



Original article

## Concordance between patient-reported and physician-reported sexual function after radical prostatectomy

Alan E. Thong, M.D., M.P.H.<sup>a</sup>, Bing Ying Poon, B.A.<sup>b</sup>, Justin K. Lee, M.D., M.P.H.<sup>a</sup>,  
Emily Vertosick, M.P.H.<sup>b</sup>, Daniel D. Sjoberg, M.A.<sup>b</sup>, Andrew J. Vickers, Ph.D.<sup>b,c</sup>,  
Behfar Ehdai, M.D., M.P.H.<sup>a,b,c,\*</sup>

<sup>a</sup> Department of Surgery, Urology Service, Memorial Sloan Kettering Cancer Center, New York, NY

<sup>b</sup> Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY

<sup>c</sup> Weill Cornell Medical College, New York, NY

Received 16 March 2017; received in revised form 15 August 2017; accepted 18 September 2017

### Abstract

**Purpose:** Accurately tracking health-related quality-of-life after radical prostatectomy is critical to counseling patients and improving technique. Physicians consistently overestimate functional recovery. We measured concordance between surgeon-assessed and patient-reported outcomes and evaluated a novel method to provide feedback to surgeons.

**Materials and methods:** Men treated with radical prostatectomy self-completed the International Index of Erectile Function-6 questionnaire at each postoperative visit. Separately, physicians graded sexual function on a 5-point scale. International Index of Erectile Function -6 score <22 and grade  $\geq 3$  defined patient-reported and physician-assessed erectile dysfunction (ED), respectively. Feedback on concordance was given to physicians starting in May 2013 with the implementation of the Amplio feedback system. Chi-square tests were used to assess agreement proportions and linear regression to evaluate changes in agreement after implementation.

**Results:** From 2009 to 2015, 3,053 men completed at least 1 postprostatectomy questionnaire and had a concurrent independent physician-reported outcome. Prior to implementation of feedback in 2013, patients and physicians were consistent as to ED 83% of the time; in 10% of cases, physicians overestimated function; in 7% of cases, physicians, but not patients reported ED. Agreement increased after implementation of feedback but this was not statistically significant, likely owing to a ceiling effect. Supporting this hypothesis, increase in agreement postfeedback was greater during late follow-up ( $\geq 12$  mo), where baseline agreement was lower compared to earlier follow-up.

**Conclusions:** Agreement was higher than expected at baseline; implementation of feedback regarding discrepancies between patient-reported and physician-assessed outcomes did not further improve agreement significantly. Our observed high rate of agreement may be partly attributed to our institutional practice of systematically capturing patient-reported outcomes as part of normal clinical care. © 2017 Elsevier Inc. All rights reserved.

**Keywords:** Prostatectomy; Outcomes; Questionnaires

### 1. Introduction

Treatment of localized prostate cancer with radical prostatectomy (RP) carries risks of functional impairment

in urinary and sexual health-related quality-of-life (HRQOL) [1]. Patient-reported outcomes (PRO) are now critical components of clinical trials used to direct patient-centered care [2]. Providing PRO to individual surgeons helps them to counsel patients, improve surgical technique, and guide follow-up care decisions.

Systematic assessment of PRO is still lacking, however, though the American Urological Association guidelines for localized prostate cancer recommend an assessment of overall health and functional status to guide treatment and follow-up care [3], assessment of patient functional status

*Funding:* Supported by funds from David H. Koch provided through the Prostate Cancer Foundation, Sidney Kimmel Center for Prostate and Urologic Cancers, P50-CA92629 National Cancer Institute SPORE Grant, P30-CA008748 National Institutes of Health cancer center support Grant, and T32 CA82088-15 National Cancer Institute training Grant.

\* Corresponding author: Tel.: +646-422-4406; fax: +646-888-2595.

E-mail address: ehdaieb@mskcc.org (B. Ehdai).

<http://dx.doi.org/10.1016/j.urolonc.2017.09.017>

1078-1439/© 2017 Elsevier Inc. All rights reserved.

has historically been sparse. In the Cancer of the Prostate Strategic Urological Research Endeavor (CaPSURE) and American College of Surgeons National Cancer Data Base cohorts, 22% to 64% of men had no documented assessment of urinary or sexual functional status [4,5].

Even when physicians do assess and document patient functional status, research has demonstrated that physicians' reports are often discordant with patient experience. For instance, in 1 study, surgeon and patient sexual function assessments were concordant for only 55% of patients [5]. These discordant physician assessments consistently underestimate the functional limitations that men experience across multiple domains before and after prostate cancer treatment [6].

Increasing awareness of the importance of PRO assessment in routine clinical care has failed to increase agreement between physician- and patient-assessed functional status. In separate reports from the CaPSURE cohort that were over a decade apart, there was no temporal convergence in the agreement of patient- and physician-reported outcomes [7]. We hypothesized that actively providing surgeons with systematic feedback as to their patients' self-reported sexual function outcomes would improve the concordance between patient- and physician-reported outcomes.

## 2. Materials and methods

### 2.1. Amplio feedback system

Starting January 2009, PRO were systematically collected as part of routine clinical care at Memorial Sloan Kettering Cancer Center (MSKCC). Patients with prostate cancer prospectively completed a validated HRQOL questionnaire assessing erectile, urinary, and bowel function as well as global quality-of-life before RP and at regular follow-up intervals [8]. The survey was administered using an interactive secure online form completed before clinical appointments via e-mail or immediately before the clinical appointment via tablet computer. A total of 6 questions from the International Index of Erectile Function-6 comprised the patient self-assessment of potency [9]. During the clinical visit also, physicians independently graded sexual function on a 5-point scale based on history and physical examination.

This routine digital collection of PRO then served as a critical component of the Amplio feedback system [10]. Beginning in May 2013, the Amplio system provided surgeons at MSKCC with biannual individualized and confidential feedback on PRO. Amplio is an interactive information technology platform developed to provide physicians across various surgical disciplines with feedback on their risk-adjusted outcomes and anonymized peer comparisons. In the case of RP, one of the metrics the Amplio system reports back to surgeons is the concordance of potency rates between patient-reported and surgeon-assessed

outcomes. The overall results are presented at surgical staff meetings, with individual surgeons encouraged to log-on and view their personal results. Log-ons are monitored, with a designated surgeon serving the role of liaison, to encourage use of the feedback tool.

### 2.2. Study cohort

After obtaining Institutional Review Board approval, we used our institutional database to identify 4,330 men who had undergone RP at MSKCC from January 2009 to April 2015. In order to compare sexual function agreement between surgeons and patients over follow-up, we omitted 1,277 men who did not have at least 1 HRQOL survey completed at the same follow-up time point by both the patient and the surgeon. Of the resulting cohort of 3,053 men, 2,359 underwent RP before Amplio concordance feedback (before May 2013) and 694 underwent RP after implementation of Amplio concordance feedback.

Patients were considered sexually potent if they self-reported a score of at least 22 points on the International Index of Erectile Function-6 (range: 1–30). Surgeons were considered to have rated their patients as potent if they reported a score of 2 or less on the 5-point erectile dysfunction (ED) survey (Supplemental material 1). If both patient and surgeon rated the patient as being potent or if they both rated the patient as having ED, this was considered agreement. Cases where the surgeon reported potency and the patient-reported ED were considered to be cases of overestimation on the part of the surgeon. Cases where the surgeon reported ED and the patient-reported potency were considered underestimation by the surgeon.

### 2.3. Statistical analysis

The chi-square test was used to assess the overall agreement between the surgeon and the patient on sexual function outcomes. To describe the level of agreement for all patient-surgeon interactions throughout follow-up, we calculated the proportion of times surgeons and patients agreed and the proportion of times surgeons overestimated sexual function.

To evaluate any changes in agreement and overestimation that resulted once the Amplio system feedback was part of the process, we used a general estimating equation with logit link to compare those surveys taken before feedback was available on May 15, 2013 and those taken after November 15, 2013. A general estimating equation was used because each patient-surgeon interaction throughout follow-up was included in the analyses, so we had to account for correlation within each patient-surgeon pair and across follow-up times. Surveys taken from May 15, 2013 to November 15, 2013 were not included in the analyses to allow for an adjustment period.

Linear regression was also used to determine the change in mean agreement and overestimation over follow-up for

patients who were ever evaluated after November 15, 2013 compared to patients who were only evaluated before May 15, 2013. The proportion of agreement and the proportion of overestimation before and after the cutoff was calculated to describe the change in agreement or overestimation before and after the implementation of concordance feedback with the Amplio system.

It is plausible that the pattern of agreement and overestimation may change over follow-up. For instance, there may be good concordance at a few months after surgery, when many patients have poor function and dysfunction may be seen less as a reflection on surgeon skill than later, when results might be seen to be more as the true result of surgery. Accordingly, an interaction analysis was used to evaluate whether the implementation of concordance feedback with the Amplio system affected agreement or overestimation at early (<12 mo) follow-up differently compared to late ( $\geq 12$  mo) follow-up using a general estimating equation.

### 3. Results

In the period from January 2009 to April 2015, a total of 3,053 men completed at least 1 post-RP HRQOL self-assessment and had a concurrent, independently reported sexual function outcome from the physician. Patient characteristics are described in Table 1. Patients who received surgery after implementation of Amplio concordance feedback more often received robot-assisted laparoscopic

surgery (71% vs. 45%;  $P < 0.0001$ ) and fewer patients harbored low-risk Gleason grade 3 + 3 prostate cancer (21% vs. 32%;  $P < 0.0001$ ). Both of these factors result from temporal trends in treatment at MSKCC. Differences among the excluded HRQOL survey nonresponders were separately analyzed. Many of the 1,277 excluded men underwent surgery in 2014 or later (34% vs. 13% for the 3,053 evaluable patients;  $P < 0.0001$ ) and have not had long enough clinical follow-up for either patients or surgeons to complete HRQOL questionnaires. Other temporal trends were demonstrated by significant differences in biopsy Gleason grade, and proportion of robotic-assisted laparoscopic RP vs. open RP in these variables, but none of these variables were expected to affect the agreement between patient and surgeon on HRQOL outcomes.

The overall degree of agreement between surgeons and patients on sexual function outcomes was unexpectedly high across all follow-up time-points (83%). However, when surgeons and patients disagreed on the patients' sexual function outcome, surgeons more often overestimated function (10%) rather than underestimated function (7%). Overall distributions of patient-reported sexual function scores across the physician-assessed sexual function scores are shown in Fig. 1.

Change in agreement and overestimation is shown with the locally weighted scatterplot smoothing curve in Fig. 2. Overall, agreement across all follow-up time points was 82% with systematic PRO collection only (before May 2013) and 84% after the addition of Amplio concordance feedback, a difference of 1.6% (95% CI: -1.4% to 3.2%).

Table 1  
Patient characteristics

	N = 3,053		P value
	Systematic PRO collection (n = 2,359; 77%)	Systematic PRO collection + Amplio concordance feedback (n = 694; 23%)	
Median age at surgery (IQR)	61 (56–66)	61 (56–66)	0.6
Median preoperative PSA, ng/ml (IQR) (N = 3034)	5 (4–7)	6 (4–8)	0.002
No. biopsy Gleason grade (%)	(n = 2,351)	(n = 688)	<0.0001
≤6	763 (32%)	143 (21%)	
7	1,267 (54%)	430 (63%)	
≥8	321 (14%)	115 (17%)	
No. clinical stage (%)	(n = 2,325)	(n = 628)	0.5
T1	1,514 (65%)	410 (65%)	
T2	715 (31%)	185 (29%)	
T3	94 (4.0%)	33 (5.3%)	
T4	2 (<0.1%)	0 (0%)	
No. surgery type (%)			<0.0001
Open	835 (35%)	137 (20%)	
Laparoscopic (nonrobotic)	462 (20%)	63 (9.1%)	
Robotic-assisted laparoscopic	1,062 (45%)	494 (71%)	

P values for age and PSA were calculated using the Wilcoxon rank sum test; the other P values were calculated using Fisher exact test. IQR = interquartile range.

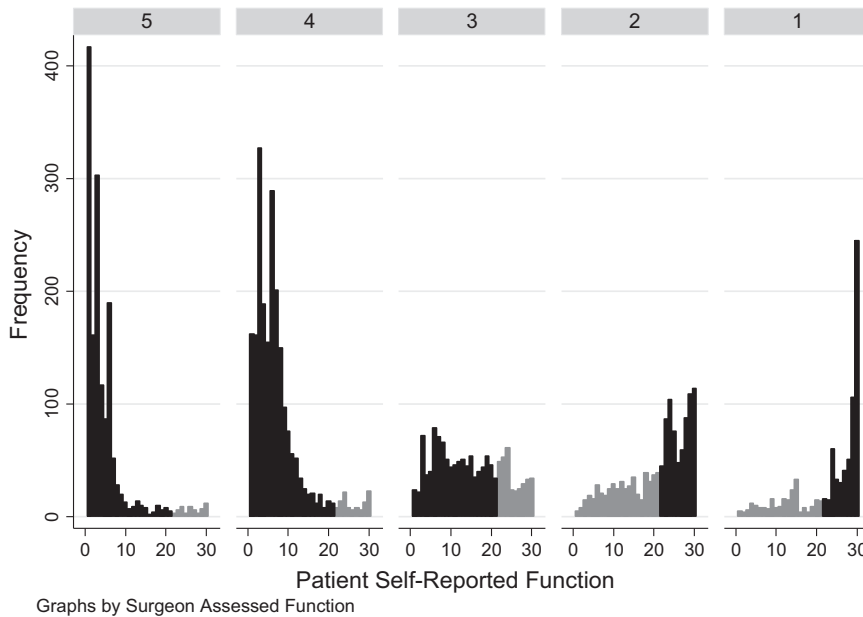


Fig. 1. Histogram of patient scores by surgeon scores. Patient scores range from 1 to 30 points where  $\geq 22$  points indicates potency. Surgeon scores range from 1 to 5 points where  $\leq 2$  points indicates potency. Black areas show agreement between patients and surgeon, and gray areas show disagreement.

This was not statistically significant using a general estimating equation ( $P = 0.5$ ) or linear regression on mean agreement ( $P = 0.3$ ). Changes in overestimation were also not found to be statistically significant, changing from only 10.7%–10.0%, a difference of 0.7% (95% CI: -2.0% to 1.6%), (11%; 95% CI: 10%–12% to 10%; 95% CI: 8%–12%) with a general estimating equation ( $P = 0.8$ ). As shown in Fig. 3, implementation of Amplio concordance feedback and phase of follow-up was not found to have significant interaction for overestimation ( $P = 0.2$ ), though we did see evidence that the influence of feedback on agreement was higher on longer-term follow-up ( $P = 0.071$ ).

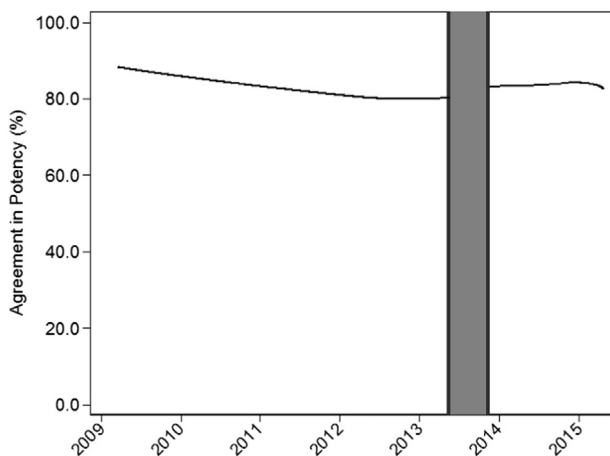


Fig. 2. Lowess curve for agreement in sexual function outcomes. The gray window marks the 6-month adjustment period (May 15–November 15, 2013) not included in the comparison of agreement before and after feedback. Lowess = locally weighted scatterplot smoothing.

#### 4. Discussion

Despite the importance of accurately assessing functional status in prostate cancer treatment decision-making and survivorship, there have been few studies examining the issue of patient-physician agreement on outcomes and possible overestimation of functional recovery by surgeons. Prior work from Litwin et al. and Sonn et al. using the CaPSURE cohort demonstrated significant and persistent discordance between physician- and patient-reported functional assessments using the University of California at Los Angeles Prostate Cancer Index and 4-point physician scales

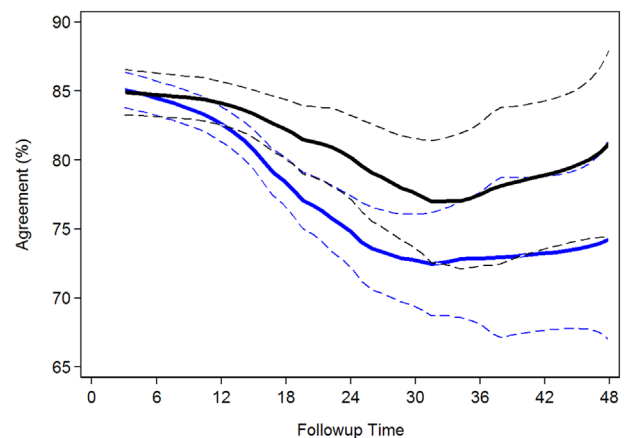


Fig. 3. Lowess curve for agreement in sexual function outcomes by follow-up period. Solid blue line shows agreement between surgeons and patients who were evaluated with systematic PRO collection only; solid black line shows agreement between surgeons and patients who were ever evaluated after Amplio concordance feedback. Dashed lines represent CIs around the Lowess curves. Lowess = locally weighted scatterplot smoothing. (Color version of figure is available online.)

for patient symptoms including impotence [6,7]. Importantly, a process to provide real-time or aggregate feedback to physicians regarding PRO data were not implemented in any of these studies.

We found high patient-physician concordance rates both before and after the implementation of feedback. Given the high-concordance rates before implementation, we failed to find a significant increase in concordance after the implementation of the Amplio feedback system. Overall, agreement between physician-reported outcomes and PRO was higher in our cohort than previously reported in prostate cancer patients. While we observed patient and physician agreement averaging 83% at follow-up, concordance in the CaPSURE cohort was as low as 55% [6]. This difference may result from our cohort being based on the RP-only experience of a single academic institution vs. CaPSURE, a cohort encompassing academic and community practices treating prostate cancer with a variety of modalities. But we think a more likely explanation is that we collect PRO as part of routine clinical care with point-of-care integration of the PRO questionnaire results into surgeons' clinic visit workflow.

The correlation between physician and patient assessments has also been called into question in the broader oncology and surgical literature [11–16]. Studies examining patient-physician agreement on general metrics of functional status such as the Karnofsky performance status scale [15] and Eastern Cooperative Oncology Group performance status scale [16] have shown that nearly half of patients disagree with their physicians' functional assessment, with physicians tending to overestimate function.

Even within the controlled clinical care and documentation of randomized trials, physicians underreport subjective toxicities when compared with PRO. Di Maio et al. [11] reported the rate of agreement in 3 randomized chemotherapy trials that collected data on toxicities from both patient-reported European Organization for Research and Treatment of Cancer quality-of-life questionnaires and physician-documented toxicities using the National Cancer Institute Common Toxicity Criteria or Common Terminology Criteria for Adverse Events. Physician-reported toxicities were consistently lower than those reported by patients, with rates of physician underreporting approaching 50%. Similarly, Gravis et al. [12] reported a physician discordance rate of 81% for sexual dysfunction when compared to patient-reported sexual dysfunction in a phase III trial of androgen deprivation therapy with or without docetaxel for metastatic noncastrate resistant prostate cancer.

Only when using nearly identical physician questionnaires and patient-derived versions of these questionnaires have physician-reported outcomes more closely paralleled those reported by patients. Smith et al. [13] found that a high level of agreement was attained using an identical shoulder replacement HRQOL questionnaire for patients and physicians, and Basch et al. [14] reported on the high observed agreement between physician and patient versions of the Common Terminology Criteria for Adverse Events.

Despite our using differing survey instruments and scales between patients and physicians, we still observed a high overall agreement on sexual function outcomes after RP. With only the systematic collection of PRO in this cohort, before the implementation of Amplio concordance feedback, 82% of patients agreed with the physician assessment. After the implementation of feedback, a ceiling effect was observed without any further significant improvement in concordance.

It is possible that in this cohort, the baseline systematic collection of PRO and real-time integration of PRO into the clinical visit workflow had already reduced or eliminated potential patient-physician communication barriers that may be responsible for the underreporting of ED observed in other studies. In this setting where PRO collection, documentation, and workflow integration are already part of routine clinical care for the physician and patient, additional anonymized feedback on concordance may not make an already high level of agreement much higher. Nevertheless, linking similar patient and physician HRQOL domains across disparate questionnaires and scales required setting discrete cutoffs points to indicate ED.

While directly surveying patients is the most valid and reliable means of determining HRQOL, obtaining both patient and physician assessments of HRQOL may help facilitate shared decision-making. Underestimating sexual function impairment after prostate cancer treatment may affect counseling men with low-risk prostate cancer regarding active surveillance and undermine shared decision-making. Greater concordance between physicians and patients on functional outcomes has been positively associated with outcomes. Chamie et al. [5] previously suggested that pretreatment assessment of function is an essential quality of care indicator. The majority of men in the CaPSURE cohort of men treated for localized prostate cancer did not have pretreatment assessment of function and these men were at higher risk of sexual and bowel dysfunction after treatment.

This positive effect on outcomes may be causally owing to a number of factors, including increased physician questioning, documentation, and discussion of patient symptoms when PRO collection is systematically incorporated into routine clinical practice. In a medical oncology setting, routine assessment of cancer patients' HRQOL had a positive effect on physician-patient communication and improved HRQOL and emotional functioning [17]. Improved patient and physician engagement has also been suggested to improve patient treatment decision satisfaction [18] and decreased resource usage [19].

Usage of PRO feedback to improve the actual quality of outcomes data collected is a novel quality improvement application. In a theoretical review of quality improvement driven by PRO feedback, Greenhalgh et al. [20] highlight the lack of research in the utility of aggregated PRO feedback to providers. Provider feedback of PRO to improve surgical outcomes has largely proved unsuccessful either owing to a ceiling effect, lack of effective outcomes communication, or lack of a perceived direction on how to

improve outcomes [21,22]. We focused on a more proximate effect, improving concordance between patient and provider reported outcomes, and also found a ceiling effect with an already high degree of interrater agreement.

Newer multi-institutional prostate cancer registries including AQUA (AUA Quality Registry), CEASAR (Comparative Effectiveness Analysis of Surgery and Radiation), and MUSIC (Michigan Urological Surgery Improvement Collaborative) have the potential to provide the urological community with increased information about general HRQOL after prostate cancer treatments, though the application of this aggregate data towards quality improvement is yet to be determined.

The Amplio system is being expanded through MSKCC surgery. At the time of writing, Amplio modules are available for 12 procedures including cystectomy, nephrectomy, esophagectomy, Whipple, and gastrectomy, among others. An additional 4 procedures will be added to Amplio by the end of 2017. Some of these, such as breast reconstruction, include PROs. The continued implementation of electronic medical records across many different types of practice settings nationwide should also continue to facilitate wider application of such routine PRO collection and feedback tools. It remains to be seen in future research if going beyond the collection of PRO to providing active, risk-adjusted, anonymized surgeon PRO feedback will result in wider quality improvement in physician-patient agreement.

## 5. Conclusions

In a large institutional cohort of men undergoing RP, standardized and systematic collection of patient- and physician-reported outcomes demonstrated a high degree of interrater agreement. Actively providing feedback to surgeons on their degree of agreement with patient self-assessments did not further improve the degree of agreement. There is ample reason to believe that the systematic collection of PRO as part of routine clinical care contributes to improved delivery of patient-centered prostate cancer care.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.urolonc.2017.09.017>.

## References

- [1] Singh J, Trabulsi EJ, Gomella LG. The quality-of-life impact of prostate cancer treatments. *Curr Urol Rep* 2010;11:139–46.
- [2] Efficace F, Feuerstein M, Fayers P, et al. Patient-reported outcomes in randomised controlled trials of prostate cancer: methodological quality and impact on clinical decision making. *Eur Urol* 2014;66:416–27.
- [3] Thompson I, Thrasher JB, Aus G, et al. Guideline for the management of clinically localized prostate cancer: 2007 update. *J Urol* 2007;177:2106–31.
- [4] Spencer BA, Miller DC, Litwin MS, et al. Variations in quality of care for men with early-stage prostate cancer. *J Clin Oncol* 2008;26:3735–42.
- [5] Chamie K, Sadetsky N, Litwin MS. Physician assessment of pretreatment functional status: a process-outcomes link. *J Urol* 2011;185:1229–33.
- [6] Litwin MS, Lubeck DP, Henning JM, et al. Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: results of the capsure database. *J Urol* 1998;159:1988–92.
- [7] Sonn GA, Sadetsky N, Presti JC, et al. Differing perceptions of quality of life in patients with prostate cancer and their doctors. *J Urol* 2009;182:2296–302.
- [8] Vickers AJ, Savage CJ, Shouery M, et al. Validation study of a web-based assessment of functional recovery after radical prostatectomy. *Health Qual Life Outcomes* 2010;8:82.
- [9] Rosen RC, Cappelleri JC, Gendrano N. The International Index of Erectile Function (IIEF): a state-of-the-science review. *Int J Impot Res* 2002;14:226–44.
- [10] Vickers AJ, Sjoberg D, Basch E, et al. How do you know if you are any good? A surgeon performance feedback system for the outcomes of radical prostatectomy. *Eur Urol* 2012;61:284–9.
- [11] Di Maio M, Gallo C, Leighl NB, et al. Symptomatic toxicities experienced during anticancer treatment: agreement between patient and physician reporting in three randomized trials. *J Clin Oncol* 2015;33:910–5.
- [12] Gravis G, Marino P, Joly F, et al. Patients' self-assessment versus investigators' evaluation in a phase III trial in non-castrate metastatic prostate cancer (GETUG-AFU 15). *Eur J Cancer* 2014;50:953–62.
- [13] Smith AM, Barnes SA, Sperling JW, et al. Patient and physician-assessed shoulder function after arthroplasty. *J Bone Joint Surg Am* 2006;88:508–13.
- [14] Basch E, Iasonos A, McDonough T, et al. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol* 2006;7:903–9.
- [15] Slevin ML, Plant H, Lynch D, et al. Who should measure quality of life, the doctor or the patient? *Br J Cancer* 1988;57:109–12.
- [16] Lee CF, Ng R, Luo N, et al. Patient-reported outcomes are associated with patient-oncologist agreement of performance status in a multi-ethnic Asian population. *Support Care Cancer* 2014;22:3201–8.
- [17] Velikova G, Booth L, Smith AB, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 2004;22:714–24.
- [18] Martinez LS, Schwartz JS, Freres D, et al. Patient-clinician information engagement increases treatment decision satisfaction among cancer patients through feeling of being informed. *Patient Educ Couns* 2009;77:384–90.
- [19] Stewart M, Brown JB, Donner A, et al. The impact of patient-centered care on outcomes. *J Fam Pract* 2000;49:796–804.
- [20] Greenhalgh J, Dalkin S, Gooding K, et al. Functionality and feedback: a realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. Southampton (UK): NIHR Journals Library 2017. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK409450/>. Accessed August 15, 2017.
- [21] Varagunam M, Hutchings A, Neuburger J, et al. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy* 2014;19:77–84.
- [22] Boyce MB, Browne JP. The effectiveness of providing peer benchmarked feedback to hip replacement surgeons based on patient-reported outcome measures—results from the PROFILE (Patient-Reported Outcomes: Feedback Interpretation and Learning Experiment) trial: a cluster randomised controlled study. *BMJ Open* 2015;5:e008325.