziya Kirkali MD; Project Officer: Christopher Mullins PhD; NIH Personnel: Tamara Bavendam, MD, Robert Star, MD, Jenna Norton

Arbor Research Collaborative for Health, Data Coordinating Center (DK097776 and DK099879): PI: Robert Merion, MD, FACS; Co-Is: Victor Andreev, PhD, DSc, Brenda Gillespie, PhD, Gang Liu, PhD, Abigail Smith, PhD; Project Manager: Melissa Fava, MPA, PMP; Clinical Study Process Manager: Peg Hill-Callahan, BS, LSW; Clinical Monitor: Timothy Buck, BS, CCRP; Research Analysts: Margaret Helmuth, MA, Jon Wiseman, MS; Project Associate: Julieanne Lock, MLitt

Funding/Support

This study is supported by the National Institute of Diabetes & Digestive & Kidney Diseases through cooperative agreements (grants DK097780, DK097772, DK097779, DK099932, DK100011, DK100017, DK097776, DK099879).

Abbreviations Key

AUA-SI	American Urological Association Symptom Index
CASUS	Comprehensive Assessment of Self-reported Urinary Symptoms
CI	Confidence interval
ICIQ-LUTS	International Consultation on Incontinence Modular Questionnaire
IQR	Interquartile range
LURN	Symptoms of Lower Urinary Tract Dysfunction Research Network
LUTS	Lower urinary tract symptoms
NIDDK	National Institute of Diabetes and Digestive and Kidney Diseases
OAB-q	Overactive Bladder Symptom and Health-Related Quality of Life Questionnaire

References

1. Wein AJ, Kavoussi LR. Campbell-Walsh urology ninth edition review. WB Saunders Company; 2007.
2. Coyne KS, Sexton CC, Kopp Z, et al. Assessing patients' descriptions of lower urinary tract symptoms (LUTS) and perspectives on treatment outcomes: results of qualitative research. *Int J Clin Pract.* 2010; 64 :1260. [PubMed: 20579138]
3. Coyne KS, Barsdorf AI, Thompson C, et al. Moving towards a comprehensive assessment of lower urinary tract symptoms (LUTS). *Neurourol Urodyn.* 2012; 31: 448. [PubMed: 22396308]
4. Barry MJ, Fowler FJ Jr, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. The Measurement Committee of the American Urological Association. *J Urol.* 1992; 148: 1549; discussion 1564. [PubMed: 1279218]
5. Abrams P, Avery K, Gardener N, et al. The International Consultation on Incontinence Modular Questionnaire: www.icicq.net. *J Urol.* 2006; 175: 1063; discussion 1066. [PubMed: 16469618]
6. Avery K, Donovan J, Peters TJ, et al. ICIQ: a brief and robust measure for evaluating the symptoms and impact of urinary incontinence. *Neurourol Urodyn.* 2004; 23: 322. [PubMed: 15227649]
7. Coyne K, Revicki D, Hunt T, et al. Psychometric validation of an overactive bladder symptom and health-related quality of life questionnaire: the OAB-q. *Qual Life Res.* 2002; 11 :563. [PubMed: 12206577]
8. Shumaker SA, Wyman JF, Uebersax JS, et al. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. Continence Program in Women (CPW) Research Group. *Qual Life Res.* 1994; 3: 291. [PubMed: 7841963]
9. Norquist JM, Girman C, Fehnel S, et al. Choice of recall period for patient-reported outcome (PRO) measures: criteria for consideration. *Qual Life Res.* 2012; 21: 1013. [PubMed: 21909804]
10. Stav K, Dwyer PL, Rosamilia A. Women overestimate daytime urinary frequency: the importance of the bladder diary. *J Urol.* 2009; 181: 2176. [PubMed: 19296975]
11. Bradley CS, Brown JS, Van Den Eeden SK, et al. Urinary incontinence self-report questions: reproducibility and agreement with bladder diary. *Int Urogynecol J.* 2011; 22: 1565. [PubMed: 21796472]

12. Yap TL, Cromwell DA, Brown C, et al. The relationship between objective frequency–volume chart data and the I-PSS in men with lower urinary tract symptoms. *Eur Urol.* 2007; 52: 811. [PubMed: 17276583]
13. Latini JM, Mueller E, Lux MM, et al. Voiding frequency in a sample of asymptomatic American men. *J Urol.* 2004; 172: 980. [PubMed: 15311017]
14. Wyman JF, Choi SC, Harkins SW, et al. The urinary diary in evaluation of incontinent women: a test-retest analysis. *Obstet Gynecol.* 1988; 71: 812. [PubMed: 3368165]
15. Yang CC, Weinfurt KP, Merion RM, et al. Symptoms of Lower Urinary Tract Dysfunction Research Network. *J Urol.* 2016; 196: 146. [PubMed: 26784646]
16. Weinfurt KP, Griffith JW, Flynn KE, et al. The Comprehensive Assessment of Self-Reported Urinary Symptoms (CASUS): A New Tool for Research on Subtypes of Patients with Lower Urinary Tract Symptoms. *J Urol.* 2019; Epub ahead of print.
17. Redelmeier DA, Kahneman D. Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain.* 1996; 66: 3. [PubMed: 8857625]
18. Fredrickson BL. Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cogn Emot.* 2000; 14: 577.
19. Stone AA, Schwartz JE, Broderick JE, et al. Variability of momentary pain predicts recall of weekly pain: a consequence of the peak (or salience) memory heuristic. *Pers Soc Psychol Bull.* 2005; 31: 1340. [PubMed: 16143666]
20. Broderick JE, Schwartz JE, Vikingstad G, et al. The accuracy of pain and fatigue items across different reporting periods. *Pain.* 2008; 139: 146. [PubMed: 18455312]
21. Cinar A, Hall SA, Link CL, et al. Cluster analysis and lower urinary tract symptoms in men: findings from the Boston Area Community Health Survey. *BJU Int.* 2008; 101: 1247. [PubMed: 18419698]
22. Coyne KS, Sexton CC, Irwin DE, et al. The impact of overactive bladder, incontinence and other lower urinary tract symptoms on quality of life, work productivity, sexuality and emotional well-being in men and women: results from the EPIC study. *BJU Int.* 2008; 101: 1388. [PubMed: 18454794]
23. Coyne KS, Sexton CC, Vats V, et al. National community prevalence of overactive bladder in the United States stratified by sex and age. *Urology.* 2011; 77: 1081. [PubMed: 21256571]
24. Weinfurt KP, Lin L, Dombeck CB, et al. Accuracy of 30-Day Recall for Components of Sexual Function and the Moderating Effects of Gender and Mood. *J Sex Med.* 2014; 11: 678. [PubMed: 23802907]
25. Kenton K, FitzGerald MP, Brubaker L. What is a clinician to do—believe the patient or her urinary diary? *J Urol.* 2006; 176: 633. [PubMed: 16813908]

Figure 1.

Correlations between averaged daily reports and 7- or 30-day recall of lower urinary tract symptoms for women and men. Item text is given in Table 2 and Supplemental Table 1; sample sizes by question and sex are given in Supplemental Tables 3–4. [CI=confidence interval]

Figure 2.

Median bias (IQR) in 7- and 30-day recall of lower urinary tract symptoms compared to averaged daily reports for women and men. Positive bias is systematic over-reporting of 7- or 30-day recall compared to averaged daily recall, negative bias is systematic under-reporting. Item text is given in Table 2 and Supplemental Table 1; sample sizes by question and sex are given in Supplemental Tables 3–4. [IQR=Interquartile range]

Figure 3.
Comparison of average daily reports with most recent and most severe reports over 7- and 30-day recall periods for women. [CI=confidence interval]

Figure 4.
Comparison of average daily reports with most recent and most severe reports over 7- and 30-day recall periods for men. [CI=confidence interval]

Table 1.

Recall Study Design

	Week 1							Week 2							Week 3							Week 4										
	0	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
Group 1	0	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D
	0	B	B					W	D	D	D	D	D	D	D	D						W										
Group 2A	0	B	B	B				W	D	D	D	D	D	D	D							W							W			M
Group 2B	0	D	D	D	D	D	D	D	B	B	B				W							W							W			M

0=Baseline (includes 7-day CASUS), D=24-hour recall, B=3-day daily bladder diary, 3=3-day recall, W=7-day recall, M=30-day recall and final assessment; CASUS=Comprehensive Assessment of Self-reported Urinary Symptoms

Table 2.

Variability of Symptoms over 30 days for Women and Men, by Item (Group 1 participants only)

Symptom	Item	Percent “highly variable” ¹ (N) ²	
		Women	Men
Daytime frequency	During waking hours, how many times did you typically urinate?	49% (98)	33% (99)
	During a typical day, how much time typically passes between urinations?	21% (126)	6% (127)
Nocturia	During a typical night, how many times did you wake up and urinate?	48% (124)	56% (125)
	Did you wake up because you had to urinate? (Y/N)	12% (124)	10% (125)
	When you woke up and urinated, did you leak urine on your way to the bathroom? (Y/N)	5% (124)	2% (125)
Urgency	How often did you feel a sudden need to urinate?	52% (127)	34% (115)
	Once you noticed the need to urinate, how difficult was it to wait more than a few minutes?	55% (124)	39% (113)
Incontinence	Did you completely lose control of your bladder? (Y/N)	7% (59)	6% (16)
	Did you leak urine or wet a pad while laughing, sneezing, or coughing? (Y/N)	7% (82)	0% (15)
	Did you leak urine or wet a pad when doing physical activities, such as exercising or lifting a heavy object? (Y/N)	9% (74)	8% (13)
	Did walking at your usual speed cause you to leak urine or wet a pad? (Y/N)	9% (77)	0% (25)
	Did you leak urine or wet a pad after feeling a sudden need to urinate? (Y/N)	13% (106)	8% (51)
	Did you leak urine or wet a pad without any reason you could identify? (Y/N)	6% (72)	6% (33)
	Did you leak urine or wet a pad without feeling it? (Y/N)	11% (56)	9% (22)
	Did getting up from a chair cause you to leak urine or wet a pad? (Y/N)	5% (62)	0% (20)
Slow/weak stream	How often was your urine flow slow or weak?	43% (101)	38% (115)
Incomplete emptying	How often did you feel that your bladder was not completely empty after urination?	37% (110)	35% (99)
Post-micturition dribble	How often did you dribble urine just after zipping your pants or pulling up your underwear?	32% (75)	32% (97)

¹ Highly variable defined as 40–60% of responses were ‘Yes’ for Yes/No (Y/N) items, or the range of answers spanning 40% of the scale for ordinal items.

² Ns are the number of participants who ever endorsed mild or higher symptoms for that item during the study.