



Published in final edited form as:

*Int J Gynecol Cancer*. 2015 March ; 25(3): 512–520. doi:10.1097/IGC.0000000000000378.

## Evaluation of the National Surgical Quality Improvement Program Universal Surgical Risk Calculator for a gynecologic oncology service

J. Brian Szender, MD<sup>a,\*</sup>, Peter J. Frederick, MD<sup>a</sup>, Kevin H. Eng, PhD<sup>b</sup>, Stacey N. Akers, MD<sup>a</sup>, Shashikant B. Lele, MD<sup>a</sup>, and Kunle Odunsi, PhD, MD<sup>a</sup>

<sup>a</sup>Division of Gynecologic Oncology, Department of Surgery, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

<sup>b</sup>Department of Biostatistics, Roswell Park Cancer Institute, Buffalo, NY 14263, USA

### Abstract

**Objectives**—The National Surgical Quality Improvement Program (NSQIP) is aimed at preventing perioperative complications. An online calculator was recently published but the primary studies used limited gynecologic surgery data. The purpose of this study is to evaluate the performance of the NSQIP Universal Surgical Risk Calculator (URC) on the patients of a gynecologic oncology service.

**Study Design**—We reviewed 628 consecutive surgeries performed by our gynecologic oncology service between July 2012 and June 2013. Demographic data including diagnosis and cancer stage, if applicable, were collected. Charts were reviewed to determine complication rates. Specific complications were: death, pneumonia, cardiac complications, surgical site or urinary infections (SSI, UTI), renal failure, or thromboemboli (VTE). Data were compared with modeled outcomes using Brier scores and ROC curves. Significance was declared based on  $p < 0.05$ .

**Results**—The model accurately predated death and VTE, with Brier scores of 0.004 and 0.003, respectively. Predicted risk was 50% greater than experienced for UTI; the experienced SSI and pneumonia rates were 43% and 36% greater than predicted. For any complication, the Brier score, 0.023, indicates poor performance of the model.

**Conclusions**—In this study of gynecologic surgeries, we could not verify the predictive value of the URC for cardiac complications, SSI, and pneumonia. One disadvantage of applying a URC to multiple subspecialties is that with some categories, complications are not accurately estimated. Our data demonstrate that some predicted risks reported by the calculator need to be interpreted with reservation.

\*Corresponding author: J. Brian Szender, MD, Division of Gynecologic Oncology, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, Phone: (716) 845-5776, Fax: (716) 845-7608, James.Szender@RoswellPark.org.  
For Reprints: Please contact: Kunle Odunsi, MD, PhD, Chair, Division of Gynecologic Oncology, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, Phone: (716) 845-8376, Fax: (716) 845-1595, Kunle.Odunsi@RoswellPark.org

The authors report no conflict of interest

There are no relevant conflicts of interest regarding the publication of this paper.

## Keywords

Postoperative complications; Risk assessment; Outcomes Research; Gynecologic Surgery; Perioperative care

---

## Introduction

Perioperative complications are a significant cost to the healthcare system<sup>1</sup>. Whether deserved or not, the frequency of peri-operative morbidity and mortality has been put forth as a reflection of surgical and hospital performance. With the implementation of the Affordable Care Act and a link between physician performance and reimbursement there has been a new focus on reducing the frequency of complications related to surgery<sup>2</sup>. The American College of Surgeons (ACS) in 1994 formed the National Surgical Quality Improvement Program (NSQIP), whose goal it is to track and monitor performance of hospitals, surgical services, and individual surgeons<sup>3</sup>. The NSQIP monitors specific quality measures identified as preventable post-operative outcomes.

In 2013 the ACS and NSQIP released a Universal Risk Calculator (URC) built by analyzing the outcomes of 1,414,006 individual procedures performed across 9 surgical specialties in 393 hospitals across the country<sup>4</sup>. The calculator uses 14 different patient-related factors to determine a probability of perioperative morbidity and mortality and serves as a discussion tool for consulting with patients and perioperative planning. However, indication for surgery (or pre-operative diagnosis) is not factored as one of the variables. Physicians can further adjust the probability scores reported by the calculator based on their clinical judgment about how “high risk” a patient is, as this has been shown to predict outcomes<sup>5</sup>. The procedures tracked are at community as well as academic hospitals and they include procedures performed by surgeons at all levels of training/experience.

Because of the pressure that hospitals face related to improving performance and reducing costs, a logical next step following introduction of a calculator is to require pre-operative risk assessment of patients to determine if a non-surgical treatment alternative, if available, should be pursued. The purpose of this study was to report on the performance of the NSQIP URC for surgical procedures performed by gynecologic oncologists.

## Patients and Methods

### Patient Selection

Approval of this study was granted by the institutional review board of the Roswell Park Cancer Institute (Buffalo, NY). The surgical case logs for the 2013 academic year (July 1, 2012 – June 30, 2013) were examined and charts were eligible for further evaluation if (1) they underwent a surgical procedure and (2) the gynecologic oncology service was the primary surgical team caring for the patient. We excluded patients undergoing procedures performed at hospitals away from the primary site to remove differences in institutional nursing care practices as a confounding factor for patient outcomes. Because this is a descriptive study, a power calculation was not necessary to determine sample size. Rather, a

one year sample was chosen to avoid differences in the seasonality of complications based on surgeon experience at our academic institution.

### Chart Abstraction

We reviewed the charts of each eligible patient and recorded pre-operative risk factors as well as surgical history and whether the patient had a history of prior chemotherapy or radiation. We recorded operative details including diagnosis, cancer stage (if applicable), procedure(s) performed, and estimated blood loss. We reviewed all post-operative progress notes and clinic visit notes up to 30 days after the surgical procedure. We recorded all documented post-operative morbidities tracked by NSQIP according to established diagnostic guidelines<sup>6</sup>. We made note of positive and negative cultures, as well as initiation of any antibiotic therapy.

Functional status was coded as “independent” if the Eastern Cooperative Oncology Group (ECOG) score was 0 or 1, corresponding to no limitations or only limited in strenuous activity<sup>7</sup>. ECOG score 2, meaning the patient can carry out self-care but not work activities, was coded as “partially dependent.” An ECOG 3 or greater was coded as “totally dependent” corresponding to limited self-care (ECOG 3) or no self-care (ECOG 4)<sup>7</sup>. Disseminated cancer was defined by International Federation of Gynecology and Obstetrics (FIGO) staging of III or IV. A sensitivity analysis was performed to evaluate the appropriate definition of “disseminated cancer” as Stage III and IV or simply Stage IV. Prior abdominal surgery was coded as “0” for no prior abdominal surgery, “1” for laparoscopic surgery only, “2” for minor open abdominal surgery such as a cesarean section, tubal ligation, or appendectomy, and “3” for a history of major open abdominal surgery including an abdominal hysterectomy, cancer staging surgery, or prior bowel surgery (other than laparoscopic appendectomy for an un-ruptured appendix).

The identified risk factors and morbidities were compared against the hospital's independent NSQIP auditor's database that is maintained in an office and by personnel separate from each surgical service.

### Risk Assessment

With the assistance of the hospital coding specialist, Current Procedural Terminology (CPT) codes were assigned for each of the surgical procedures performed. Where multiple procedures were performed that could not be captured by a single code, multiple codes were assigned. After all procedures had codes assigned, the procedure as well as the pre-operative demographics were input into the calculator and the percent predicted risk for each outcome was recorded. The risk calculator also allows for alteration of the patient's pre-operative risk based on the surgeon's assessment of their overall health status and to incorporate any other comorbid conditions not included in the NSQIP calculator. Patients could be identified as requiring “no adjustment necessary”, “risk somewhat higher than estimate”, or “risk significantly higher than estimate.”

We recorded the outcomes predicted for “average” patient, as well as for the individual patients at each of the “risk” levels allowed by the calculator. For patients who underwent

multiple procedures we multiplied the complements of the individual probabilities and used the complement of that product as a rough estimate of the combined probability.

$$\text{probability}_{\text{combined}} = 1 - [(1 - \text{probability}_1)(1 - \text{probability}_2)]$$

This calculation assumes that the probability of a complication due to one procedure is not related to the probability of a complication due to the second procedure. We also established performance of the calculator for patients with a cancer diagnosis versus those that underwent surgery for benign indications. We evaluated the performance of gynecologic surgeons performing “general surgery” procedures including bowel resection and hernia repairs. We defined general surgery CPT codes as those less than 55000.

### Statistical Analysis

Categorical variables were compared using the Chi-squared test and continuous variables were evaluated using the Student's t test. Significance was confirmed for  $p < 0.05$ . Receiver Operator Characteristic (ROC) curves, which evaluate the ability of the forecast model to discriminate between events and non-events, along with corresponding c-statistics (the area under the curve), were calculated<sup>8</sup>. The Brier score<sup>9</sup>, a statistical method used to evaluate the accuracy of forecasting models, was employed to compare the predicted outcome with the experienced outcomes. The Brier score ranges from 0 (best) to 1 (worst) and is calculated by assigning a value of zero to each non-event and one to each event. The Brier score is a mean-squared error value and in the previously reported NSQIP related literature by Bilimoria, Cohen, *et al.*<sup>4,10</sup> a theoretical threshold of 0.01 (which corresponds to 90% forecast accuracy) was considered “good performance”. The Brier Skill Score is related to the Brier Score and answers the question “does the new model perform better than the current standard?” It is calculated as one minus the ratio of new Brier score to the standard Brier score and values range from negative infinity (worst), to 1 (best).

### Results

We identified 628 patients who underwent surgery during the defined time period. Table 1 summarizes the demographic data for the patients. The average age was 53 years. There were 286 patients that had a cancer diagnosis and of these 38% had advanced cancer (Stage III or IV disease). There were 51 patients excluded from analysis because their procedure did not have a CPT code recognized by the URC and an additional 32 patients were analyzed using a “similar” CPT code because there were inadequate numbers of patients in the calculator dataset with the reported CPT code. Thus, there were 577 patients included in the remainder of the analysis. The most common surgery performed was a total abdominal hysterectomy with bilateral salpingo-oophorectomy, totaling 63 out of the 577 (10.9%) of cases. Seven patients had synchronous ovarian and endometrial cancers and were analyzed in the ovarian cohort.

The patients excluded due to unrecognized CPT codes had a better functional status ( $p=0.02$ ) and less disseminated cancer ( $p<0.0001$ ) cancer than the group that was able to be

analyzed with the URC. Furthermore, experienced complication rates differed ( $p=0.02$ ) with more complications in the analyzed group than the excluded group.

The complications are summarized in Table 2 along with the associated model predictions; the rate for any complication was 13.7%, similar to the rate published by Uppal *et al.* for another academic gynecologic oncology service<sup>11</sup>. Predicted complication rates by the URC are summarized in Supplemental Table A.1. The average predicted complication rates were  $9.1\pm 0.8\%$  for the procedures performed (no patient specific information included),  $10.7\pm 0.8\%$  without risk alteration,  $14.2\pm 1.1\%$  at the middle level, and  $18\pm 1.4\%$  at the highest level of risk alteration. The Brier scores and c-statistics for the URC predicted outcomes are summarized in Supplemental Tables A.2 and A.3, respectively. The predicted rates and Brier scores are additionally displayed in Figure 1. The Brier scores for all NSQIP outcomes are presented in Figure 2. The horizontal dashed line indicates the NSQIP published performance of 0.01, which corresponds to 90% forecast accuracy<sup>9</sup>. The Brier score approaches the threshold for death in cancer patients, pneumonia in all patients, cardiac complications in non-cancer patients, urinary tract infection (UTI) in cancer patients, venous thromboembolic event (VTE) in all patients, and renal failure in non-cancer patients. Surgical Site Infection (SSI) was not accurately predicted. The model did not accurately predict outcomes for “any complication” or “serious complication” in any subgroup analysis. When Stage III cancer was treated as a “disseminated” cancer for purposes of the risk calculator, the rates were further from the experienced outcome based on a Brier score that was further from 0. This difference, however, was only present when all patients were included. During individual subgroup analysis there was no improvement in the performance of the model based on stage, in fact, increasing stage corresponded with trends toward worse performance (Figure 1d). When Brier scores were computed for each pre-operative risk stratum, a surgeon's assessment of higher pre-operative risk led to worse performance of the model for all complications except renal failure (Table A.2).

The sensitivity analysis for classification of “disseminated cancer” as Stage III versus Stage IV is also included in the supplementary material, with stratification by primary tumor site (Tables A.4 and A.5). The model predicts outcomes for patients with benign disease but in patients with a cancer diagnosis the model does not accurately predict post-operative morbidity and mortality. When predictions are stratified by primary cancer site, as represented in Figure 1c, the model more accurately predicts outcomes in uterine, cervical, and vulvar cancers but not ovarian (with Fallopian tube and primary peritoneal) cancer. For Stage I cancer the model predicts four types of complications, for stage II it predicts three. Whether Stage III disease is treated as disseminated cancer or not the model predicts three types of complications. Finally, for all procedures performed on patients with stage IV cancer the URC predicts four types of complications.

ROC analysis failed to demonstrate any significant difference from the unity line ( $A = 0.5$ ). Representative ROC curves are presented in Figure 3. While some absolute values of the c-statistic were less than 0.5, meaning the ROC curve is below the unity line the confidence intervals always included the unity line.

When a comparison was made between patients that experienced a complication and those that did not experience a complication, as illustrated in Figure 4, there were no other differences between the predicted rates of any complications. Two exceptions were the generic (Figure 4a) and high risk (Figure 4d) predictions for death, with  $p < 0.001$  and  $p = 0.002$ , respectively.

## Discussion

The URC functions in a limited capacity for gynecologic oncology cancer cases. When gynecologic oncologists perform surgery for benign conditions, the URC does a better job predicting outcomes. The reason that URC does not perform as well for cancer cases may be related to systemic changes that coexist with the disease independent of cancer stage. Unfortunately in many situations the URC does not accurately predict outcomes and more importantly the URC only performs as well as simply taking the overall procedure risks and in some cases the performance is actually worse than just quoting the general risks of a procedure, as evidenced by a higher Brier score and a lower c-statistic when patient-specific parameters are included. Further, Brier Skill Scores were typically negative, indicating worse performance of the patient specific model than the generic procedure-related risks.

The reasons the URC is not directly applicable to gynecologic oncology surgeries are likely multifactorial. First, the creation of a universal calculator that incorporates multiple and heterogeneous surgical specialties and that is applied to private as well as academic settings necessitates that the degree of precision of the calculator must decrease. Second, the URC seems to perform better when there is not a coexistent cancer diagnosis. Interestingly, if all cancer patients except those with ovarian cancer (including Fallopian tube and primary peritoneal) are included in the analysis, the URC actually performs better at predicting that some morbidity will occur after surgery, although not for predicting which complication (pneumonia, SSI, UTI, or VTE) a patient will encounter. This finding suggests that there might be some process related to intraperitoneal disease that is too complex for the calculator to compensate for. Finally, the URC forces some CPT codes to be re-assigned to “related” codes because the numbers of individual procedures were not large enough to reach statistical significance. This was true for 5.5% of our analyzed procedures. With re-assignment of some surgical procedures to “related” codes there might be some deviation from the actual expected probabilities, because while some of the re-assigned codes are similar, such as analyzing ovarian transposition (CPT code 58825) as bilateral salpingo-oophorectomy (CPT code 58940), our impression of other code re-assignments does not make the same clinical sense, such as vulvar laser ablation (CPT code 56515) treated as partial vulvectomy (CPT code 56620). Additionally, there are some procedure translations that do not seem to be clinically similar procedures, such as re-assigning a dilation and curettage (CPT code 58120) as an abdominal myomectomy with removal of one to four fibroids weighing 250 g or less (CPT code 58140). When these re-assigned surgeries were removed from the analysis the Brier scores did not significantly change.

Because none of the 51 patients excluded from analysis due to unrecognized CPT codes had disseminated cancer, a factor shown above to induce worse performance in the model, it is possible that inclusion of the would have improved the overall performance of the URC.

However, the lack of ability to analyze those cases underscores the lack of direct applicability of the URC to a gynecologic oncology service.

This URC also only allows input of a single procedure CPT code, which means that the CPT code entered for a complex tumor debulking surgery that involves hysterectomy as well as bowel resection will not be completely encompassed with a single calculation. We overcame this issue by combining the risk predictions from multiple procedures, rather than selecting the highest risk procedure performed. For example, the typical patient undergoing a radical ovarian debulking surgery (CPT code 58952) with end colostomy and a Hartmann pouch (CPT code 44143) is predicted to have an 18.3% chance of any complication based on her debulking surgery and a 41.9% chance of any complication because of her colectomy. For most of the reported complications, the risks of the colectomy are higher; however the risk of urinary tract infection is 2% higher (5.8% vs. 3.8%) for the debulking part of the procedure. By combining the complements of the percentages, the predicted risk of any complication becomes 52.5% and the risk of urinary tract infection becomes 9.4%. This calculation predicts the probability for complications from a procedure with multiple CPT codes greater than either individual probability. While this methodology has not been previously validated with the URC, there is a logical need to adjust the prediction for more complex procedures, since the chance of post-operative morbidities can be expected to increase with more procedures. With increased numbers of complex procedures with multiple CPT codes, further refinement in the calculation can be performed by subtracting a factor representing the interaction between each of the procedures.

Other risk prediction calculator and models have been attempted in the past<sup>10-17</sup> including two by the designers of the URC<sup>10,14</sup>. Models that are more specialized, such as the colorectal surgery and pancreatectomy perioperative morbidity prediction models do a good job predicting the outcomes they were designed for, but they have a more narrow scope. The creators of the URC compared their URC with a specialized risk calculator for colorectal surgery. As expected their colorectal surgery predictor was more precise than the URC, but both performed reasonably well. Other pre-operative risk evaluation models, such as stratification by patient albumin<sup>11</sup> predict complications when below a critical values, however these values were not found to be independently predictive of outcomes<sup>4</sup>.

As the Affordable Care Act becomes implemented and outcome measures are tied to physician and hospital reimbursement, external pressure on hospitals to limit performance of high risk surgeries will lead to hospital-based strategies to stratify the risk of patients before undergoing surgery. The URC is a good step in the direction of achieving this goal; however the results must be interpreted with caution, especially when there is a concurrent cancer diagnosis. Furthermore, with changes in the Medicare oversight of the 11 largest cancer centers in the US that have otherwise been exempted from some of the same outcome-based measures that are in place at other institutions around the country, we as a specialty must remain involved in the discussion so that surgeries deemed clinically necessary can be performed on our patients without undue penalties levied on the hospital or surgeon<sup>18</sup>.

As stated previously, a limitation of this study is there were only 577 cases included in the analysis compared with the more than 1 million cases used to build the actual URC<sup>4</sup>. This

number reflects only slightly more favorably when considering that only 5.3% or 74,737 of the cases were performed by gynecologic surgical services<sup>4</sup>. Another limitation of this study is that charts were abstracted retrospectively rather than collected in a prospective database. Retrospective reviews have the recognized possibility of misclassification error due to missing information and subsequent recall bias; however, these risks are minimized in the current study since abstracted data were compared against our institution's prospectively collected NSQIP database and found to be similar. Since this was a descriptive study no power analysis was performed, despite this lack of analysis there were numerous differences identified between the model predictions and clinical experience, suggesting the number of patients in the study was adequate to detect differences. Furthermore, all procedures were performed by a single gynecologic oncology service with five attending surgeons at a single academic institution, which may introduce some bias for individual practice styles and with the presence of trainees, some surgeon inexperience for cancer surgeries.

The ACS NSQIP collects clinically relevant pre-operative characteristics; however the URC fails to adequately incorporate these factors into a useful forecast model for gynecologic cancer patients. Further work is necessary to accurately integrate the site of origin of a gynecologic cancer to solidify the URC as a useful tool for the gynecologic oncologist.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

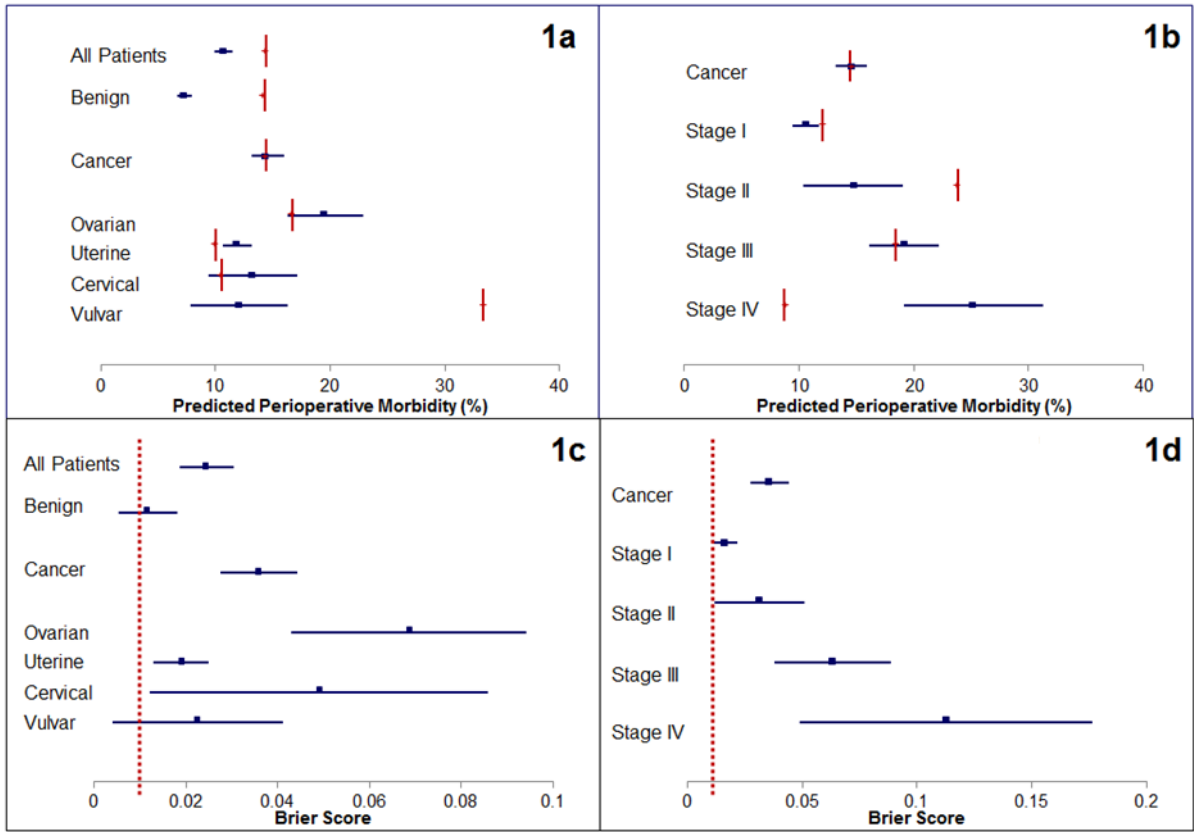
This work was supported by Roswell Park Cancer Institute grant NCI P30CA016056 and NIH 5T32CA108456.

## References

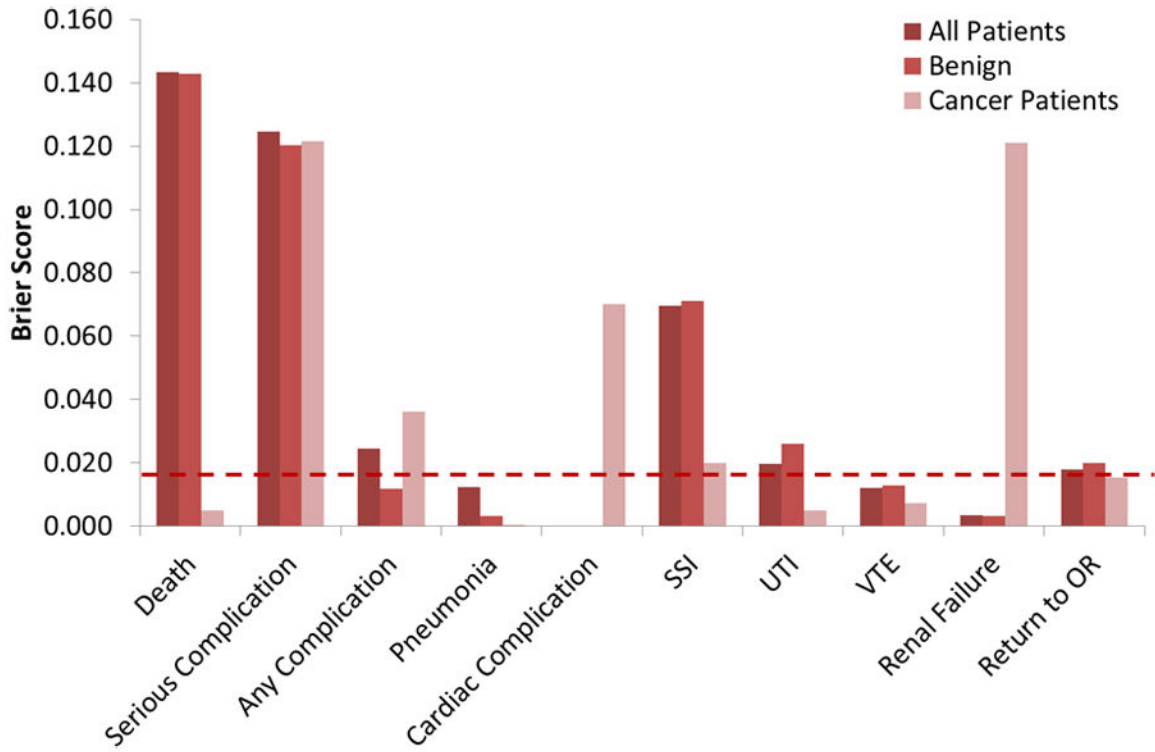
1. Dimick JB, Chen SL, Taheri PA, Henderson WG, Khuri SF, Campbell DA Jr. Hospital costs associated with surgical complications: a report from the private-sector National Surgical Quality Improvement Program. *J Am Coll Surg*. 2004 Oct; 199(4):531–1.
2. Internal Revenue Service, Treasury. Computation of, and rules relating to, medical loss ratio. Final regulations. *Fed Regist*. 2014 Jan 7; 79(4):755–8. [PubMed: 24396939]
3. Ingraham AM, Richards KE, Hall BL, Ko CY. Quality Improvement in Surgery: the American College of Surgeons National Surgical Quality Improvement Program Approach. *Ann Surg*. 2010; 44:251–67.
4. Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmieciak TE, Ko CY, Cohen ME. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons. *J Am Coll Surg*. 2013 Nov; 217(5):833–42. [PubMed: 24055383]
5. Cohen ME, Bilimoria KY, Ko CY, Richards K, Hall BL. Effect of Subjective Preoperative Variables on risk-Adjusted Assessment of Hospital Morbidity and Mortality. *Ann Surg*. 2009; 249:682–9. [PubMed: 19300217]
6. [Accessed December 12, 2013] ACS NSQIP participant use file user's guide. [http://site.acsnsqip.org/wp-content/uploads/2013/10/acsnsqip.puf\\_UserGuide.2012.pdf](http://site.acsnsqip.org/wp-content/uploads/2013/10/acsnsqip.puf_UserGuide.2012.pdf)
7. Oken MM, Creech RH, Tormey DC, et al. Toxicity and Response Criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol*. 1982; 5:649–55. [PubMed: 7165009]



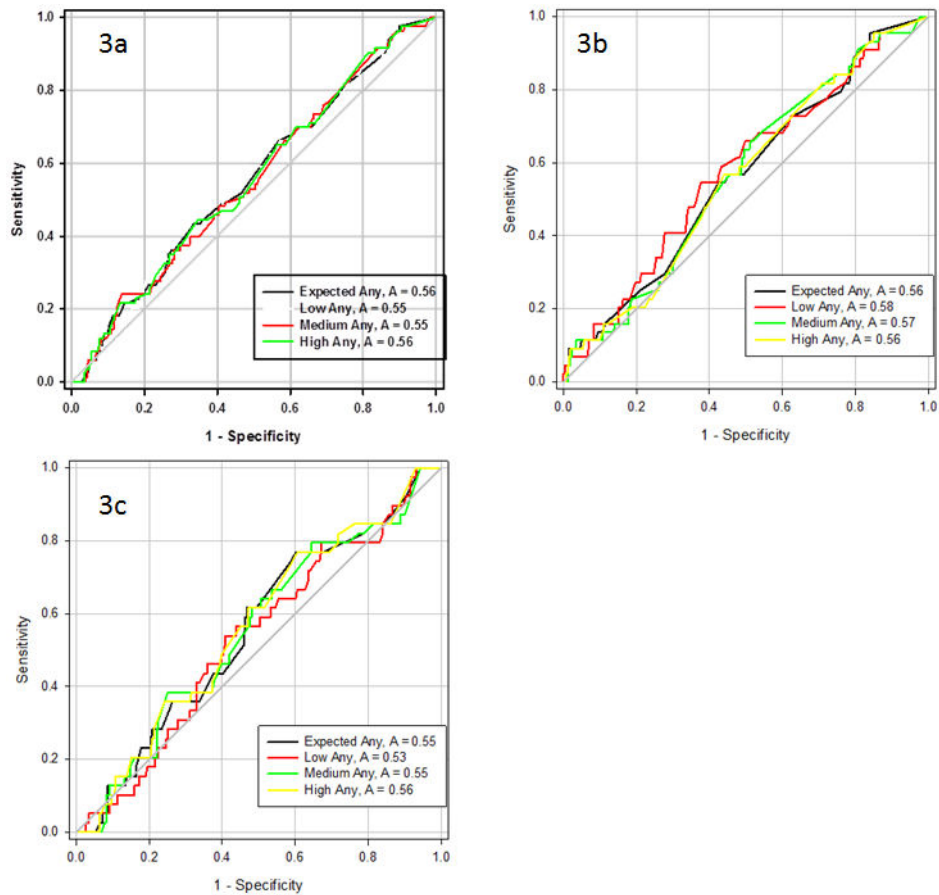
8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. 1988 Sep.;44, 837–45.
9. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Rev.* 1950; 78:1–3.
10. Cohen ME, Bilimoria KY, Ko CY, Hall BL. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg.* 2009; 208:1009–16. [PubMed: 19476884]
11. Uppal S, Al-Niaimi A, Rice LW, et al. Preoperative hypoalbuminemia is an independent predictor of poor perioperative outcomes in women undergoing open surgery for gynecologic malignancies. *Gyn Onc.* 2013; 131:416–22.
12. Finks JF, Kole KL, Yenumula PR, et al. Predicting risk for serious complications with bariatric surgery: results from the Michigan Bariatric Surgery Collaborative. *Ann Surg.* 2011; 254:633–40. [PubMed: 21897200]
13. Gupta PK, Gupta H, Sundaram A, et al. Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation.* 2011; 124:381–7. [PubMed: 21730309]
14. Parikh P, Shiloach M, Cohen ME, et al. Pancreatectomy risk calculator: an ACS-NSQIP resource. *HPB (Oxford).* 2010; 12:488–97. [PubMed: 20815858]
15. Biagioli B, Scolletta S, Cevenini G, Barbini E, Giomarelli P, Barbini P. A multivariate Bayesian model for assessing morbidity after coronary artery surgery. *Critical Care.* 2006; 10:R94. [PubMed: 16813658]
16. Prasad SM, Ferreria M, Berry AM, et al. Surgical Apgar Outcome Score: Perioperative Risk Assessment for Radical Cystectomy. *J Urol.* 2009 Mar.;181, 1046–53.
17. Olsen MA, Higham-Kessler J, Yokoe DS, et al. Developing a Risk Stratification Model for Surgical Site Infection after Abdominal Hysterectomy. *Infect Control Hosp Epidemiol.* 2009 Nov; 30(11):1077–83. [PubMed: 19803722]
18. Cohen DE, Leitao M, Levenback C, et al. Reporting of quality measures in gynecologic oncology programs at Prospective Payment System (PPS)-Exempt Cancer Hospitals: An early glimpse into a challenging initiative. *Gyn Onc.* 2013; 130:403–6.



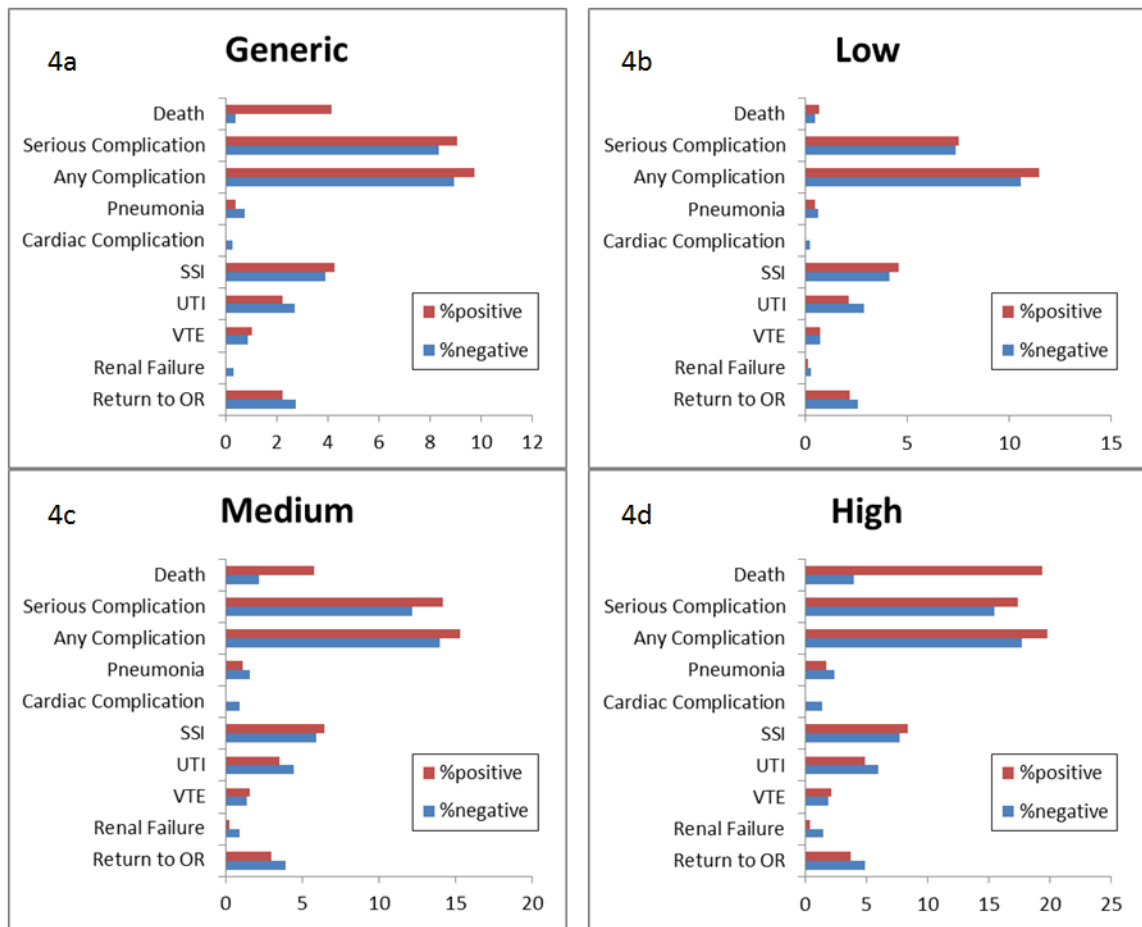
**Figure 1.** Predicted outcomes (blue lines) presented with 95% confidence intervals and experienced outcomes (red vertical lines). Figure 1a also illustrates the predicted outcome based on tissue of origin for different cancers. Figure 1b presents predicted and experienced outcomes of cancer procedures stratified by stage of disease. Figures 1c and 1d illustrates the Brier score result (blue lines) with 95% confidence intervals and the NSQIP expected model outcome (0.01) illustrated by the red vertical line, which corresponds to 90% forecast accuracy.



**Figure 2.** Representative Brier scores for all patients, patients who had surgery for benign indications, and cancer patients. The horizontal dashed line indicates the NSQIP published performance of 0.01, which corresponds to 90% forecast accuracy. The Brier score approaches the threshold for death in cancer patients, pneumonia in all patients, cardiac complications in non-cancer patients, urinary tract infection (UTI) in cancer patients, venous thromboembolic event (VTE) in all patients, and renal failure in non-cancer patients. Surgical Site Infection (SSI) was not accurately predicted.



**Figure 3.** Representative ROC curves for any complication for all patients undergoing surgery (3a), those with benign conditions (3b), and those with malignant conditions (3c). There is no difference between the ROC curves for generic, low, medium, and high risk conditions. There is also no difference between the curves and the unity line ( $A = 0.5$ ).



**Figure 4.** Comparison between predicted rates of complications between those who experienced complications (red) and those who did not experience complications (blue). Data represented are for generic risks (4a), low risk (4b), medium risk (4c), and high risk (4d). Only the generic (4a) and high (4d) predicted risks of death were different between the two populations ( $p < 0.05$ ).

Table 1

## Patient Characteristics

<b>Demographic</b>	<b>mean ± SD</b>
Age (years)	53±14.2
Height (inches)	63.9±2.8
Weight (pounds)	186±62.7
BMI (kg/m <sup>2</sup> )	32±11
Blood Loss (mL)	160±320
Hospital Stay (days)	2.9±5.6
<b>Medical History</b>	<b>Count (%)</b>
Emergency Surgery	5 (0.8)
Steroid Use	8 (1.3)
Ascites	28 (4.5)
Sepsis	1 (0.2)
Ventilator Dependent	1 (0.2)
Disseminated Cancer	23 (3.7)
Diabetes (Oral Medications)	52 (8.3)
Diabetes (Insulin)	33 (5.3)
Hypertension	244 (38.9)
Prior Cardiac Event	35 (5.6)
Congestive Heart Failure	10 (1.6)
Dyspnea	13 (2.1)
Current Smoker	140 (22.3)
COPD	30 (4.8)
Dialysis	2 (0.3)
Acute Renal Failure	3 (0.5)
<b>Prior Abdominal Surgery</b>	<b>Count (%)</b>
None	234 (37)
Laparoscopic Only	64 (10)
Minor Open Abdominal	167 (27)
Major Open Abdominal	163 (26)
<b>Functional Status (ECOG)</b>	<b>Count (%)</b>
0-1	609 (97)
2	14 (2)
3+	5 (1)
<b>Cancer Type</b>	<b>Count (%)</b>
Ovarian	94 (33)
Uterine	126 (44)

<b>Demographic</b>	<b>mean ± SD</b>
Cervix	39 (14)
Vulvar	17 (6)
Vaginal	1 (0.4)
Non-Gyn	8 (3)
Total Cancer	285 (45)
Benign	343 (55)
<b>Cancer Stage (FIGO)</b>	<b>Count (%)</b>
I	156 (55)
II	22 (8)
III	85 (30)
IV	23 (8)

NSQIP defined pre-operative risk factors as well as other demographic information. NSQIP – National Surgical Quality Improvement Program

**Table 2**  
**Experienced and Calculated Complication Rates**

Complication	Actual Complication n (%)	URC Expected Complication n (%)
Death	3 (0.5)	3 (0.5±0.2)
Serious Complication *	81 (2.9)	43 (7.4±0.5)
Any Complication *	86 (13.7)	62 (10.7±0.8)
Pneumonia	7 (1.1)	3 (0.6±0.1)
Cardiac Complication *	0 (0)	2 (0.3±0.04)
Surgical Site Infection *	44 (7)	24 (4.2±0.4)
Urinary Tract Infection	11 (1.8)	17 (2.9±0.2)
Venous Thromboembolic Event	7 (1.1)	4 (0.7±0.1)
Renal Failure	2 (0.3)	2 (0.3±0.06)
Return to Operating Room	11 (1.8)	15 (2.6±0.2)

Summary statistics for all surgical procedures. URC, Universal Risk Calculator. Percent prediction is reported as a point estimate ± 95% confidence intervals.

\* prediction and complication significantly different