



Published in final edited form as:

Eur Urol. 2012 February ; 61(2): 284–289. doi:10.1016/j.eururo.2011.10.039.

How do you know if you are any good? A surgeon performance feedback system for the outcomes of radical prostatectomy

Andrew J. Vickers, Daniel Sjoberg, Ethan Basch, Frank Sculli, Marwan Shouery, Vincent Laudone, Karim Touijer, James Eastham, and Peter T. Scardino

Memorial Sloan-Kettering Cancer Center

Introduction

Variation in outcomes

Surgery remains a mainstay of initial treatment for prostate cancer, with an estimated 85,000 operations per year in the US. Radical prostatectomy is associated with important risks of erectile dysfunction, urinary incontinence and, naturally, cancer recurrence. Given the possible consequences, it would be reassuring were it known that urologic surgeons offer uniformly high-quality care. Unfortunately, the data suggest that this is far from the case. There is copious evidence that surgeons with greater case volume or total lifetime experience have better outcomes. For example, low volume surgeons have complication rates 6 to 8% greater than their higher volume counterparts[1]; in studies on the learning curve, the risk of recurrence is about 7% higher for a typical patient treated by an inexperienced surgeon than if treated by a more experienced surgeon[2, 3] There are also data that differences in outcome go over and above characteristics such as volume or experience, with large variations between surgeons even within volume categories[1, 4], with one study reporting a five-fold variation in potency rates between surgeons at a single institution[5].

The extent of differences in outcome between different surgeons, as demonstrated in these studies, suggests that many patients are receiving suboptimal care. This raises the issue of how to improve the results of underperforming surgeons.

Feedback for quality improvement

It is an axiom of education theory that performance feedback is critical to learning: consider if a schoolchild was never told whether his math answers were correct, or a basketball player could not tell whether practice free throws had been made. Yet there is a strong case to be made that usable performance feedback is largely absent for surgeons. Or to put it another way: how does a surgeon know whether he or she is any good?

A surgeon must first *obtain data on patient outcomes*. This might be relatively straightforward for an outcome such as perioperative mortality, but for many types of surgery, patient reported outcomes are critical. In the case of radical prostatectomy, erectile and urinary function are key endpoints; it is hugely time consuming and expensive to track patients over time, administer questionnaires, chase non-compliant patients and then enter results into a database.

Even were a surgeon to go to the time and trouble of collecting patient reported outcomes, there would be a need to *benchmark results*. As a simple example, if 45% of a surgeon's patients had recovered potency at one year, would this constitute a good result or a poor one? In theory a surgeon could compare results to those published in the literature, but it is unlikely that published results constitute a representative sample of surgeon outcomes. Moreover, published papers involve a wide variety of different methods of assessing erectile function and of defining potency. One review paper reported definitions of potency that included >17 on IIEF6; 26 on IIEF6; >21 on SHIM; grade 1 – 2 on a 5 point scale; grade 1 – 3 on a 5 point scale; “ability to have sexual intercourse” and several others[6].

Yet even benchmarked, standardized patient reported outcomes would be inadequate in the absence of *risk adjustment*. If 30% of surgeon A's patients were potent at one year, but in comparison, surgeon B had a potency rate of 45%, we might be tempted to recommend remedial action for surgeon A. Yet if surgeon B treated predominately young, healthy men with low risk cancer and surgeon A focused on older men with high risk disease, this advice might be entirely misplaced.

In sum, feedback to surgeons as to their outcomes is only possible if a sophisticated system of data gathering, risk adjustment and benchmarking has been implemented. Here we describe the development of such a system at Memorial Sloan-Kettering Cancer Center (MSKCC).

Methods

The design of the MSKCC surgeon performance feedback system was informed by both theoretical considerations from the literature and practical experience of a pilot at MSKCC.

Theoretical requirements for a performance feedback system

There is a developing literature on performance feedback that suggests a number of factors critical to success[7–9]. First, feedback must be confidential. This is largely because public reporting creates perverse incentives, such as doctors switching to low risk cases or gaming risk data[10]. Second, feedback must be credible, that is, the “data must be perceived by physicians as valid to motivate change”[8]. In one well-known feedback project that is regarded as being unsuccessful, almost all participants (85%) stated that the method of statistically adjusting results was insufficient, such that surgeons who operated on high risk patients would appear to have worse than average results[11]. Third, feedback must focus on outcomes, such as complication rates or erectile function, rather than processes, such as use of preoperative β -blockers or postoperative PDE5 inhibitors. Fourth, feedback must be integrated into routine practice. A key predictor of the success of feedback is its persistence[7]. If established as a stand-alone “research” project, participants will perceive feedback as time-limited and have little incentive for continuous and cumulative quality improvement.

Pilot study at MSKCC

An early version of the surgeon performance feedback system was piloted in July 2009. All surgeons at MSKCC who had data on at least 30 procedures were included. Each surgeon was sent details of their adjusted rates for recurrence, erectile dysfunction and continence, along with their rank (e.g. 4th of 13 surgeons), and the mean and best rates (e.g. “Adjusted average for all surgeons: 79%; best rate for a surgeon with at least 100 patients: 91%”). Data on functional outcomes were taken from surgeon notes.

Each surgeon was then interviewed to determine his or her opinions on the project. This provided information on how surgeons viewed the credibility of the data. In particular, the

functional outcomes used in the pilot were surgeon assessed, and this lead to concerns about biased assessment. For example, it may have been that the surgeons with good results were merely those who were most overoptimistic about their patients' recovery.

The pilot also allowed surgeons to provide ideas on how the feedback should be presented. These included provision of feedback on perioperative outcomes; giving a distribution of results (“if I am the 8th surgeon in term of continence [it makes a difference if] the first 7 first have a continence rate [slightly or much higher] than mine”); showing results for specific periods of time such as patients treated in the last year (“[my results are an average from] many years and ... to modify them is almost impossible as a mid-term goal”).

Obtaining patient reported outcomes

In the light of the surgeons' comments on our pilot, a critical element in ensuring credibility was to obtain functional outcomes direct from patients. Our approach has been to collect patient-reported outcomes electronically, so that data can be used both in clinical practice and for the performance feedback system with a minimum burden of data collection. The interface is based on a software environment known as STAR (“Symptom Tracking and Reporting”) developed and validated at MSKCC originally for monitoring chemotherapy toxicities[12–14]. We developed a web-based questionnaire for patient-reported outcomes after radical prostatectomy that includes the six-question version of the International Index of Erectile Function (IIEF) as well as questions about urinary function, bowel function and overall health-related quality of life. This electronic questionnaire has been shown to have good psychometric properties in a validation study[15].

The current web-interface can be implemented in two distinct ways. Many radical prostatectomy patients have email addresses and thus access a web-interface via a click-through from a reminder email. Patients without email can access the web-interface via iPads available at the clinic. The interface is in principle no more complex than a case machine and we have found no significant problems in terms of computer literacy.

Data from the STAR patient-reported outcomes is transferred to the Caisis database, an open source clinical information system developed over the past decade that is currently used to help manage patient care at MSKCC and several other large academic centers. Caisis stores data on patient characteristics (such as age, family history, and co-morbidities), tumor characteristics (such as pathologic stage and Gleason score), operative characteristics (such as blood loss and nerve sparing), and outcomes (such as postoperative PSAs and complications).

A surgeon performance feedback system

A tab on Caisis allows surgeons to log on to explore their outcomes, case mix adjusted and in comparison to their peers. Figure 1 shows a screen shot in which oncologic outcomes are graphed against functional recovery. The ideal result would be in the top right, where the surgeon would high cure rates with most patients both potent and continent. The surgeon is only able to access his or her own personal results (indicated by the red triangle); he or she can see the results of other surgeons, but not who those surgeons are. The statistical methodology is to build a predictive model using covariates specific to the endpoint and include each surgeon as a fixed effect. Covariates for oncologic outcomes include stage, grade and PSA; functional outcomes also include baseline function, age and co-morbidity. Adjusted rates are then calculated by fixing covariates to the mean. For some endpoints, such as surgeon volume, no statistical adjustment is applied. Statistical code for calculation of surgeon-specific rates is written in the R programming language and is integrated in

Caisis. As such, the process is completely automated, and does not involve statistical or data management staff to download or analyze data.

Two critical features of the system are that it is interactive and multimodal. A wide variety of outcomes data are available in addition to oncologic control and functional recovery. These include perioperative data (operative time, length of hospital stay, estimate blood loss); surgical volume; reporting rates (e.g. proportion of patients providing functional data at 6 months); patient selection given as life expectancy (calculated from age and comorbidity) vs recurrence risk; rates of positive surgical margins. Outcomes are presented as scatter plots or histograms (see figures 2 – 3) as appropriate. A full list of endpoints and options is given in table 1. The importance of multimodality is that it prevents surgeons attempting to optimize some endpoints at the expense of others. For example, if the only feedback given was on surgical margins, surgeons may try to lower margin rates at the expense of functional preservation; if case mix was not included as an endpoint, a feedback system may lead surgeons to select lower risk patients[10].

The interactive nature of the system can be seen on the right hand side of figure 1, where surgeons can select different options for viewing results. These include the surgical modality (open, laparoscopic, robotically-assisted), time-point (recurrence rates at between 1 and 5 years, functional outcomes at between 3 months and two years), risk group (all patients, low risk patients only, or intermediate and high risk patients only), and time period (all patients, or those treated in the last 1, 2 or 5 years). For example, a surgeon might have made a recent change in technique and so would want to look only at results for patients treated in the last year; another might want to focus on results concerning early recovery of function; a third may feel that erectile dysfunction in some higher risk patients is an inevitable consequence of appropriate surgery and so would be particularly interested in functional outcomes in low risk patients. Figure 4 shows an example of a customized report, restricted to patients at intermediate and high risk, and showing potency and continence outcomes at 6 months. The interactive nature of the interface also serves to reinforce that the perception that the performance feedback system is an opportunity for surgeons to learn and try to improve their results, rather than a top down “report” in which the surgeon is judged by an outsider[10]. This helps improve surgeons’ credibility in the system.

Looking at time periods can also help evaluate changes at the institutional level. Figure 2 shows the overall case selection results at MSKCC. It can be seen that some surgeons are treating patients with relatively short life expectancy, with a mean of 12 or 14 years. Figure 3 shows the life expectancy results for patients treated in the last 2 years, when much more consistent institutional policies were in place, particularly with respect to the use of active surveillance for older patients with low risk disease.

Surgeon reactions

Reactions to the feedback system from surgeons at MSKCC have generally been very positive. In particular, no surgeon has raised questions about the data, on the grounds that all see it being collected from patients on a day-to-day basis. Moreover, the presentation of the system as a tool, embedded within the electronic health record, rather than as a report from a higher authority, has helped surgeons feel ownership over the process. Indeed, most surgeons made suggestions as to the content or form of the feedback system, which have either been incorporated into the existing software or are planned for the “2.0” version.

The feedback system has also prompted educational activities. The service chief asked to be unblinded to the identity of a particularly high performing surgeon and has asked that surgeon to prepare presentations about surgical technique to the faculty.

Conclusions

Surgeons typically do not know whether their results are good or bad because, to do so, would require that they obtain high quality outcome data from their patients, then apply a statistical risk adjustment and make a comparison with their peers. This is only feasible using a complex health informatics system.

We have developed a surgeon performance feedback system. This is based on prior literature on surgeon feedback, attempting to incorporate aspects that have been successful and avoid those associated with failure. The feedback system provides data on outcomes, rather than processes, and the data have high credibility because the systems for obtaining data and adjusting results for case mix have been developed in close collaboration with clinicians. Feedback is benchmarked, by comparing a surgeon's outcomes with all other surgeons in the database, and is multimodal, incorporating both functional and oncologic outcomes, and data on both outcomes and patient selection; this helps to avoid perverse incentives. Outcomes reporting is private – a surgeon's results are only accessible by that surgeon – preventing the perception that feedback is punitive, and interactive, allowing surgeons to explore their own results and customize data presentation. The system is fully integrated into routine clinical practice, being available on the very same database used to access patient data, ensuring that feedback is ongoing, and that the system can be “scaled” to other institutions. Indeed, the next stage of the project is to develop software to allow different institutional databases to talk with one another so as to allow multiple institutions to participate.

All that said, we cannot be sure that use of our system will actually improve outcomes. Accordingly, we intend to monitor outcomes before and after implementation of the surgeon feedback system at each institution. We hope to find that, once surgeons are aware of their outcomes, they will find ways to improve them.

Acknowledgments

Supported in part by funds from the Goldstein Foundation, David H. Koch provided through the Prostate Cancer Foundation, the Sidney Kimmel Center for Prostate and Urologic Cancers and a P50-CA92629 SP0RE grant from the National Cancer Institute to Dr. P. T. Scardino.

References

1. Begg CB, Riedel ER, Bach PB, Kattan MW, Schrag D, Warren JL, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med.* 2002; 346:1138–44. [PubMed: 11948274]
2. Vickers AJ, Bianco FJ, Serio AM, Eastham JA, Schrag D, Klein EA, et al. The surgical learning curve for prostate cancer control after radical prostatectomy. *J Natl Cancer Inst.* 2007; 99:1171–7. [PubMed: 17652279]
3. Vickers AJ, Savage CJ, Hruza M, Tuerk I, Koenig P, Martinez-Pineiro L, et al. The surgical learning curve for laparoscopic radical prostatectomy: a retrospective cohort study. *Lancet Oncol.* 2009; 10:475–80. [PubMed: 19342300]
4. Bianco FJ Jr, Vickers AJ, Cronin AM, Klein EA, Eastham JA, Pontes JE, et al. Variations among experienced surgeons in cancer control after open radical prostatectomy. *J Urol.* 2010; 183:977–82. [PubMed: 20083278]
5. Vickers A, Savage C, Bianco F, Mulhall J, Sandhu J, Guillonau B, et al. Cancer control and functional outcomes after radical prostatectomy as markers of surgical quality: analysis of heterogeneity between surgeons at a single cancer center. *Eur Urol.* 2011; 59:317–22. [PubMed: 21095055]
6. Mulhall JP. Defining and reporting erectile function outcomes after radical prostatectomy: challenges and misconceptions. *J Urol.* 2009; 181:462–71. [PubMed: 19084865]

7. Peterson ED. Optimizing the science of quality improvement. *JAMA*. 2005; 294:369–71. [PubMed: 16030283]
8. Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. Data feedback efforts in quality improvement: lessons learned from US hospitals. *Qual Saf Health Care*. 2004; 13:26–31. [PubMed: 14757796]
9. Bradley EH, Holmboe ES, Mattera JA, Roumanis SA, Radford MJ, Krumholz HM. A qualitative study of increasing beta-blocker use after myocardial infarction: Why do some hospitals succeed? *JAMA*. 2001; 285:2604–11. [PubMed: 11368734]
10. Vickers A, Eastham J. What has to happen before we report radical prostatectomy outcomes of individual surgeons to the public? *Urologic oncology*. 2011; 29:118–23. [PubMed: 20884248]
11. Narins CR, Dozier AM, Ling FS, Zareba W. The influence of public reporting of outcome data on medical decision making by physicians. *Arch Intern Med*. 2005; 165:83–7. [PubMed: 15642879]
12. Basch E, Artz D, Dulko D, Scher K, Sabbatini P, Hensley M, et al. Patient online self-reporting of toxicity symptoms during chemotherapy. *J Clin Oncol*. 2005; 23:3552–61. [PubMed: 15908666]
13. Basch E, Iasonos A, Barz A, Culkin A, Kris MG, Artz D, et al. Long-term toxicity monitoring via electronic patient-reported outcomes in patients receiving chemotherapy. *J Clin Oncol*. 2007; 25:5374–80. [PubMed: 18048818]
14. Basch E, Artz D, Iasonos A, Speakman J, Shannon K, Lin K, et al. Evaluation of an online platform for cancer patient self-reporting of chemotherapy toxicities. *J Am Med Inform Assoc*. 2007; 14:264–8. [PubMed: 17329732]
15. Vickers AJ, Savage CJ, Shouery M, Eastham JA, Scardino PT, Basch EM. Validation study of a web-based assessment of functional recovery after radical prostatectomy. *Health Qual Life Outcomes*. 2010; 8:82. [PubMed: 20687938]

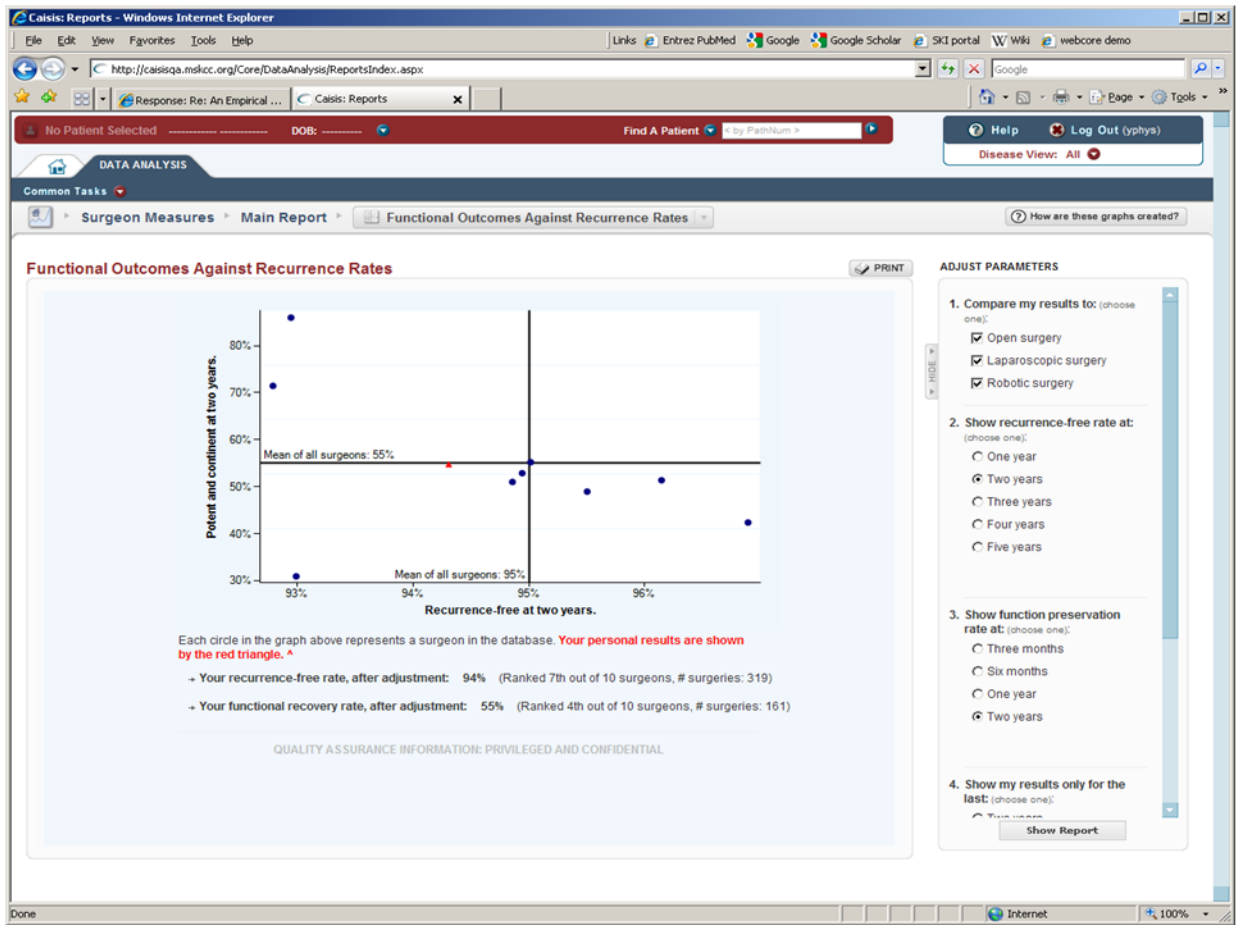


Figure 1. Screen shot from the MSKCC surgeon performance feedback system: main report
 The proportion of a surgeon’s patients who are recurrence free at two years on the x axis is graphed against the proportion of patients both potent and continent at two years (y axis). Both estimates are adjusted for case mix.

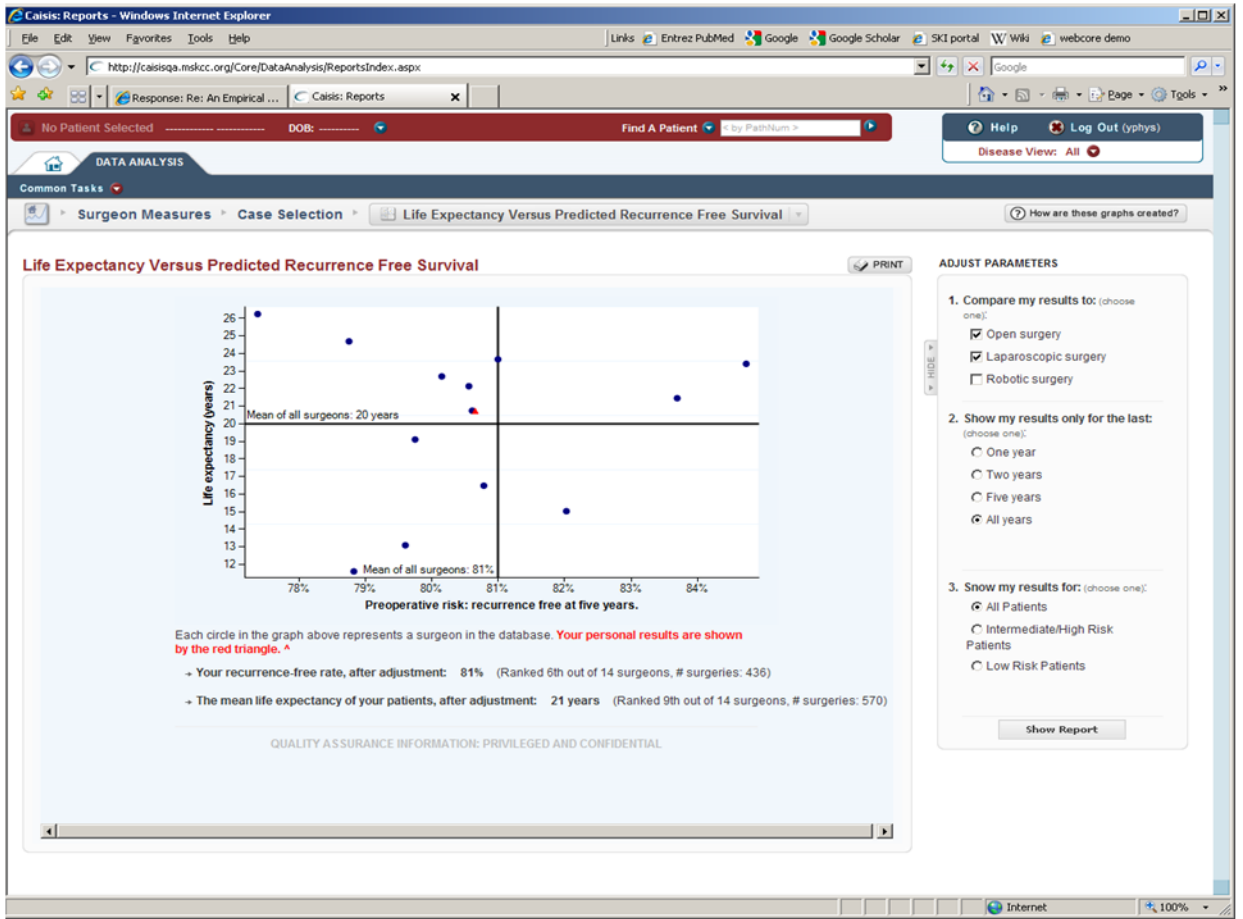


Figure 2. Screen shot from the MSKCC surgeon performance feedback system: case selection
 The mean preoperative risk of a surgeon’s patients at five years on the x axis is graphed against mean life expectancy on the y axis.

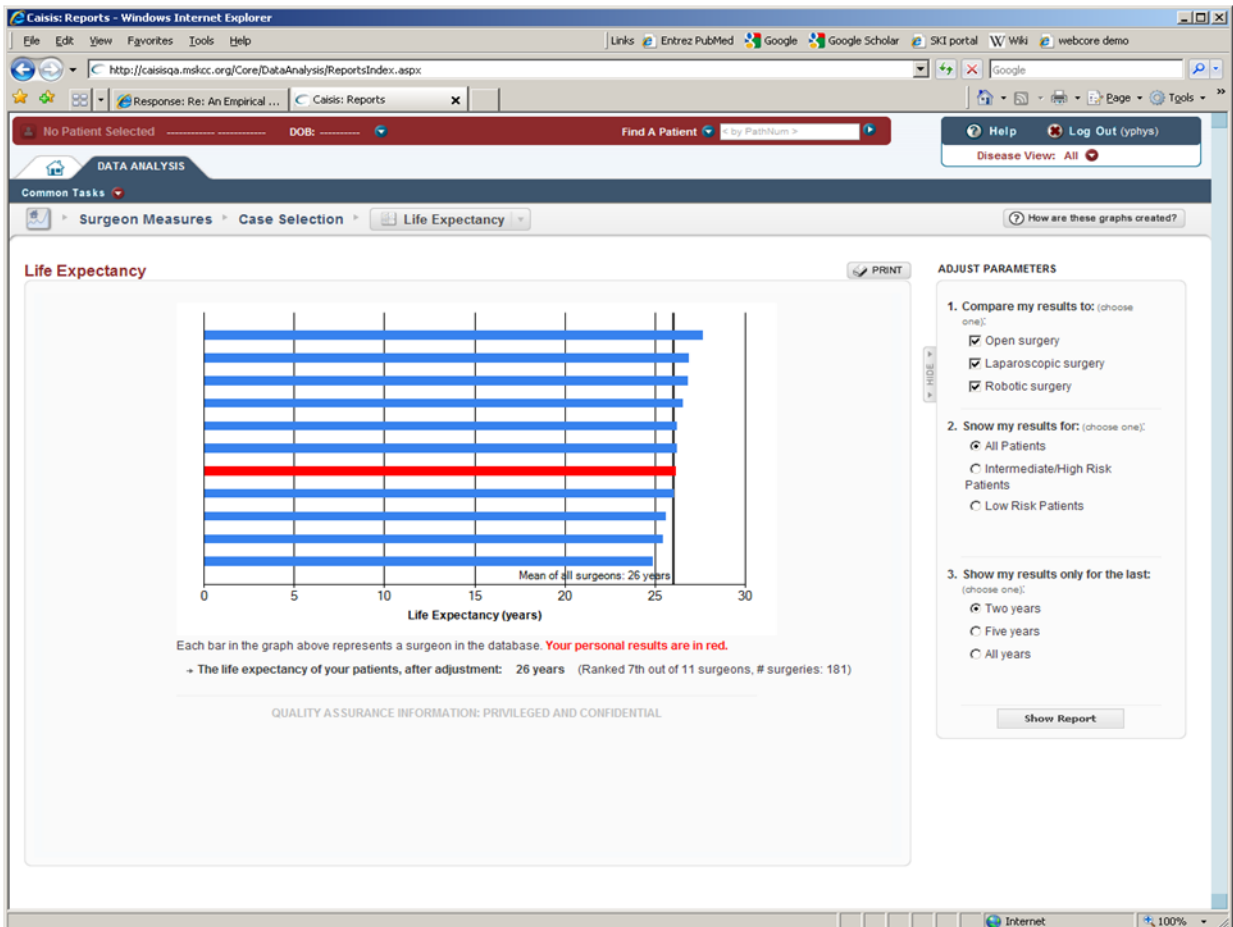


Figure 3. Screen shot from the MSKCC surgeon performance feedback system: life expectancy Mean life expectancy is shown for different surgeons. The analysis is restricted to the previous two years, and to low risk patients only.

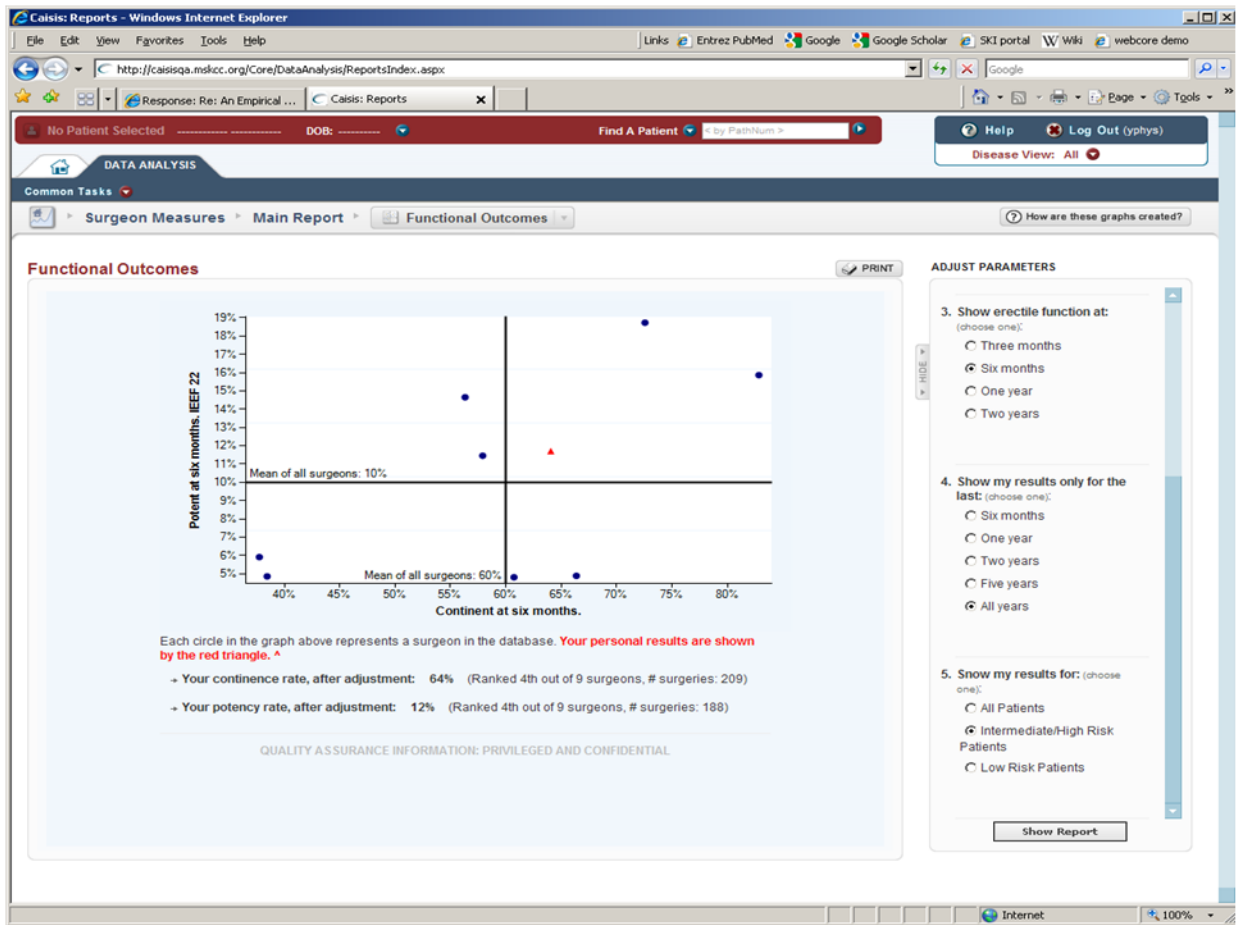


Figure 4. Screen shot from the MSKCC surgeon performance feedback system: functional outcomes

The proportion of a surgeon’s patients who are continent at six months on the x axis is graphed against the proportion of patients potent at six months on the y axis. Both estimates are adjusted for case mix. The analysis is restricted to intermediate and high risk patients.

Table 1

Description of the endpoints available on the surgeon performance feedback system. A full description of each endpoint is available as part of the online system.

Endpoint	Comments
Main outcomes	
Biochemical Recurrence Rate	Defined as a PSA of 0.2 ng/ml or more, with a confirmatory rise
Urinary Function Rate	Patients are considered functional if they score 15 or above out of 21 on a five-item scale
Erectile Function Rate	Patients are considered functional if they score 22 or above out of 30 on the International Index of Erectile Function-6
Functional Preservation Rate	Patients who have both good urinary and erectile function
Perioperative outcomes	
Estimated Blood Loss	Reported separately for open, laparoscopic and robotically- assisted surgery
Length of Stay	Calculated in terms of overnight stays
Positive Surgical Margin Rate	
Complication Rate	Given separately for major vs. all complications
Operating Room Time	Reported separately for open, laparoscopic and robotically- assisted surgery
Reporting rates and volume	
Functional reporting rate	Proportion of patients providing functional outcome data for given time-points
Biochemical Recurrence Reporting Rate	Proportion of patients with a PSA at given time-points
Surgical Volume	Number of cases treated within the time period specified by the user
Case selection	
Life Expectancy	Calculated from the US life tables, with a modification to account for lower comorbidity amongst radical prostatectomy candidates
Inappropriate Surgery Rate	Patients eligible for active surveillance, and those with low life expectancy relative to recurrence risk, are deemed inappropriate for surgery