

Tailor-made proficiency curves in laparoscopic hysterectomy: enhancing patient safety using CUSUM analysis

A. R. H. Twijnstra · M. D. Blikkendaal · S. R. C. Driessen ·
E. W. van Zwet · C. D. de Kroon · F. W. Jansen

Received: 19 March 2014 / Accepted: 14 October 2014 / Published online: 22 October 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract The objective of this study is to develop a risk-adjusted real-time quality control system in laparoscopic hysterectomy with respect to blood loss, operative time and adverse events in order to signal derailing surgical performance in a timely fashion. Based on prior research, uterus weight, body mass index, number of surgeons, prior abdominal surgery, and type of laparoscopic hysterectomy were identified as significant covariates predicting successful surgical outcome. Cumulative sum (CUSUM) analysis, a model based on dichotomous input (success or “failure”), was selected as a predictive tool for performance analysis. Cutoff values were set at blood loss <200 mL and operative time <120 min and no adverse event. Risk-adjusted CUSUM graphs were constructed. In order to detect progressive failure rates (odds ratio 2.0 compared to average) in surgical performance (for blood loss, operative time, and adverse events) within 20 procedures, as a result, surgeons with average clinical outcomes will be flagged once in every 70–75 procedures (median) without justified derailing performance. With proposed validated and risk-adjusted CUSUM graphs, gynecologists are able to continuously monitor their surgical performance in laparoscopic hysterectomy. Consequently, this identifies suboptimal factors, which allow improvement of their

surgical outcomes (by means of adjustment) and further enhancement of patient safety.

Keywords Hysterectomy · Laparoscopy · Quality control · CUSUM · Patient safety

Introduction

In order to enhance patient safety, it has become increasingly important to measure outcome in health care. Surgical outcomes such as blood loss, operative time, and the occurrence of adverse events are widespread applied instant measures. These measures, as well as skills and experience of the surgeon (usually expressed by the number of performed cases) are currently still used as quality predictors [1]. However, it is also established that surgical outcome, apart from surgical experience, is influenced by co-factors such as the makeup of the OR team and (inherently) patient factors (i.e., the case mix). These factors are not taken into account when the aforementioned crude and unadjusted parameters are used to measure and present the actual surgical outcome [2, 3].

With respect to patient-related factors, recent research in laparoscopic hysterectomy (LH) demonstrated five significant covariates predicting successful outcome: uterus weight, body mass index, number of surgeons present at surgery, prior abdominal surgery, and type of laparoscopic hysterectomy (i.e., total laparoscopic hysterectomy, supracervical laparoscopic hysterectomy, or laparoscopic-assisted vaginal hysterectomy) [4]. Moreover, experience is predicting successful surgical outcome in LH, with respect to blood loss and adverse events, up to at least a hundred procedures. This finding was also observed in the field of advanced colorectal laparoscopic surgery [5, 6]. Finally, recent research demonstrated a

A. R. H. Twijnstra · M. D. Blikkendaal · S. R. C. Driessen ·
C. D. de Kroon · F. W. Jansen (✉)
Department of Gynecology, Leiden University Medical Center, PO
Box 9600, 2300 RC Leiden, The Netherlands
e-mail: f.w.jansen@lumc.nl

E. W. van Zwet
Department of Medical Statistics, Leiden University Medical Center,
Leiden, The Netherlands

F. W. Jansen
Department of Biomechanical Engineering, Delft University of
Technology, Delft, The Netherlands

significant experience independent and case mix-adjusted surgical skills factor (SSF) with regard to successful outcome in LH [4].

The aforementioned findings support that surgical outcomes in laparoscopic hysterectomy should be monitored consecutively, as both case mix and surgeon's skills may vary over time, and experience alone is not sufficiently predicting these outcomes. Parallel to the traditional outcome measures, the traditional single outcome learning curves in surgery, which were applied in order to assess surgical proficiency, do not take these findings into account [7–10]. Monitoring tools based on cumulative sum (CUSUM) analysis, already used in obstetrics and general surgery, overcome these shortcomings [11–16]. In the industrial setting, since 1974, CUSUM charts have been shown to be ideally suited to detect relatively small persistent changes in the event rates over time [3]. Traditional CUSUM approaches, however, make no adjustment for different risk profiles because machine inputs are usually relatively homogeneous. In contrast, patients undergoing a particular surgical intervention are often very heterogeneous in their clinical presentation. Additionally, the surgical approach may vary considerably due to the clinical presentation as well as the preference of the surgeon. As a result, the probability of successful outcome may vary considerably between patients. By using a likelihood-based scoring method, the cumulative sum procedure is adapted so that it adjusts for the surgical risk of each patient estimated preoperatively [2, 17, 18]. As a result, the user will be provided with a graphical representation of its surgical outcomes corrected for patient mix and instantly compared to the national average. Trends will be visualized, and significant deterioration in surgical outcome will be noticed.

In gynecology, nowadays, a shift in implementing more advanced surgical procedures is observed. However, several studies suggest that these advanced laparoscopic surgical

procedures are characterized by a specific proficiency gaining curve due to the acquirement of unique operative skills [19]. Consequently, this learning curve is considered a barrier for widespread implementation of advanced laparoscopic surgery [5]. Other research already revealed that even in basic laparoscopy, nearly a fifth of surgeons never gain proficient skills to perform laparoscopic surgery adequately [20]. These insights, combined with the call for constant monitoring of patient safety, make us strive for risk-adjusted continuous quality assessments during mentorships and beyond in order to adjust performance when quality of surgery is at risk.

The aim of this study is to develop such a tool. In order to signal derailing surgical performance in a timely fashion, a risk-adjusted real-time quality control system for laparoscopic hysterectomy is analyzed, inquired, and launched.

Methods

A previously described data set of 1,534 LHs, performed by 79 surgeons, was used to validate and compose a risk adjusted CUSUM graph in LH [4]. Significant predicting covariates were included. These consisted of uterus weight, body mass index, number of surgeons present at surgery, prior abdominal surgery, and type of laparoscopic hysterectomy (Table 1).

The CUSUM score depends on four factors: the current average level of surgical performance, a chosen level of surgical performance deemed undesirable, the patient's surgical risk estimated preoperatively, and the actual surgical outcome in this patient. Preoperative surgical risk estimation was based on body mass index, uterus weight, and prior abdominal surgery. With respect to the continuous surgical outcomes, blood loss, and operative time, these were dichotomized using the rounded mean observed value. Consequently, successful surgical outcome was determined as blood loss <200 mL, operative time <120 min, and no adverse event. Because incidences of these outcomes varied, with accompanying

Table 1 Association between predictors and primary outcomes in laparoscopic hysterectomy

	Blood loss (>200 mL)	Operative time (>120 min)	Adverse event (yes)
Uterus weight increase per 100 g	<i>0.33 (P<0.0001)</i>	<i>0.40 (P<0.0001)</i>	<i>0.18 (P=0.0002)</i>
Body Mass Index increase per 1 point (kg/m ²)	<i>0.28 (P<0.0001)</i>	0.18 (P=0.0841)	0.02 (P=0.221)
Numbers of previous abdominal surgeries	0.19 (P=0.54)	0.78 (P=0.782)	<i>0.48 (P=0.048)</i>
Two surgeons (vs. one)	-0.47 (P=0.072)	<i>0.64 (P=0.028)</i>	0.05 (P=0.811)
LAVH vs. TLH	<i>0.91 (P=0.0274)</i>	0.04 (P=0.915)	0.33 (P=0.306)
SLH vs. TLH	-0.14 (P=0.482)	<i>-0.47 (P=0.032)</i>	<i>-0.52 (P=0.079)</i>

Positive predictors represent higher chance of suboptimal primary outcome, and negative predictors represent lower chance of suboptimal primary outcome. Italicized items are significant predictors

LAVH laparoscopic-assisted vaginal hysterectomy, TLH total laparoscopic hysterectomy, SLH supracervical laparoscopic hysterectomy

varying influences of covariates, we applied three risk-adjusted CUSUM graphs, one for each outcome.

With the chosen level of surgical performance deemed undesirable, we aimed to minimize the number of procedures before possible derailing performance is signaled, while minimizing “false alarms”. For quality control, a lower boundary line is not used. To allow a sensitive and timely detection of “eventful” procedures, this model resets itself to 0, each time the *x*-axis is hit [18]. As a consequence, the median number of procedures needed to detect an unacceptable failure rate (in case a surgeon performs below an acceptable level) is based on the upper boundary (“out of control”, odds ratio of 2 compared to average performance). Nevertheless, this model cannot prevent that also average clinical performance every once in a while is “flagged” as derailing (Fig. 2). The primary outcome of this study is the number of procedures after which surgeons are flagged, both true positive and false positive.

To apply a risk-adjusted (i.e., based on the patient’s surgical risk estimated preoperatively) CUSUM analysis, we have to estimate the logistic regression model as described earlier [4]. Based on this model, we can compute the probability of an unfavorable outcome (failure) for each procedure. For ease of notation, suppose we use only uterus weight *Ut* as a predictor. Then, provided that the surgeon is performing exactly on the national average (i.e., is in control), the probability of failure in procedure *i* is:

$$p0(i) = 1/(1 + \exp(-\beta0-\beta1 \times Ut(i)))$$

- $\beta0$ the intercept in the logistic regression model
- $\beta1$ log odds ratio for uterus weight

If the surgeon performs worse than average (OR=2 compared to the national average), the probability of failure becomes larger and is given by:

$$p1(i) = p0 = 1/(1 + \exp(-\beta0-\beta1 \times Ut(i)-\log(2)))$$

Given the outcome of procedure *i*, we can compute the log likelihood ratio as

$$W(i) = \begin{cases} \log(p0/p1) & \text{if failure} \\ \log((1-p0)/(1-p1)) & \text{if success} \end{cases}$$

Now, we construct the CUSUM graph by plotting $X(i) = \max(0, X(i-1) + W(i))$

This *X* will provide the actual direction and weight of the outcome of procedure *i* on the CUSUM graph corrected for uterus weight. In our model, we included all covariates (uterus weight increase per 100 g, BMI increase per 5 points, numbers

of prior abdominal surgeries, 1 or 2 performing surgeons, and type of laparoscopic hysterectomy).

Results

Figure 1 provides an example of the principle of a risk-adjusted CUSUM graph of 21 consecutive LHs in one surgeon with respect to blood loss <200 mL. The horizontal axis represents the numbers of consecutive procedures. The vertical axis represents the cumulative sum of the risk-adjusted scores per procedure. As can be seen at “no. 1,” the fourth procedure was complicated by blood loss >200 mL in a regular patient, followed by three regular procedures with blood loss <200 mL. The eighth procedure (no. 2) was performed uneventful in a “challenging patient” (e.g., high BMI and large uterus weight). At the 13th procedure (no. 3), blood loss >200 mL occurred; however, this occurred in a challenging case (high BMI and large uterus weight compared to no. 1). At the 15th procedure, another failure occurred (no. 4), however, because of average patient characteristics (see also Table 1); this procure was expected to be performed uneventful. A steep rise on the curve represents this discordance between the observed and expected outcome. At attempt number 21 (no. 5), the CUSUM graph goes out of control. Consequently, the chart signals.

For the defined outcomes of LH, respectively blood loss, operative time, and adverse events separate risk-adjusted CUSUM graphs that were constructed. In order to detect unacceptable failure rates (clinical performance OR 2.0 compared to average clinical performance) in surgical

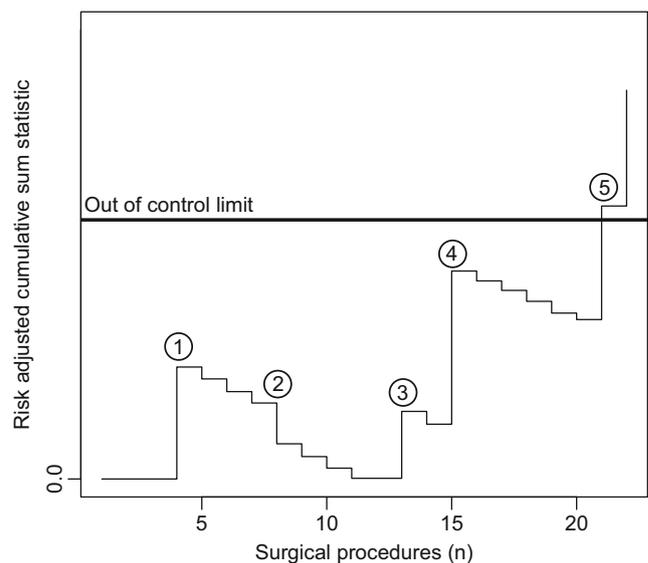


Fig. 1 Example of a cumulative summation analysis graph in one gynecologist with respect to blood loss <200 mL (see “Results” for explanation)

performance within 20 procedures, as a result, a surgeon with average surgical outcomes will be flagged without justified bad performance once in approximately every 70–75 procedures respectively (Fig. 2). Reference values are based on the previously described cohort of 1,534 procedures performed by 79 gynecologists.

Once one of the three CUSUM graphs signals, one should analyze at least 20 of its past performed procedures using a concise checklist, as depicted in Table 2. Five fields address possible causes. If one or more fields are ticked once or more, this field should be studied and addressed in particular. This checklist is not validated yet.

Web-based non-commercial and protected application is available in order to process the proposed CUSUM graphs in the field of LH in order to provide the surgeon his/her performance statistics at a glance (<https://www.qusum.org>). The program is primarily designed for a national multicenter validation study; however, one is free to register and apply the application. This software should be easily integrated with (existing) data recording systems in the near future. The five characteristics (uterus weight in grams, body mass index (kg/m^2), number of previous abdominal surgeries, one or two surgeons, type of LH, and the three primary outcomes (operative time in minutes, blood loss in milliliters, and adverse event) can be entered immediately postoperatively or at any given moment.

Fig. 2 Threshold curves for blood loss, operative time, and adverse events. Horizontal axis represents amount procedures before flagging in case of out-of-control performance (OR 2.0 compared to average performance). When performing exactly on average, flagging will occur as frequent as depicted on the vertical axis

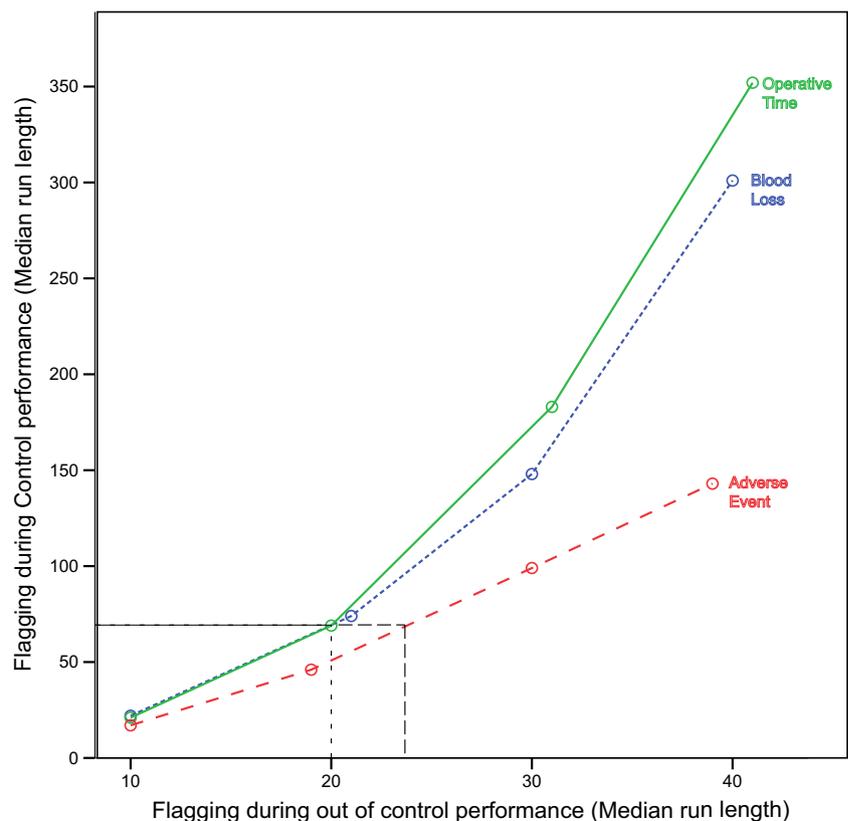


Table 2 Check list after signaling of CUSUM graph

Factor	Example
Patient	Unexpected co-morbidity
Surgeon	Fatigue, stress, and inaccurate indication
Team	Communication and staff's experience
Equipment	Altered vision and new coagulation device
Logistic	Tight scheduled operation programs

Discussion

With proposed validated and risk-adjusted CUSUM graphs, gynecologists have the ability to continuously monitor their surgical performance in laparoscopic hysterectomy, consequently identifying suboptimal factors with respect to operative time, blood loss, and adverse events. As a result, they are able to enhance patient safety.

Despite correction for patient case mix (i.e., identified risk factors), this analysis model still inevitably yields flagging of surgeons with average clinical performance. This is due to the sensitivity of the model. If the CUSUM analysis has to identify derailing performance (OR 2 compared to average performance) in surgeons within a reasonable number of procedures (i.e., 20 laparoscopic hysterectomies), occasional flagging of surgeons with average clinical performance is inevitable.

These proposed cutoff limits are set primarily to identify possible suboptimal situations and to enhance patient safety. The goal is twofold. Firstly, by alarming out of control limits in a timely fashion, the surgeon can evaluate his/her performance as well as of its surgical team and even its equipment and act if necessary. Secondly, by providing (national) averages as a standard of care, hypothetically at long-term, also suboptimal performing surgeons that do not cross the out-of-control line will improve their outcomes.

Although this proposed CUSUM system for laparoscopic hysterectomy is based on national averages of Dutch cohort in 2009, we suggest that the reference values are applicable to every gynecologist. The proposed cutoff values might appear “mild.” However, if these values are raised, as a consequence, signaling will be delayed. This will result in less adequate flagging of potentially derailing performance.

If implemented in a straightforward digital registry tool (or stand alone computer program), this CUSUM for LH provides easy to understand and swift to apply insight into tailor-made proficiency curves. We suggest that out-of-control signaling should primarily be discussed internally and only after a certain acclimatizing period should be discussed with expert peers in order to identify suboptimal care and to provide “Best Practices.”

A number of aspects of the proposed model should be addressed. Firstly, is the average signaling rate of one in 75 procedures in surgeons with average clinical performance acceptable? Yes, however, proper information and efficient evaluation are a prerequisite. Time-consuming evaluation will harm initial motivation. When a CUSUM chart goes out of control, one should be provided with a concise check box-based questionnaire in order to signal the origin of derailing performance (Table 2). This could be due to skills, technical issues, misjudging of a series of cases, problems with the OR team, etc. These issues should be directed. Secondly, ideally, the CUSUM chart (and preferably also its evaluation system) should be integrated and implemented in an already existing electronic patient file system. Registration of patient data in multiple sources will affect quality and quantity of data. Thirdly, the national averages set in this tool should be updated on a frequent basis, preferably every 5 years. Hypothetically, the cohort will improve its surgical outcomes over time. As a result, averages and out-of-control limits should be finetuned as well.

An example is found in the field of (surgical) oncology in which the value of continuous quality assurance is well studied [21–23]. However, these examples use evaluation of care on a yearly basis and often lack correction for patient case mix. Furthermore, most of these registries use adverse events as sole primary outcome and direct hospitals rather than surgeons personally. Some registries reflect hospital outcomes to national averages; however, most systems compare to

(outdated) literature. CUSUM analysis addresses all abovementioned points of interest.

For a start, the CUSUM should be applied and compared indoors only. By means of a multicenter prospective cohort study, the proposed cutoff values are validated as well as the feasibility of this system should be researched. More information as well as the web-based CUSUM tool can be found on www.qusum.org. In conclusion, applying CUSUM charts as quality assurance for the surgical performance and clinical outcome measures in LH might enhance patient safety.

Conflict of interest Andries Twijnstra, Mathijs Blikkendaal, Sara Driessen, Erik van Zwet, Cor de Kroon, and Frank Willem Jansen declare that they have no conflict of interest.

Informed consent All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. The Ethical Committee decided that no informed consent was mandatory in this observational study.

References

- Hopper AN, Jamison MH, Lewis WG (2007) Learning curves in surgical practice. *Postgrad Med J* 83(986):777–779
- Steiner SH, Cook RJ, Farewell VT, Treasure T (2000) Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 1(4):441–452
- Biau DJ, Resche-Rigon M, Godiris-Petit G, Nizard RS, Porcher R (2007) Quality control of surgical and interventional procedures: a review of the CUSUM. *Qual Saf Health Care* 16(3):203–207
- Twijnstra AR, Blikkendaal MD, van Zwet EW, van Kesteren PJ, de Kroon CD, Jansen FW (2012) Predictors of successful surgical outcome in laparoscopic hysterectomy. *Obstet Gynecol* 119(4):700–708
- Park IJ, Choi GS, Lim KH, Kang BM, Jun SH (2009) Multidimensional analysis of the learning curve for laparoscopic colorectal surgery: lessons from 1,000 cases of laparoscopic colorectal surgery. *Surg Endosc* 23(4):839–846
- Cheng JM, Duan H, Wang JJ, Zhang HT, Liu Y (2007) Clinical analysis of conversion from gynecological laparoscopic surgery to laparotomy. *Zhonghua Fu Chan Ke Za Zhi* 42(3):173–175
- Perino A, Cucinella G, Venezia R, Castelli A, Cittadini E (1999) Total laparoscopic hysterectomy versus total abdominal hysterectomy: an assessment of the learning curve in a prospective randomized study. *Hum Reprod* 14(12):2996–2999
- Wattiez A, Soriano D, Cohen SB, Nervo P, Canis M, Botchorishvili R et al (2002) The learning curve of total laparoscopic hysterectomy: comparative analysis of 1647 cases. *J Am Assoc Gynecol Laparosc* 9(3):339–345
- Leminen A (2000) Comparison between personal learning curves for abdominal and laparoscopic hysterectomy. *Acta Obstet Gynecol Scand* 79(12):1100–1104
- Altgassen C, Michels W, Schneider A (2004) Learning laparoscopic-assisted hysterectomy. *Obstet Gynecol* 104(2):308–313
- de Saintonge DM, Vere DW (1974) Why don't doctors use CUSUMs? *Lancet* 1:120–121

12. Schlachta CM, Mamazza J, Seshadri PA, Cadeddu M, Gregoire R, Poulin EC (2001) Defining a learning curve for laparoscopic colorectal resections. *Dis Colon Rectum* 44(2): 217–222
13. Bolsin S, Colson M (2000) The use of the Cusum technique in the assessment of trainee competence in new procedures. *Int J Qual Health Care* 12(5):433–438
14. Weerasinghe S, Mirghani H, Revel A, bu-Zidan FM (2006) Cumulative sum (CUSUM) analysis in the assessment of trainee competence in fetal biometry measurement. *Ultrasound Obstet Gynecol* 28(2):199–203
15. Boulkedid R, Sibony O, Bossu-Salvador C, Oury JF, Alberti C (2010) Monitoring healthcare quality in an obstetrics and gynaecology department using a CUSUM chart. *BJOG*
16. Lindenburg IT, Wolterbeek R, Oepkes D, Klumper FJ, Vandebussche FP, van Kamp IL (2011) Quality control for intravascular intrauterine transfusion using cumulative sum (CUSUM) analysis for the monitoring of individual performance. *Fetal Diagn Ther* 29(4):307–314
17. Steiner SH, Cook RJ, Farewell VT (2001) Risk-adjusted monitoring of binary surgical outcomes. *Med Dec Making* 21(3):163–169
18. Grigg OA, Farewell VT, Spiegelhalter DJ (2003) Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res* 12(2):147–170
19. Aggarwal R, Moorthy K, Darzi A (2004) Laparoscopic skills training and assessment. *Br J Surg* 91(12):1549–1558
20. Schijven MP, Jakimowicz J (2004) The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. *Surg Endosc* 18(1):121–127
21. Landheer ML, Therasse P, van de Velde CJ (2002) The importance of quality assurance in surgical oncology. *Eur J Surg Oncol* 28(6):571–602
22. Peeters KC, van de Velde CJ (2003) Surgical quality assurance in breast, gastric and rectal cancer. *J Surg Oncol* 84(3):107–112
23. Verleye L, Vergote I, Reed N, Ottevanger PB (2009) Quality assurance for radical hysterectomy for cervical cancer: the view of the European Organization for Research and Treatment of Cancer–Gynecological Cancer Group (EORTC-GCG). *Ann Oncol* 20(10): 1631–1638