Benchmarking of surgical complications in gynaecological oncology: prospective multicentre study

Matthew Burnell[1], Rema Iyer[1], Aleksandra Gentry-Maharaj*[1], Andy Nordin[2], Robert Liston[1], Ranjit Manchanda[1,12], Nagindra Das[3], Rob Gornall[4], Alice Beardmore-Gray[1], Kathryn Hillaby[4], Simon Leeson[5], Anders Linder[6], Alberto Lopes [3], David Meechan[7], Tim Mould[8], James Nevin[9], Adeola Olaitan[8], Barnaby Rufford [6], Smruta Shanbhag[10], Alexandra Thackeray[7], Nick Wood[11], Karina Reynolds[12], Andy Ryan[1], Usha Menon[1]

*joint

Affiliation

[1]Gynaecological Cancer Research Centre, Dept of Women's Cancer, Institute for Women's Health, University College London, Maple House, 149 Tottenham Court Road, London W1T 7DN

[2]East Kent Gynaecological Oncology Centre, Queen Elizabeth the Queen Mother Hospital, St Peter's Road, Margate, Kent CT9 4AN

[3]Dept of Gynaecological Cancer, Royal Cornwall Hospitals NHS Trust, Truro, Cornwall TR1 3LJ

[4]Dept of Gynaecological Oncology, Cheltenham General Hospital, Sandford Road, Cheltenham, Gloucestershire, GL53 7AN

[5]Dept of Obstetrics and Gynaecology, BetsiCadwaladr University Health Board, Penrhosgarnedd, Bangor, Gwynedd, North Wales, LL57 2PW

[6]Dept of Gynaecological Oncology, The Ipswich Hospital NHS Trust, Heath Road, Ipswich, Suffolk, IP4 5PD

[7]Public Health England, 5 Old Fulwood Road, Sheffield S10 3TG

[8]Dept of Gynaecological Oncology, University College London Hospital NHS Foundation Trust, 2nd Floor North, 250 Euston Road, London NW1 2PG

[9]Pan Birmingham Gynaecological Cancer Centre, Birmingham City Hospital, Dudley Road, Birmingham, West Midlands, B18 7QH

[10]Dept of Gynaecological Oncology, Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER

[11]Dept of Gynaecological Oncology, Lancashire Teaching Hospitals NHS Foundation trust, Royal Preston Hospital, Sharoe Green Lane, North Fulwood, Preston Lancashire PR2 9HT

[12]Dept of Gynaecological Cancer, Barts Cancer Centre, Barts and the London NHS Trust, West Smithfield, London EC1A 7BE

*Corresponding author: Usha Menon, Head Gynaecological Cancer Research Centre, Department of Women's Cancer, 1st Floor Maple House, 149 Tottenham Court Road, University College London, London W1T 7DN, United Kingdom

Email address: u.menon@ucl.ac.uk

Telephone: +44 203 447 2108/2112

Fax: +44 203 447 2129

Abstract

Objective: To explore the impact of risk-adjustment on surgical complication rates (CRs) for benchmarking gynaecological-oncology centres.

Design: Prospective cohort study

Setting: Ten UK accredited gynaecological-oncology centres

Population: Women undergoing major surgery on a gynaecological-oncology operating list

Methods: Patient co-morbidity, surgical procedures and intraoperative (IntraOp) complications were recorded contemporaneously by surgeons for 2948 major surgical procedures. Postoperative (PostOp) complications were collected from hospitals and patients. Risk-prediction models for IntraOp and PostOp complications were created using penalised (lasso) logistic regression using over 30 potential patient/surgical risk-factors.

Main outcome measures: Observed and risk-adjusted IntraOp and PostOp CRs for individual hospitals were calculated. Benchmarking using colour-coded funnel plots and observed-to-expected ratios was undertaken.

Results: Overall, IntraOp CR was 4.7%(95%CI 4.0-5.6) and PostOp CR 25.7%(95%CI 23.7-28.2). The observed CRs for all hospitals were under the upper 95% control limit for both IntraOp and PostOp funnel plots. Risk-adjustment and use of observed-to-expected ratio resulted in one hospital moving to the >95-98%CI(red) band for IntraOp CRs. Use of only hospital-reported data for PostOp CRs would have resulted in one hospital being unfairly allocated to the red band. There was little concordance between IntraOp and PostOp CRs.

Conclusions: The funnel plots and overall IntraOp(≈5%) and PostOp(≈26%) CRs could be used for benchmarking gynaecological-oncology centres. Hospital benchmarking using risk-adjusted CRs allows fairer institutional comparison. IntraOp and PostOp CRs are best assessed separately. As hospital under-reporting is common for post-operative complications, use of patient-reported outcomes is important.

Keywords: UKGOSOC, complications, gynaecological oncology, risk adjustment, centres, benchmarking, comparison, surgery

Tweetable Abstract: Risk-adjusted benchmarking of surgical complications for ten UK gynaecological oncology centres allows fairer comparison

Introduction

There is a drive within the National Health Service (NHS) to increase transparency and improve quality and safety. To this end, one of the initiatives in surgery has been to publish outcomes data for hospitals and more recently for individual surgeons which have been sourced from national clinical audits in some specialties and in most from administrative data. [1, 2]

While surgical data on a national level has been collected in specialities such as cardiothoracic [3] and orthopaedic [4] surgery and certain cancers such as lung [5], colorectal [6] and head and neck [7], there is paucity of such data in gynaecological oncology (GO). To address this, the United Kingdom Gynaecological Oncology Surgical Outcomes and Complications (UKGOSOC) [8] study was undertaken to prospectively capture data on surgery with a view to setting benchmarking standards. In this cohort, the overall intraoperative (IntraOp) complication rate (CR) was 4.7% [8] and the postoperative (PostOp) CR was 25.7%. [9] However use of observed complication rates (CRs) for centre level comparisons does not take into account patient comorbidity, underlying disease or surgical complexity, all of which can impact on the risk of a complication. [8] The use of unadjusted crude CRs has resulted in significant unease amongst surgeons and hospitals due to the variations in prevalence of surgical risk factors. Concerns have been raised that it might deter surgery being undertaken in 'high-risk' patients with significant comorbidity. We report on the impact of risk-adjustment of surgical CRs on benchmarking of GO at hospitals participating in UKGOSOC.

Methods

Study Design

The UKGOSOC study design has been previously described. [8, 9] In brief, ten UK GO centres collected data using web-based software on consented women undergoing major gynaecological cancer surgery. Surgeons entered patient co-morbidity, surgical procedures and IntraOp complications (Table S1) contemporaneously in theatre. PostOp complications (Table S2) were defined as occurring up to eight weeks after surgery. These were entered on to the online database during the admission by the hospital team. Following discharge from hospital, patient-reported complications data was also collected using postal follow-up. [9]

Statistical Methods

All reported major surgery was used to calculate IntraOp CR. All reported PostOp complications were graded (I-V) according to severity using the Clavien and Dindo system.[10] Grade 1 complications were excluded from analysis. A PostOp complication for this analysis was defined as one of Grade II-V severity. Only those surgeries where both hospital and patient follow-up data were available, were included in the primary analysis, with PostOp complications including those reported by the clinical team (hospital-reported), the patient (patient-reported) or both. A secondary analysis based on the whole dataset and only hospital-reported PostOp complications, was also undertaken. The outcome was treated as strictly binary and a surgery was coded as having a complication irrespective of the number of complications.[8] The recorded risk factors for IntraOp and PostOp complications have been described previously [8] and are listed in Tables 1 and 3. Separate comparisons were undertaken for IntraOp and PostOp CRs. All methods described below apply to both.

Data description

Outcome and categorical risk factors were cross tabulated by hospitals. To assist identification of imbalance in risk factors across hospitals, chi-squared tests were performed. Associated p-values were not used as a formal test measure for the factors with small category counts (<5) at any hospital. Continuous risk factors were summarised by within-hospital means and standard deviations, and F test statistics and p-values from an analysis of variance were similarly used to aid assessment of hospital variation.

Risk prediction and penalised regression

Logistic regression models were used for risk prediction, though parameter estimates were based on a penalised method (least absolute shrinkage and selection operator or 'lasso') [(11)] rather than maximum likelihood (ML). A fundamental issue involved in prognostic model construction is that of 'events per variable' (EPV) [(12)], where the number of 'events' in a binary regression model is taken as the total of the less common outcome. A fitted model should have an EPV of at least 10 [(13, 14)], where the variable count includes all levels of a categorical variable. The EPV requirement holds even if variable selection (stepwise methods) is performed, so that the variable count is based on the full model. A limited number of EPV can cause validation problems when using ML for parameter estimates, as the model becomes over-fitted and prediction error is inflated. As a result many prediction models fail to be successfully validated (12).

The lasso deliberately biases (shrinks) the regression estimates toward zero, reducing the mean square prediction error (MSPE) which is a function of the variability, as well as the bias, of the predictions. As a result, despite intentional bias, penalised methods typically provide better prediction than ML. A brief description of the lasso

method is presented in Appendix 1. Formal inference for biased estimates is dubious so p-values should be used only for approximate guidance. [11] The user-written Stata commands *plogit* and *plsearch* were used to fit lasso-shrunk logistic models. Equivalent models fitted by ML are presented for comparison.

Model fit was assessed using the Hosmer-Lemeshow test and model specification with the link test. A receiver operating characteristic (ROC) curve using the predicted probabilities generated by leave-one-out cross-validation (LOO-CV) was calculated and overall performance (discrimination) assessed by the area under the curve (AUC). By regressing the outcome on bootstrapped linear predictions (log odds) for each subject, the calibration slope (CS) [12, 13] could be estimated as the mean slope of 1000 bootstrap samples, where a slope close to one suggests good calibration and (much) less than one implies over-fitting of the model. An over-fitted model will give predictions that are too narrow in range.

Hospital rate adjustments

Observed IntraOp and PostOp CRs for hospitals were compared using funnel plots, a standard approach to institutional comparison. [15, 16] They were generated by plotting each hospital's observed CR against sample size and assessed with respect to confidence bands that signify unusually high or low CRs. Control limits for the funnel plots were generated using smoothed exact confidence intervals for the overall CR. These were displayed using coloured 'warning' bands of increasing 'concern' with regard to increasing CRs : green (up to 80%), yellow (≥80%-90%), orange (≥90%-95%), and red (≥95-98%)..

We used the prediction model to produce expected IntraOp and PostOp CRs for each hospital and hence an observed-to-expected CR ratio. Bootstrapped confidence

intervals for the CR ratio were generated (see Appendix 2) with the same warning levels as for the funnel plots, and if the confidence band contained 1, the hospital was denoted with the appropriate coloured warning.

Results

Intra-operative complications

2948 surgeries undertaken across 10 hospitals were included in the analysis. 139 had at least one IntraOp complication. Although the observed IntraOp CR ranged from 2.0% to 8.0%, the variation was not significant between hospitals (p=0.052).

Modelling and fit

The distribution of risk factors and CRs across the 10 hospitals is detailed in Table S3. There was variation across hospitals for most predictors, but particularly for laparoscopic approach, grade of surgeon, surgical complexity, final diagnosis and American Society for Anaesthesiologists (ASA) grade.

Of the 2948 surgeries, missing data meant that only 2709 surgeries with 132 complications could be included when fitting the full model. This resulted in an EPV of 4.1 given the 32 variables. The optimised lasso model resulted in four of the 32 variables being excluded. (Table1) Lasso-shrunk odds ratios are presented in Table 1, alongside the ML estimates for comparison. The strongest predictors ($p < 0.05$ when estimated by ML) of IntraOp complication were surgical complexity (increased risk with increasing complexity), previous abdominal surgery (increased risk), diabetes (increased risk), metabolic-endocrine disorders other than diabetes (decreased risk) and final diagnosis (all cancer types associated with reduced risk relative to ovarian cancer).

The ROC curve based on LOO-CV produced an AUC_{lasso}=0.663 (95%CI: 0.616-0.710), similar to AUC_{ML}=0.659 (95% CI: 0.611-0.706), although ROC curves are

affected only by rank order and not magnitude. The mean CS_{lasso} of 0.871 suggested a slight narrowness of predictions, although the 2.5th-97.5th centile of the slopes (0.717-1.068) contain the optimal value of 1. In contrast, the CS_{ML}=0.712 (95%CI: 0.364-0.887) indicated that the prediction range was limited, and hence the model over-fitted, without parameter shrinkage.

Hospital rate adjustments

Figure 1a shows the funnel plot allowing a simple comparison of observed IntraOp CR by hospital. A majority of the hospitals are within the green band with some, such as Hospital F, having a significantly low IntraOp CR outside the 95% (≈2 standard deviations) control limits. Hospitals J and E have CR higher than the overall IntraOp CR but the moderate number of surgeries analysed from these hospitals (150 and 181, respectively) means a reduced confidence in their outlier status and they lie within the yellow band (control limits ≥80%-90%). Figure 1b shows the observed-to-expected CR ratio, adjusted for risk factor prevalence in the 10 hospitals. The spread of expected IntraOp CR for hospitals is between 3.9% and 5.4% (Table 2a). In Figure 1b Hospital F is confirmed as having a low IntraOp CR after adjustment. The 95% confidence interval for Hospital E is completely above 1, the line of equality (observed=expected), and it is coded red (confidence interval >95-98%) marking it as a centre with high IntraOp CR. The ratio for Hospital G is also high at 1.8, but with wide confidence intervals, hence it remains coded green. Hospital J, which had the highest observed IntraOp CR, only has the 3rd highest ratio (Table 2a) and remains coded yellow, indicating that its high CR is partially mitigated by a relatively high-risk case-mix of surgeries.

Post-operative complications

The primary PostOp CR analysis was restricted to the subset of 1462 surgeries with both hospital and patient reported outcomes. 376 had at least one PostOp complication. Individual hospital statistics are presented in Table S4. Estimated blood loss (EBL) and duration of surgery varied significantly by hospital. The PostOp CR varied from 15.6% to 36.2% between hospitals but the difference was not significant (p=0.096). The findings were similar for the full dataset (Table S3).

Modelling and fit

Of the 1462 surgeries missing data meant only 1371 surgeries with 346 events were included when fitting the full model. This resulted in an EPV of 9.9 given the 35 variables. The optimised lasso model resulted in 15 variables out of 35 being removed from the model (Table 3). Only duration of surgery appeared to be a strong predictor of PostOp complications, though coagulation-thrombosis, and musculoskeletal disorders and diabetes (all increase risk), laparoscopic approach (decreases risk) and final diagnosis (cervical and vulval cancer increase risk relative to ovarian cancer) were significant at the 5% level using ML. The ROC curve based on LOO-CV produced an AUC_{lasso}=0.659 (95%CI:0.585-0.733), significantly larger than AUC_{ML}=0.569 (95%CI:0.487-0.652) as tested using the method described by DeLong et al (17)(p=0.0003). The mean CS_{lasso}=1.008 (95%CI:0.799-1.264) suggested near-perfectly calibrated predictions. By comparison, a mean CS_{ML}=0.689 (95%CI: 0.562-0.835) strongly indicated that the ML-based prediction range was too narrow, and hence the model was over-fitted.

Hospital rate adjustments

Figure 1c compares the observed PostOp CRs of the 10 hospitals using a funnel plot. The overall CR was 25.7%. None of the hospitals had a PostOp CR that was significantly higher than the overall CR. Hospital J had the highest PostOp CR (36.2%) but as this was based on only 58 surgeries, it lies within the yellow (control limits ≥80%-90%) band. Hospitals G (15.6%) and I (17.1%) have notably low CR. Figure 1d shows the observed-to-expected PostOp CR ratio, with actual values found in Table 2b. The spread of expected PostOp CR for hospitals is between 20.1% and 28.5%. None of the hospitals have a CR ratio significantly different above 1, the line of equality (observed=expected) at the 5% level, though Hospital J remains within the yellow band.

All PostOp CRs are lower if hospital reported statistics alone are used. (Figure S1-faded colour-coding) Hospital D is in the red band when hospital reported rates alone are used. However, when surgeries with both hospital and patient reported statistics are used, hospital D is in the green band suggesting that differences in under-reporting can significantly impact on hospital comparisons. (Figure S1 - bright colour-coding)

Generally, there was little concordance between IntraOp and PostOp CR. Hospitals G and E which were among the centres with high IntraOp CR, had some of the lowest PostOp CRs, both observed and risk-adjusted. (Figure 1)

Discussion

Main Findings

Risk-adjustment did not make a significant difference to the CRs for majority of centres, but helped to delineate the outliers better. The shaded funnel plots and observed versus expected ratios generated made comparisons easy to comprehend.

Hospital under-reporting is common for PostOp complications and inclusion of patient-reported outcomes is important to ensure valid comparison between institutions. Risk factors for IntraOp and PostOp CRs were largely different and even after adjustment there was no concordance between hospital IntraOp and PostOp CRs. The overall IntraOp (≈5%) and PostOp (≈26%) CR derived from this study could be used to benchmark performance in GO.

Strength and Limitations

This is the first large prospective multicentre study in GO to develop risk-adjusted CRs for inter-institutional comparison of surgical outcomes. Although such data is available in other specialties, [3, 5, 7, 18] in GO, it is limited to a retrospective study comparing outcomes of ovarian cancer surgery between three US tertiary centres.[19]

Whilst the limited number of surgeries entered is not entirely representative of all GO operations performed in the UK, this was a huge undertaking for the clinicians involved. For the 7 centres we have data for, the inclusion rate of cases ranged from 64.6% to 97.6% and all cases were prospectively registered prior to surgery.

We have previously described the risk predictors for IntraOp and PostOp CRs based on univariable and multivariable regression.[8] Few of the factors appeared important across either model reflecting the difficulty in developing risk-prediction tools. While significant effort was required for complete prospective data collection on 2948 surgeries, given the high EPV rate had significant implications for estimation.

Use of a data-dependent internal measure (observed overall CR) in lieu of a pre-specified target rate based on external data and expert opinion was a limitation, and

a hospital with a particularly high rate will help push up that value to which all hospitals are compared to. Unfortunately there was insufficient published data on GO CRs to utilise a prior target rate. Given that there are no apparent institutional outliers in our dataset, the observed CRs might be reasonably used as future target rates. A related issue is that the data used to estimate the prediction model was the same to which the model was then applied, though cross-validation methods were employed. An external dataset is therefore required for proper model validation.

Interpretation

Figure 1 suggests the adjustment process may appear to add little value when comparing centres, given that most of the predictors appeared to have limited impact on outcome. However adjustment helps to better define the level of excess surgical complications at a given hospital, and could therefore provide an earlier intimation of potential issues. Hospital E was only flagged as having a statistically high IntraOp CR ($p=0.03$) following adjustment. The IntraOp and PostOp CRs did not vary significantly between the 10 hospitals, so that for the majority, the observed CRs were within the funnel plot control limits. In the broader healthcare community, where the spread of quality and CRs is likely to be wider, it is likely that there will be institutions beyond the various safety bounds, requiring more precise performance monitoring.

By contrast, nearly all the predictors varied considerably by hospital, especially those involving an element of surgical decision making (laparoscopic approach, surgeon grade, surgical complexity). This ‘inter-hospital’ variability in risk factor prevalence is a strong argument in itself for the need to adjust for fairer comparison. That risk-adjustment did not substantively affect results is partially due to parameter shrinkage caused by lack of statistical power. Based on our results, a subsequent study

modelling IntraOp CRs, would need n≈12000 for hypertension (OR_{ML}=1.28; p=0.26) to achieve power=80%, for example. With more data some of these factors may contribute significantly to CR prediction, both statistically and clinically. The lack of association between CR and factors like BMI, especially after open surgery, are contrary to previous reports. [20, 21] However, it is evident that much of the outcome variability is related to unmeasured (perhaps unobservable) phenomena, and we do not expect a surgical complication to be ever predicted with a high degree of confidence.

The difference in ranking order of hospitals for IntraOp and PostOp CRs, for example, hospitals G and E, could be due to various reasons including surgical skill, post-operative care and under-reporting of PostOp complications. Also, since IntraOp and PostOp CRs had differing risk factors, we recommend assessing hospitals separately for IntraOp and PostOp complications as combining them could mask deficiencies in perioperative care at certain hospitals.

Analysis of only hospital-reported PostOp complications demonstrated that hospital D had the highest CR outside the control limits (Figure S1). However, with inclusion of patient-reported data hospital D was no longer an outlier, but hospital J's CR increased from close to the 50th to the 90th centile, suggesting hospital D had been more diligent at recording all PostOp complications. These findings further substantiate the need for including patient-reported PostOp complications to overcome clinician under-reporting.

Penalized models may appear complicated but in a limited event situation it is known that selection methods may moderate predictors and include noise predictors. [22] However, it is straightforward to input predictor values into, say, an Excel sheet pre-prepared with the necessary formula to calculate risk scores, and calculate confidence

limits treating the expected rate as fixed [23] (Table S5). Alternatively, the model parameters presented could be used as informative priors for subsequent model building by other researchers in a Bayesian context.

Since morbidity is the main yardstick for benchmarking surgical performance, moving forwards, it would be important to have complete and accurate data in a larger database. The drawback of clinician-led databases is that all surgical episodes may not get recorded due to its heavy reliance on voluntary data entry. [24] Also, some centres with high morbidity rates may be hesitant to enter all their data.[25] The alternative would be an administrative database like Hospital Episode Statistics (HES) where all surgical episodes are automatically recorded, which, might however have incomplete complications data. Although Nouraei et al [7] found CRs derived from HES comparable to the clinician-led head and neck surgery database, this has not been the case with other specialties.[26] A re-audit in one of the participating centres demonstrated higher CRs than that derived from HES, but comparable to UKGOSOC. [27] Therefore, a reasonable compromise may be to combine HES with a clinician-led database. To ensure completeness it would be prudent to minimise the data fields requiring clinician entry. It is hoped that mandatory requirement to publish individual surgeon's outcomes will encourage complete and accurate data entry.[1] Valid case-mix risk-adjustment is also likely to reassure surgeons and encourage active participation in surgical outcomes assessment audits.

Conclusion

Risk-adjustment had a modest effect on the rankings of the individual centres based on their CRs. However, by accounting for the prevalence of potential risk factors we may be able to estimate an adjusted CR that ensures fairer inter-institutional

comparison. The overall IntraOp (≈5%) and PostOp (≈26%) CRs and funnel plots could be used to benchmark performance of GO centres and even individual surgeons with a larger dataset. The risk factors for IntraOp and PostOp complications are different and it is important to report on the two CRs separately.

Disclosure of interests

The authors have no conflicts of interest to declare.

Contribution to authorship

UM was the principal investigator of the study. UM, AN, AGM, were responsible for conceptualisation and design of the study. RL, AGM, RI, AN and UM designed the database. RM, ND, RG, ABG, KH, SL, An L, Al L, TM, JN, AO, BR, SS, NW, KR were involved with recruitment and data entry from individual centres. AT and DM helped with hosting the database on the cancer-registry server, issuing passwords for all users and sending patient follow-up letters. RI was responsible for cleaning and preparing the data for the analysis, analysing patient follow-up questionnaires and writing of the manuscript with MB. MB performed the statistical analysis and drafted the manuscript. AR helped with cleaning of the data. All the authors provided regular feedback on the manuscript.

Acknowledgements

This study was supported by researchers at the National Institute for Health Research, University College London Hospitals Biomedical Research Centre. The authors are thankful to the Eve appeal for their support of the study. The authors are also very grateful to all the women who participated and to all the medical, nursing and

administrative staff who worked on UKGOSOC. In addition the authors would like to thank the following members of the hospital teams who contributed to data collection: Sara Roberts, Philip Toon, Richard Peevor (Betsi Cadwaladr University Health Board), Charlie Chan, Janos Balega, Kavita Singh, Sudha Sundar, Ahmed Elattar, Esther Moss, Mary Wright (City Hospital Birmingham), Alta Viljoen (Cheltenham General Hospital), Branislav Potancok, Mohamed Ismail, Vivek Nama, Ruth Lomas, Cheryl Walke (East Kent University Hospital NHS Foundation Trust), Deborah Woods, Alison Garnham (Ipswich General Hospital), KetanGajjar, Deborah Parkinson (Royal Preston Hospital), Emma Arthur, Arnold Kruse (Royal Cornwall Hospital), Kostas Doufekas (University College London Hospital). The study was supported by researchers at the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

Ethics Approval

The joint UCL and UCLH committees decided that this study did not require ethics approval. However consent was taken from the women who participated in the study to include their identifiers so that they could be sent a follow-up letter following their discharge from hospital.

Funding

We are grateful to the Eve Appeal for the funding of this study.

References

1. http://www.nhs.uk/service-search/performance/Consultants#view-the-data.

2. http://www.hqip.org.uk/consultant-outcomes-publication/.

3. Bridgewater B. Cardiac registers: the adult cardiac surgery register. Heart. 2010 September 1, 2010;96(18):1441-3.

4. Hunt LP, Ben-Shlomo Y, Clark EM, Dieppe P, Judge A, MacGregor AJ, et al. 90-day mortality after 409 096 total hip replacements for osteoarthritis, from the National Joint Registry for England and Wales: a retrospective analysis. The Lancet. //28;382(9898):1097-104.

5. Beckett Pcp, Woolhouse Icp, Stanley R, Peake MDcp. Exploring variations in lung cancer care across the UK - the 'story so far' for the National Lung Cancer Audit. Clinical medicine. 2012;12(1):14-8.

6. The National Bowel Cancer Audit Report 2010 http://www.ic.nhs.uk/webfiles/Services/NCASP/audits%20and%20reports/NHS_Bowel_Cancer_2010_INTERACTIVE.pdf. 2010.

7. Nouraei SAR, Middleton SE, Hudovsky A, Darzi A, Stewart S, Kaddour H, et al. A national analysis of the outcome of major head and neck cancer surgery: implications for surgeon-level data publication. Clinical otolaryngology & allied sciences. 2013;38(6):502-11.

8. Iyer R, Gentry-Maharaj A, Nordin A, Burnell M, Liston R, Manchanda R, et al. Predictors of complications in gynaecological oncological surgery: a prospective multicentre study (UKGOSOC[mdash]UK gynaecological oncology surgical outcomes and complications). Br J Cancer. 2014 12/23/online.

9. Iyer R, Gentry Maharaj A, Nordin A, Liston R, Burnell M, Das N, et al. Patient-reporting improves estimates of postoperative complication rates: A prospective cohort study in gynaecological oncology. British Journal of Cancer. 2013;109(3):623.

10. Dindo D, Demartines N, Clavien PA. Classification of surgical complications - A new proposal with evaluation in a cohort of 6336 patients and results of a survey. Annals of surgery. 2004;240(2):205-13.

11. Tibshirani R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society Series B (Methodological). 1996;58(1):267-88.

12. Ambler G, Seaman S, Omar RZ. An evaluation of penalised survival methods for developing prognostic models with rare events. Statistics in medicine. 2012 May 20;31(11-12):1150-61. PubMed PMID: 21997569.

13. Ambler G, Brady AR, Royston P. Simplifying a prognostic model: a simulation study based on clinical data. Statistics in medicine. 2002 Dec 30;21(24):3803-22. PubMed PMID: 12483768.

14. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart. 2012 May;98(9):683-90. PubMed PMID: WOS:000303859300004. English.

15. Spiegelhalter D. Funnel plots for institutional comparison. Quality & safety in health care. 2002 Dec;11(4):390-1. PubMed PMID: 12468705. Pubmed Central PMCID: 1757996.

16. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Statistics in medicine. 2005 Apr 30;24(8):1185-202. PubMed PMID: 15568194.

17. Delong ER, Delong DM, Clarkepearson DI. Comparing the Areas under 2 or More Correlated Receiver Operating Characteristic Curves - a Nonparametric Approach. Biometrics. 1988 Sep;44(3):837-45. PubMed PMID: WOS:A1988Q069100016. English.

18. National Bowel Cancer Audit Programme : Healthcare Quality Improvement Partnership (HQIP). http://www.hqip.org.uk/national-bowel-cancer-audit-programme-nbocap.Reports of the National Bowel Cancer Audit Project 2002–2006. .

19. Aletti GD, Santillan A, Eisenhauer EL, Hu J, Aletti G, Podratz KC, et al. A new frontier for quality of care in gynecologic oncology surgery: Multi-institutional assessment of short-term outcomes for ovarian cancer using a risk-adjusted model. Gynecologic Oncology. 2007;107(1):99-106.

20. Kumar A, Bakkum-Gamez JN, Weaver AL, McGree ME, Cliby WA. Impact of obesity on surgical and oncologic outcomes in ovarian cancer. Gynecologic Oncology. 2014 10//;135(1):19-24.

21. Osler M, Daugbjerg S, Frederiksen B, Ottesen B. Body mass and risk of complications after hysterectomy on benign indications. Human reproduction.26(6):1512-8.

22. Royston P, Sauerbrei W. Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables. Wiley: Chichester, UK. 2008.

23. DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. Comparing risk-adjustment methods for provider profiling. Statistics in Medicine. 1997;16(23):2645-64.

24. RCOG. Hospital Episode Statistics as a source of information on safety and quality in gynaecology to support revalidation. 2012.

25. Almoudaris AM, Burns EM, Bottle A, Aylin P, Darzi A, Faiz O. A colorectal perspective on voluntary submission of outcome data to clinical registries. British Journal of Surgery. 2011;98(1):132-9.

26. Cockbain AJ, Carolan M, Berridge D, Toogood GJ. Performance and Quality Indicators: The importance of accurate coding. Bulletin of the Royal College of Surgeons of England. 2012;94(2):46-50.

27. Rajkumar S, Lee MJ, Nordin A. Perioperative morbidity and mortality audit strategies in gynaecological oncology surgery: The east Kent gynaecological oncology centre experience. International journal of gynecological cancer. 2014;24(9 SUPPL. 4):1160-1.

Appendix S1: The Lasso method

The lasso (least absolute shrinkage and selection operator) estimator [1] employs a penalty term in the likelihood function that is then maximised subject to a constraint on the (absolute) sum of the regression coefficients. The penalty term is a function of shrinkage parameter (λ) chosen by the investigator, which when equal to zero reduces to ML estimation and when tending to infinity results in estimates tending to zero. In contrast to the similar ridge regression method, where all the coefficients of the full model are partially shrunk, the lasso actually performs a type of variable selection. Strong and moderate predictors are shrunk by a certain amount dependent on λ, whilst weak predictors may be shrunk to exactly zero and so drop out of the model. The choice of λ here was based on a grid search that minimised the generalized cross-validation error [1]. Note that inference, such as confidence intervals and p-values, based on standard errors from the lasso variance-covariance matrix should be treated with caution and used only for approximate guidance. Standard errors are not particularly meaningful for (deliberately and quite strongly) biased estimates as they will exclude the inaccuracy caused by bias [1].

In contrast to penalized regression methods, model selection procedures are known to result in selection- or omission-bias [2], whereby weakly significant variables will be infrequently selected, dependent on chance variation, and when selected, they will typically have overestimated coefficients.

Appendix S2: The observed-expected complication rate and confidence intervals

We used the risk prediction model to produce expected IntraOp and PostOp CRs for each hospital by summing the predicted risk for each surgery over each hospital. We then compared the expected with the observed CR, and if the confidence band for the CR ratio (defined as for the funnel plots: green (up to 80%), yellow (≥80%-90%), orange (≥90%-95%), and red (≥95-98%)) contained 1, we denoted the hospital with the appropriate coloured warning. To incorporate the uncertainty involved in estimating the expected CR, as well as the choice of λ, the sampling distribution of the observed-to-expected CR ratio was estimated by taking 1000 bootstrap samples of the full dataset. For each bootstrap sample the new observed-to-expected CR ratio was calculated for each hospital, based on a refitting of the lasso model with λ optimally selected. Confidence intervals for the ratio were based on the appropriate bias-corrected centiles of the bootstrap-derived sampling distribution.

References:

1. Tibshirani R. Regression shrinkage and selection via the Lasso. J Roy Stat Soc B Met 1996;**58**(1):267-88
2. Royston P, Sauerbrei W. Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables. Wiley: Chichester, UK 2008

Figure 1: Comparison of intraoperative (a and b) and patient reported postoperative (c and d) complication rates. Left hand panels plot observed rates over colour-coded funnel plots where the yellow band/line represents a 80-90% control limit, the orange band/line a 90-95% control limit, and the red band/line a 95-98% control limit around overall rate. Right hand panels plot ratio of observed to expected rate against the null value of one, with colour-coded confidence intervals representing the same interval range as for the funnel plots. All plotted rates are also colour-coded, reflecting the position of the observed rate (left panels) or null value (right panels) within the appropriate coloured band.

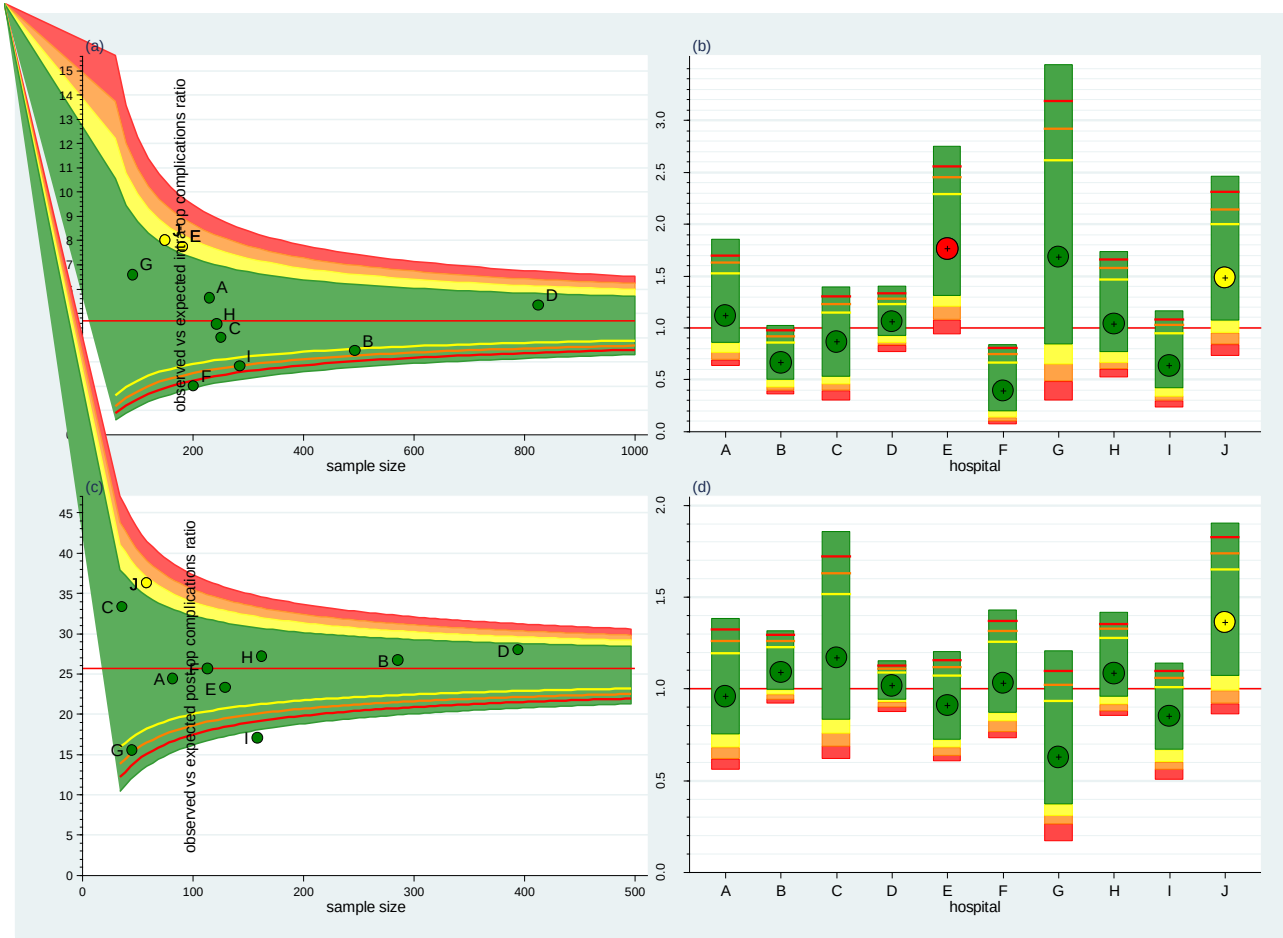

Table 1: Risk prediction model for intra-operative complications - full sample of hospital reported outcomes n=2948

Variables	Lasso				Maximum Likelihood				Shrinkage
	OR	p-value*	L95% CI*	U95% CI*	OR	p-value	L95% CI	U95% CI	
Age at surgery	1.000	0.977	0.985	1.015	1.001	0.859	0.986	1.017	-84.0%
BMI	*removed*				1.003	0.867	0.974	1.032	-100.0%
Previous abdominal surgery	1.426	0.058	0.988	2.057	1.459	0.045	1.008	2.111	-6.0%
Low albumin	3.916	0.118	0.709	21.645	4.461	0.080	0.836	23.799	-8.7%
Coagulation-thrombosis	1.052	0.910	0.436	2.540	1.148	0.755	0.483	2.729	-63.1%
Diabetes	1.804	0.032	1.052	3.095	1.923	0.018	1.118	3.306	-9.7%
Cardiac	1.462	0.205	0.812	2.632	1.572	0.128	0.878	2.814	-15.9%
Respiratory	0.676	0.266	0.339	1.348	0.573	0.133	0.277	1.185	-29.8%
Gastrointestinal	1.065	0.893	0.425	2.668	1.188	0.703	0.490	2.879	-63.5%
Genitourinary	0.699	0.679	0.129	3.805	0.486	0.483	0.065	3.651	-50.4%
Musculoskeletal	*removed*				1.091	0.791	0.574	2.071	-100.0%
Neurology-psychiatric	*removed*				1.028	0.940	0.501	2.109	-100.0%
Vascular	0.849	0.775	0.276	2.607	0.675	0.527	0.199	2.286	-58.3%
Auto-immune	1.968	0.253	0.617	6.282	2.132	0.191	0.685	6.642	-10.6%
Metabolic-endocrine	0.412	0.027	0.187	0.906	0.329	0.010	0.141	0.768	-20.3%
Integumentary-dermatology	*removed*				0.964	0.965	0.191	4.873	-100.0%
Hypertension	1.239	0.325	0.808	1.899	1.279	0.263	0.831	1.969	-13.0%
Smoking	0.978	0.969	0.323	2.961	0.828	0.760	0.246	2.788	-88.2%
Other neoplasms	1.506	0.246	0.755	3.004	1.590	0.182	0.805	3.140	-11.8%
Laparoscopic approach	1.021	0.935	0.618	1.689	1.240	0.403	0.749	2.051	-90.2%
ASA									
ASA grade 1	1				1				
ASA grade 2	1.103	0.699	0.670	1.816	1.250	0.401	0.742	2.106	-56.0%
ASA grade 3+	1.039	0.908	0.539	2.003	1.183	0.628	0.599	2.336	-77.0%
Surgeon grade									

Consultant	1				1				
Sub-specialty trainee	0.716	0.604	0.202	2.535	0.614	0.460	0.168	2.243	-31.6%
General Obstetrics & Gynaecology Trainee	1.286	0.673	0.399	4.144	1.243	0.720	0.378	4.083	15.8%
Surgical complexity									
Complexity score 1&2	1				1				
Complexity score 3&4	1.097	0.692	0.695	1.731	1.263	0.325	0.794	2.009	-60.4%
Complexity score 5&6	1.905	0.016	1.130	3.212	2.208	0.003	1.298	3.756	-18.6%
Complexity score 7&8	2.666	0.012	1.242	5.725	3.080	0.004	1.434	6.612	-12.8%
Complexity score >8	4.005	0.003	1.626	9.865	4.561	0.001	1.850	11.242	-8.6%
Final diagnosis									
Ovarian	1				1				
Uterine	0.600	0.050	0.360	1.001	0.504	0.011	0.296	0.856	-25.5%
Cervical	0.834	0.636	0.393	1.769	0.696	0.361	0.320	1.514	-49.9%
Vulval	0.289	0.049	0.084	0.993	0.195	0.026	0.046	0.826	-24.1%
Benign	0.567	0.041	0.329	0.976	0.508	0.017	0.291	0.887	-16.2%
Constant	0.033	0.000	0.007	0.146	0.025	0.000	0.004	0.145	-32.8%

* for approximate guidance only

Table 2: Summary of intraoperative and postoperative complications by centre

a) Intraoperative complications summary - full sample of hospital reported outcomes n=2948

Hospital	Number of surgeries	Number IO[1] complications	Observed IO CR[2]	Expected number IO complications	Expected IO CR	O/E[3] IO CR ratio	Lower 95% CI[4] for O/E ratio	Upper 95% CI for O/E ratio
A	230	13	5.7%	11.7	5.1%	1.116	0.689	1.698
B	493	17	3.4%	25.6	5.2%	0.664	0.390	0.977
C	251	10	4.0%	11.6	4.6%	0.865	0.389	1.307
D	825	44	5.3%	41.7	5.1%	1.055	0.826	1.333
E	181	14	7.7%	8.0	4.4%	1.761	1.072	2.554
F	201	4	2.0%	10.2	5.1%	0.393	0.102	0.804
G	91	6	6.6%	3.6	3.9%	1.681	0.481	3.189
H	242	11	4.5%	10.6	4.4%	1.035	0.598	1.659
I	284	8	2.8%	12.6	4.4%	0.635	0.298	1.081
J	150	12	8.0%	8.1	5.4%	1.479	0.838	2.313

b) Postoperative complications summary – sample restricted to hospital and patient reported outcomes n=1462

Hospital	Number of surgeries	Number PO[5] complications	Observed PO CR	Expected number PO complications	Expected PO CR	O/E PO CR ratio	Lower 95% CI for O/E ratio	Upper 95% CI for O/E ratio
A	82	20	24.4%	20.9	25.4%	0.958	0.616	1.325
B	285	76	26.7%	69.7	24.4%	1.091	0.942	1.296
C	36	12	33.3%	10.3	28.5%	1.171	0.688	1.723
D	394	110	27.9%	108.2	27.5%	1.017	0.902	1.128
E	129	30	23.3%	33.0	25.5%	0.910	0.638	1.157
F	113	29	25.7%	28.2	24.9%	1.029	0.769	1.370

G	45	7	15.6%	11.1	24.8%	0.628	0.265	1.098
H	162	44	27.2%	40.5	25.0%	1.086	0.881	1.355
I	158	27	17.1%	31.8	20.1%	0.850	0.562	1.099
J	58	21	36.2%	15.4	26.6%	1.361	0.917	1.828

Table 3: Risk prediction model for Postoperative complications – sample restricted to hospital and patient reported outcomes n=1462

Variables	Lasso				Maximum Likelihood				Shrinkage
	OR	p-value*	L95% CI*	U95% CI*	OR	p-value	L95% CI	U95% CI	
Age at surgery	0.997	0.552	0.986	1.008	0.991	0.160	0.979	1.003	-57.3%
BMI	1.012	0.213	0.993	1.031	1.020	0.059	0.999	1.041	-38.6%
Previous abdominal surgery	1.008	0.954	0.774	1.313	1.096	0.501	0.838	1.434	-90.4%
Coagulation-thrombosis	1.510	0.202	0.802	2.842	2.130	0.022	1.115	4.072	-45.1%
Diabetes	1.355	0.145	0.901	2.038	1.565	0.038	1.024	2.392	-31.3%
Cardiac	*removed*				1.036	0.878	0.660	1.627	-100.0%
Respiratory	*removed*				1.146	0.536	0.744	1.763	-100.0%
Gastrointestinal	0.916	0.798	0.467	1.796	0.673	0.291	0.322	1.405	-77.4%
Genitourinary	*removed*				1.371	0.490	0.560	3.354	-100.0%
Musculoskeletal	1.254	0.265	0.842	1.868	1.555	0.033	1.037	2.333	-50.3%
Neurology-psychiatric	0.908	0.722	0.533	1.546	0.693	0.217	0.387	1.241	-69.0%
Vascular	0.926	0.847	0.423	2.024	0.692	0.412	0.287	1.669	-78.6%
Auto-immune	*removed*				0.531	0.366	0.134	2.098	-100.0%
Metabolic-endocrine	*removed*				1.120	0.586	0.744	1.688	-100.0%
Integumentary-dermatology	*removed*				0.981	0.977	0.262	3.672	-100.0%
Hypertension	*removed*				1.129	0.431	0.834	1.530	-100.0%
Smoking	*removed*				1.341	0.467	0.609	2.955	-100.0%
Other neoplasms	*removed*				1.167	0.577	0.678	2.009	-100.0%
Laparoscopic approach	0.739	0.084	0.525	1.042	0.649	0.020	0.451	0.935	-28.7%
ASA									
ASA grade 1	1								
ASA grade 2	*removed*				0.942	0.745	0.659	1.348	-100.0%
ASA grade 3+	*removed*				0.812	0.407	0.497	1.328	-100.0%
Surgeon grade									
Consultant	1								

Sub-specialty trainee	*removed*				0.864	0.658	0.453	1.648	-100.0%
General Obstetrics & Gynaecology Trainee	*removed*				0.873	0.656	0.480	1.588	-100.0%
Surgical complexity									
Complexity score 1&2	1								
Complexity score 3&4	1.112	0.437	0.851	1.454	1.322	0.078	0.969	1.804	-61.8%
Complexity score 5&6	*removed*				1.054	0.819	0.670	1.659	-100.0%
Complexity score 7&8	1.056	0.881	0.516	2.163	1.480	0.313	0.691	3.169	-83.3%
Complexity score >8	1.004	0.970	0.828	1.217	1.748	0.322	0.579	5.279	-90.6%
Final diagnosis									
Ovarian	1								
Uterine	0.982	0.911	0.719	1.343	0.951	0.790	0.658	1.376	-82.5%
Cervical	1.606	0.094	0.923	2.794	2.099	0.016	1.148	3.836	-33.4%
Vulval	1.779	0.030	1.056	2.999	2.274	0.003	1.311	3.943	-27.7%
Benign	*removed*				1.058	0.775	0.720	1.554	-100.0%
Duration of surgery (hrs)	1.086	0.003	1.028	1.146	1.081	0.018	1.014	1.152	-6.5%
Estimated blood loss									
<500ml	1								
500ml-1000ml	1.267	0.208	0.876	1.833	1.405	0.077	0.963	2.048	-32.1%
1000ml-2500ml	1.052	0.860	0.600	1.843	1.249	0.442	0.709	2.202	-76.4%
>2500ml	0.997	0.962	0.867	1.146	0.506	0.417	0.098	2.623	-86.6%
constant	0.167	0.000	0.066	0.422	0.179	0.003	0.057	0.564	-25.5%

* for approximate guidance only. Low albumin not included in the model as only one instance of it amongst 1462 surgeries

Figure S1: Comparison of observed postoperative complication rates against colour-coded funnel plots. Bright colour-coding and circle markers represent patient-reported statistics and faded colour-coding and square markers represent hospital reported statistics only.

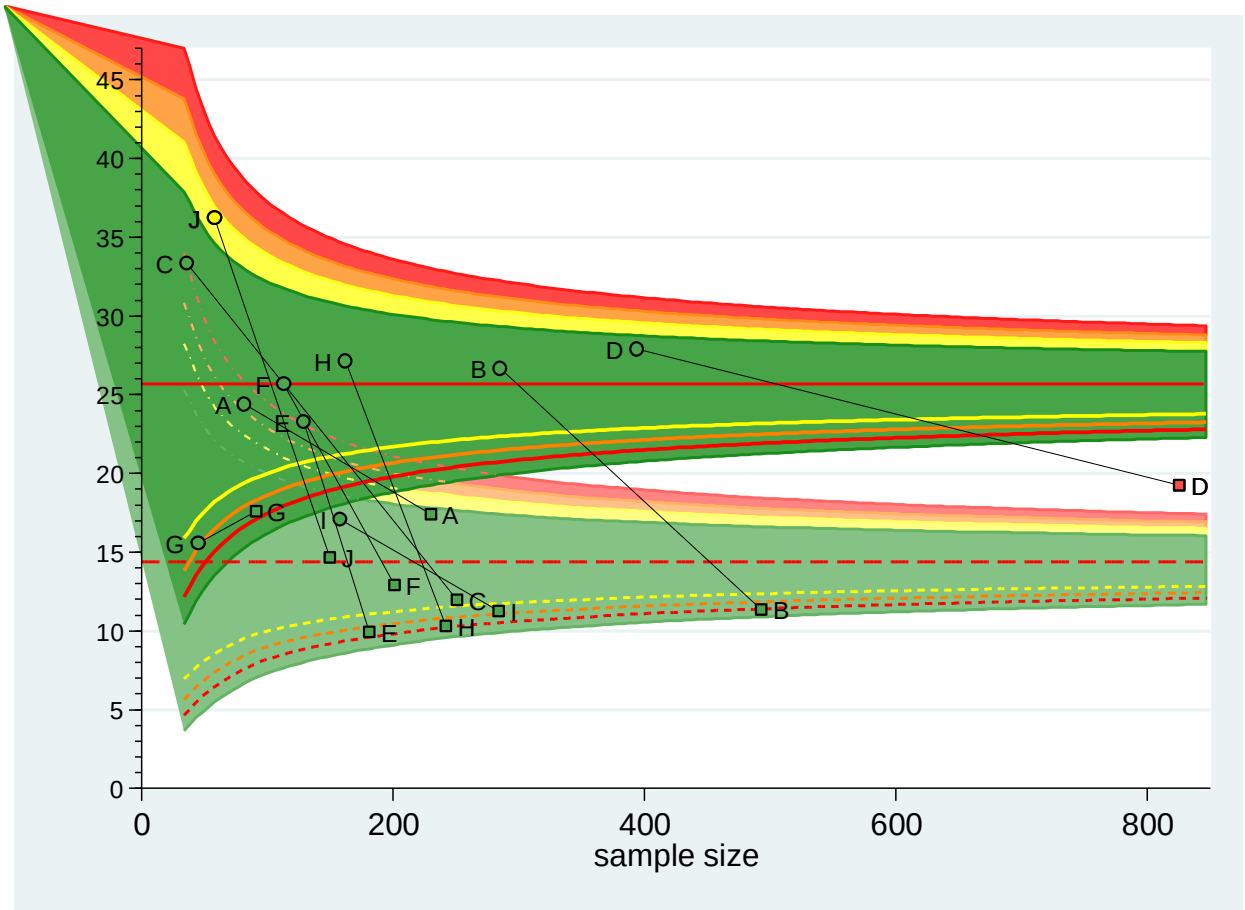

Table S1: Intra-operative complications	
Anaesthetic complications	
Cardiac	e.g. Cardiac arrythmias, Intra-operative cardic arrest
Respiratory	e.g. Aspiration, pneumothorax, pulmonary oedema
Allergic reactions	Allergic reactions including anaphylaxis
Injury to viscera	
Uterine perforation	Perforation of uterus during instrumentation
Vascular injury	Injury to major blood vessel e.g. superior and inferior mesenteric, renal, aorta, Inferior vena cava, iliacs, femorals,
GI tract injury – Stomach	Accidental injury involving complete penetration Into the lumen: Stomach
GI tract injury – Small bowel	Accidental injury involving complete penetration Into the lumen:Small bowel
GI tract injury – Large bowel	Accidental injury involving complete penetration Into the lumen: Large bowel
Bladder injury	Accidental bladder injury (full thickness)
Ureteric injury	Ligation / Transection / Diathermy burn
Intra-operative Haemorrhage	Estimated blood loss >2.5l
Other intra-operative complications (give details)	Other intraoperative complications not included in the list

Table S2: Post-operative complications	
Abscess/Haematoma	Pelvic or abdominal abscess / haematoma
Anastomotic leak	Anastomotic leak: Small bowel
	Anastomotic leak: Large bowel
Ileus	Post op Ileus requiring NG tube / Total parental nutrition
Bowel obstruction	Bowel Obstruction – small bowel
	Bowel Obstruction – large bowel
Bowel perforation	Small / large bowel
Bowel - other	Constipation / Diarrhoea / faecal incontinence/urgency
Bladder	Urinary retention requiring catheterisation
	Urinary obstruction
	Incontinence- stress / urge
Cardiac	Atrial fibrillation, Myocardial infarction, Cardiac failure & other cardiac problems
DVT	Confirmed DVT on imaging / Doppler
PE	Confirmed PE on imaging
Fistula	Enterocutaneous
	Enterovaginal
	Vesicovaginal
	Ureterovaginal
	Other types of fistula
Hernia	Hernia as a result of surgery
Infection	Pyrexia (>38.5°C on 2 separate occasions) after 48 hours post op requiring antibiotics or infection confirmed by culture
	MRSA/ C. difficile
Lymphocyst/Lymphoedema	Lymphoedema
	Lymphocyst
Neurological	Neuropathic pain/ paraesthesia / nerve palsy

Psychiatric	unexpected psychiatric problems postoperatively e.g. Delirium, Psychosis, Depression and other
Primary haemorrhage	Haemorrhage within 24 hours of surgery
Secondary haemorrhage	Haemorrhage after 24hours of surgery
Respiratory	Pulmonary oedema, Pneumothorax, Atelectasis, Pleural effusion and other respiratory problems excluding pneumonia (to be included in infections)
Ureteric Obstruction	Ureteric obstruction postoperative
Wound breakdown	Wound breakdown: Superficial - skin & subcutaneous tissue
	Wound breakdown: Deep - involving fascia / muscle
	Burst abdomen requiring repair under anaesthesia
Other	Other postoperative complications not included in the list

Table S3: Full dataset used for intra-operative complication analysis and hospital reported post operative complications n=2948																									
	Hospital																				Overall		chi2	df	pvalue
	A		B		C		D		E		F		G		H		I		J						
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%			
Intraop complications	13	*5.7*	17	*3.5*	10	*4.0*	44	*5.3*	14	*7.7*	4	*2.0*	6	*6.6*	11	*4.6*	8	*2.8*	12	*8.0*	139	*4.7*	16.8	9	0.0517
Postop complications (hospital reported)	40	*17.4*	56	*11.4*	30	*12.0*	159	*19.3*	18	*9.9*	26	*12.9*	16	*17.6*	25	*10.3*	32	*11.3*	22	*14.7*	424	*14.4*	32.0	9	0.0002
Previous abdominal surgery	83	*36.1*	168	*34.1*	64	*25.5*	306	*37.1*	42	*23.2*	84	*41.8*	33	*36.3*	109	*45.0*	117	*41.2*	19	*12.7*	1,025	*34.8*	75.6	9	0.0000
Low Albumin	0	*0.0*	5	*1.0*	0	*0.0*	1	*0.1*	0	*0.0*	0	*0.0*	1	*1.1*	3	*1.2*	0	*0.0*	1	*0.7*	11	*0.4*	17.7	9	0.0391
Coagulation-thrombosis	9	*3.9*	16	*3.3*	12	*4.8*	24	*2.9*	8	*4.4*	9	*4.5*	6	*6.6*	17	*7.0*	9	*3.2*	6	*4.0*	116	*3.9*	11.9	9	0.2180
Diabetes	21	*9.1*	52	*10.6*	30	*12.0*	94	*11.4*	12	*6.6*	15	*7.5*	5	*5.5*	30	*12.4*	18	*6.3*	21	*14.0*	298	*10.1*	17.2	9	0.0455
Cardiac	17	*7.4*	55	*11.2*	32	*12.8*	75	*9.1*	21	*11.6*	23	*11.4*	7	*7.7*	29	*12.0*	27	*9.5*	22	*14.7*	308	*10.5*	10.5	9	0.3081
Respiratory	21	*9.1*	58	*11.8*	21	*8.4*	74	*9.0*	14	*7.7*	20	*10.0*	5	*5.5*	37	*15.3*	21	*7.4*	16	*10.7*	287	*9.7*	16.6	9	0.0553
Gastrointestinal	10	*4.4*	10	*2.0*	6	*2.4*	21	*2.6*	9	*5.0*	17	*8.5*	6	*6.6*	13	*5.4*	3	*1.1*	9	*6.0*	104	*3.5*	35.2	9	0.0001
Genitourinary	2	*0.9*	5	*1.0*	2	*0.8*	7	*0.9*	6	*3.3*	5	*2.5*	0	*0.0*	14	*5.8*	8	*2.8*	3	*2.0*	52	*1.8*	37.2	9	0.0000
Musculoskeletal	15	*6.5*	23	*4.7*	15	*6.0*	98	*11.9*	13	*7.2*	19	*9.5*	4	*4.4*	34	*14.1*	21	*7.4*	19	*12.7*	261	*8.9*	38.7	9	0.0000
Neurology-psychiatric	18	*7.8*	37	*7.5*	12	*4.8*	65	*7.9*	8	*4.4*	15	*7.5*	7	*7.7*	17	*7.0*	22	*7.8*	7	*4.7*	208	*7.1*	6.7	9	0.6652
Vascular	4	*1.7*	18	*3.7*	5	*2.0*	18	*2.2*	3	*1.7*	9	*4.5*	4	*4.4*	8	*3.3*	11	*3.9*	6	*4.0*	86	*2.9*	9.5	9	0.3916
Infections	2	*0.9*	8	*1.6*	2	*0.8*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	1	*0.4*	0	*0.0*	13	*0.4*	24.9	9	0.0031
Auto-immune	4	*1.7*	8	*1.6*	2	*0.8*	6	*0.7*	3	*1.7*	5	*2.5*	0	*0.0*	2	*0.8*	5	*1.8*	3	*2.0*	38	*1.3*	8.5	9	0.4874
Metabolic-endocrine	25	*10.9*	47	*9.5*	25	*10.0*	67	*8.1*	15	*8.3*	27	*13.4*	15	*16.5*	35	*14.5*	35	*12.3*	11	*7.3*	302	*10.2*	18.7	9	0.0282
Integumentary-dermatology	4	*1.7*	2	*0.4*	3	*1.2*	6	*0.7*	1	*0.6*	3	*1.5*	0	*0.0*	8	*3.3*	3	*1.1*	0	*0.0*	30	*1.0*	19.7	9	0.0199
Hypertension	61	*26.5*	186	*37.7*	72	*28.7*	257	*31.2*	66	*36.5*	78	*38.8*	19	*20.9*	110	*45.5*	71	*25.0*	53	*35.3*	973	*33.0*	48.4	9	0.0000
Smoking	1	*0.4*	25	*5.1*	15	*6.0*	3	*0.4*	5	*2.8*	10	*5.0*	16	*17.6*	14	*5.8*	1	*0.4*	5	*3.3*	95	*3.2*	113.7	9	0.0000
Other neoplasms	24	*10.4*	15	*3.0*	6	*2.4*	40	*4.9*	11	*6.1*	6	*3.0*	5	*5.5*	16	*6.6*	19	*6.7*	6	*4.0*	148	*5.0*	27.4	9	0.0012
Surgeon grade																							302.4	18	0.0000
General Obstetrics & Gynaecology Trainee	3	*1.3*	22	*4.6*	3	*1.3*	15	*1.8*	25	*15.0*	3	*1.5*	0	*0.0*	2	*0.9*	30	*11.1*	5	*3.4*	108	*3.8*			
Sub-specialty trainee	87	*37.8*	95	*19.9*	65	*27.2*	172	*20.9*	8	*4.8*	3	*1.5*	23	*25.3*	80	*35.1*	35	*12.9*	5	*3.4*	573	*20.0*			
Consultant	140	*60.9*	360	*75.5*	171	*71.6*	638	*77.3*	134	*80.2*	192	*97.0*	68	*74.7*	146	*64.0*	206	*76.0*	136	*93.2*	2,191	*76.3*			
Laparoscopic approach	109	*47.4*	40	*8.1*	44	*17.5*	208	*25.2*	8	*4.4*	11	*5.5*	26	*28.6*	62	*25.6*	152	*53.5*	21	*14.0*	681	*23.1*	373.2	9	0.0000
ASA Grade																							145.5	27	0.0000

ASA grade 1	44	*19.2*	142	*28.8*	107	*43.3*	194	*23.5*	32	*17.7*	58	*28.9*	23	*25.3*	19	*7.9*	97	*34.2*	38	*25.3*	754	*25.6*			
ASA grade 2	149	*65.1*	262	*53.1*	115	*46.6*	438	*53.1*	93	*51.4*	101	*50.3*	56	*61.5*	156	*65.0*	138	*48.6*	79	*52.7*	1,587	*54.0*			
ASA grade 3+	36	*15.7*	89	*18.1*	25	*10.1*	193	*23.4*	56	*30.9*	42	*20.9*	12	*13.2*	65	*27.1*	48	*16.9*	33	*22.0*	599	*20.4*			
Surgical complexity																							228.9	36	0.0000
Complexity score 1&2	97	*42.2*	149	*30.2*	160	*63.8*	395	*47.9*	118	*65.2*	72	*35.8*	48	*52.8*	124	*51.2*	181	*63.7*	54	*36.0*	1,398	*47.4*			
Complexity score 3&4	79	*34.4*	199	*40.4*	62	*24.7*	276	*33.5*	38	*21.0*	88	*43.8*	39	*42.9*	79	*32.6*	70	*24.7*	52	*34.7*	982	*33.3*			
Complexity score 5&6	41	*17.8*	111	*22.5*	17	*6.8*	105	*12.7*	23	*12.7*	37	*18.4*	3	*3.3*	35	*14.5*	19	*6.7*	39	*26.0*	430	*14.6*			
Complexity score 7&8	9	*3.9*	24	*4.9*	5	*2.0*	34	*4.1*	2	*1.1*	4	*2.0*	1	*1.1*	3	*1.2*	6	*2.1*	5	*3.3*	93	*3.2*			
Complexity score >8	4	*1.7*	10	*2.0*	7	*2.8*	15	*1.8*	0	*0.0*	0	*0.0*	0	*0.0*	1	*0.4*	8	*2.8*	0	*0.0*	45	*1.5*			
Final diagnosis																							274.2	36	0.0000
Ovarian	94	*40.9*	123	*25.0*	99	*39.4*	305	*37.0*	57	*31.5*	82	*40.8*	27	*29.7*	78	*32.2*	64	*22.5*	60	*40.0*	989	*33.6*			
Uterine	70	*30.4*	119	*24.1*	73	*29.1*	243	*29.5*	56	*30.9*	51	*25.4*	37	*40.7*	60	*24.8*	62	*21.8*	49	*32.7*	820	*27.8*			
Cervical	18	*7.8*	19	*3.9*	20	*8.0*	82	*9.9*	16	*8.8*	9	*4.5*	5	*5.5*	16	*6.6*	11	*3.9*	11	*7.3*	207	*7.0*			
Vulval	3	*1.3*	18	*3.7*	14	*5.6*	69	*8.4*	12	*6.6*	12	*6.0*	4	*4.4*	20	*8.3*	12	*4.2*	12	*8.0*	176	*6.0*			
Benign	45	*19.6*	214	*43.4*	45	*17.9*	126	*15.3*	40	*22.1*	47	*23.4*	18	*19.8*	68	*28.1*	135	*47.5*	18	*12.0*	756	*25.6*			
Estimated blood loss																							246.2	27	0.0000
<500ml	155	*67.4*	400	*82.0*	196	*78.7*	709	*87.4*	93	*51.4*	161	*80.9*	69	*75.8*	198	*81.8*	261	*93.2*	93	*62.0*	2,335	*79.9*			
500ml-1000ml	50	*21.7*	49	*10.0*	33	*13.3*	62	*7.6*	69	*38.1*	23	*11.6*	12	*13.2*	31	*12.8*	11	*3.9*	38	*25.3*	378	*12.9*			
1000ml-2500ml	19	*8.3*	33	*6.8*	16	*6.4*	37	*4.6*	17	*9.4*	15	*7.5*	7	*7.7*	10	*4.1*	7	*2.5*	13	*8.7*	174	*6.0*			
>2500ml	6	*2.6*	6	*1.2*	4	*1.6*	3	*0.4*	2	*1.1*	0	*0.0*	3	*3.3*	3	*1.2*	1	*0.4*	6	*4.0*	34	*1.2*			
	mean	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**F**	**df***	**pvalue**
Age at surgery	58.6	*13.8*	59.5	*15.4*	58.4	*15.7*	60.6	*15.0*	63.2	*12.1*	63.1	*13.6*	57.2	*12.0*	63.3	*16.0*	59.6	*14.4*	64.1	*14.0*	60.6	*14.8*	5.2	9	0.0000
BMI	27.6	*6.7*	28.8	*6.9*	28.2	*8.3*	29.1	*7.2*	30.0	*6.8*	28.5	*6.1*	29.6	*8.4*	29.2	*6.7*	28.9	*6.4*	28.4	*6.7*	28.8	*7.0*	2.1	9	0.0304
Duration of surgery (hrs)	141	*59.1*	114	*53.7*	151	*85.1*	159	*83.4*	116	*52.0*	122	*57.0*	136	*56.8*	121	*58.8*	101	*47.9*	128	*55.9*	133.3	*69.9*	31.4	9	0.0000

Table S4: Subset used for post-operative complication analysis where both hospital and patient reported data was available n=1462																									
	Hospital																				Overall		chi2	df	pvalue
	A		B		C		D		E		F		G		H		I		J						
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%			
Intraop complications	3	*3.7*	10	*3.5*	2	*5.6*	22	*5.6*	10	*7.8*	3	*2.7*	3	*6.7*	8	*4.9*	5	*3.2*	2	*3.5*	68	*4.7*	7.1	9	0.6276
Postop complications (hospital and patient reported)	20	*24.4*	76	*26.7*	12	*33.3*	110	*27.9*	30	*23.3*	29	*25.7*	7	*15.6*	44	*27.2*	27	*17.1*	21	*36.2*	376	*25.7*	14.8	9	0.0960
Previous abdominal surgery	23	*28.1*	99	*34.7*	13	*36.1*	141	*35.8*	30	*23.3*	42	*37.2*	14	*31.1*	70	*43.2*	66	*41.8*	9	*15.5*	507	*34.7*	27.9	9	0.0010
Low Albumin	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	1	*0.6*	0	*0.0*	0	*0.0*	1	*0.1*	8.0	9	0.5311
Coagulation-thrombosis	3	*3.7*	9	*3.2*	2	*5.6*	12	*3.1*	7	*5.4*	3	*2.7*	3	*6.7*	11	*6.8*	5	*3.2*	1	*1.7*	56	*3.8*	8.3	9	0.4997
Diabetes	7	*8.5*	28	*9.8*	5	*13.9*	43	*10.9*	6	*4.7*	7	*6.2*	3	*6.7*	22	*13.6*	13	*8.2*	11	*19.0*	145	*9.9*	15.8	9	0.0713
Cardiac	5	*6.1*	36	*12.6*	10	*27.8*	29	*7.4*	15	*11.6*	12	*10.6*	3	*6.7*	17	*10.5*	18	*11.4*	8	*13.8*	153	*10.5*	20.4	9	0.0157
Respiratory	7	*8.5*	40	*14.0*	3	*8.3*	26	*6.6*	9	*7.0*	11	*9.7*	2	*4.4*	22	*13.6*	10	*6.3*	4	*6.9*	134	*9.2*	19.0	9	0.0254
Gastrointestinal	1	*1.2*	3	*1.1*	1	*2.8*	12	*3.1*	8	*6.2*	7	*6.2*	5	*11.1*	7	*4.3*	1	*0.6*	4	*6.9*	49	*3.4*	26.7	9	0.0016
Genitourinary	2	*2.4*	2	*0.7*	0	*0.0*	2	*0.5*	4	*3.1*	1	*0.9*	0	*0.0*	11	*6.8*	6	*3.8*	0	*0.0*	28	*1.9*	34.3	9	0.0001
Musculoskeletal	7	*8.5*	17	*6.0*	2	*5.6*	57	*14.5*	8	*6.2*	11	*9.7*	3	*6.7*	28	*17.3*	13	*8.2*	7	*12.1*	153	*10.5*	26.4	9	0.0017
Neurology-psychiatric	5	*6.1*	14	*4.9*	1	*2.8*	22	*5.6*	5	*3.9*	7	*6.2*	2	*4.4*	12	*7.4*	16	*10.1*	4	*6.9*	88	*6.0*	8.0	9	0.5329
Vascular	2	*2.4*	9	*3.2*	1	*2.8*	6	*1.5*	3	*2.3*	4	*3.5*	2	*4.4*	5	*3.1*	6	*3.8*	2	*3.5*	40	*2.7*	4.1	9	0.9046
Infections	0	*0.0*	6	*2.1*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	1	*0.6*	0	*0.0*	7	*0.5*	20.8	9	0.0136
Auto-immune	1	*1.2*	3	*1.1*	1	*2.8*	4	*1.0*	3	*2.3*	2	*1.8*	0	*0.0*	1	*0.6*	2	*1.3*	0	*0.0*	17	*1.2*	4.4	9	0.8794
Metabolic-endocrine	9	*11.0*	22	*7.7*	4	*11.1*	32	*8.1*	12	*9.3*	14	*12.4*	7	*15.6*	23	*14.2*	22	*13.9*	5	*8.6*	150	*10.3*	11.3	9	0.2567
Integumentary-dermatology	2	*2.4*	0	*0.0*	1	*2.8*	1	*0.3*	1	*0.8*	1	*0.9*	0	*0.0*	6	*3.7*	1	*0.6*	0	*0.0*	13	*0.9*	23.7	9	0.0048
Hypertension	25	*30.5*	113	*39.7*	12	*33.3*	127	*32.2*	47	*36.4*	47	*41.6*	12	*26.7*	74	*45.7*	51	*32.3*	17	*29.3*	525	*35.9*	17.2	9	0.0458
Smoking	0	*0.0*	14	*4.9*	0	*0.0*	1	*0.3*	2	*1.6*	2	*1.8*	7	*15.6*	9	*5.6*	1	*0.6*	1	*1.7*	37	*2.5*	58.1	9	0.0000
Other neoplasms	10	*12.2*	9	*3.2*	1	*2.8*	27	*6.9*	6	*4.7*	3	*2.7*	3	*6.7*	13	*8.0*	10	*6.3*	5	*8.6*	87	*6.0*	15.6	9	0.0768
Surgeon grade																							201.5	18	0.0000
General Obstetrics & Gynaecology Trainee	2	*2.4*	17	*6.2*	1	*2.9*	6	*1.5*	21	*17.8*	3	*2.7*	0	*0.0*	2	*1.3*	19	*12.5*	3	*5.3*	74	*5.2*			
Sub-specialty trainee	35	*42.7*	55	*20.2*	16	*47.1*	82	*20.8*	5	*4.2*	1	*0.9*	15	*33.3*	58	*38.2*	21	*13.8*	1	*1.8*	289	*20.4*			
Consultant	45	*54.9*	201	*73.6*	17	*50.0*	306	*77.7*	92	*78.0*	107	*96.4*	30	*66.7*	92	*60.5*	112	*73.7*	53	*93.0*	1,055	*74.4*			
Laparoscopic approach	37	*45.1*	21	*7.4*	4	*11.1*	99	*25.1*	8	*6.2*	6	*5.3*	15	*33.3*	43	*26.5*	91	*57.6*	9	*15.5*	333	*22.8*	220.3	9	0.0000

ASA Grade																						49.1	18	0.0001
ASA grade 1	10	*12.4*	80	*28.1*	12	*33.3*	90	*22.8*	23	*17.8*	29	*25.7*	9	*20.0*	15	*9.3*	45	*28.5*	17	*29.3*	330	*22.6*			
ASA grade 2	58	*71.6*	155	*54.4*	21	*58.3*	230	*58.4*	70	*54.3*	64	*56.6*	32	*71.1*	106	*65.8*	82	*51.9*	29	*50.0*	847	*58.0*			
ASA grade 3+	13	*16.1*	50	*17.5*	3	*8.3*	74	*18.8*	36	*27.9*	20	*17.7*	4	*8.9*	40	*24.8*	31	*19.6*	12	*20.7*	283	*19.4*			
Surgical complexity																						125.8	36	0.0000
Complexity score 1&2	36	*43.9*	89	*31.2*	25	*69.4*	179	*45.4*	87	*67.4*	41	*36.3*	22	*48.9*	84	*51.9*	106	*67.1*	21	*36.2*	690	*47.2*			
Complexity score 3&4	30	*36.6*	115	*40.4*	9	*25.0*	143	*36.3*	27	*20.9*	46	*40.7*	20	*44.4*	53	*32.7*	38	*24.1*	23	*39.7*	504	*34.5*			
Complexity score 5&6	11	*13.4*	60	*21.1*	1	*2.8*	50	*12.7*	13	*10.1*	22	*19.5*	2	*4.4*	23	*14.2*	10	*6.3*	14	*24.1*	206	*14.1*			
Complexity score 7&8	3	*3.7*	16	*5.6*	0	*0.0*	15	*3.8*	2	*1.6*	4	*3.5*	1	*2.2*	2	*1.2*	1	*0.6*	0	*0.0*	44	*3.0*			
Complexity score >8	2	*2.4*	5	*1.8*	1	*2.8*	7	*1.8*	0	*0.0*	0	*0.0*	0	*0.0*	0	*0.0*	3	*1.9*	0	*0.0*	18	*1.2*			
Final diagnosis																						163.5	36	0.0000
Ovarian	39	*47.6*	70	*24.6*	17	*47.2*	157	*39.9*	39	*30.2*	46	*40.7*	11	*24.4*	54	*33.3*	34	*21.5*	14	*24.1*	481	*32.9*			
Uterine	25	*30.5*	68	*23.9*	13	*36.1*	124	*31.5*	40	*31.0*	33	*29.2*	22	*48.9*	41	*25.3*	40	*25.3*	21	*36.2*	427	*29.2*			
Cervical	7	*8.5*	7	*2.5*	0	*0.0*	29	*7.4*	12	*9.3*	4	*3.5*	3	*6.7*	10	*6.2*	3	*1.9*	5	*8.6*	80	*5.5*			
Vulval	0	*0.0*	14	*4.9*	1	*2.8*	26	*6.6*	7	*5.4*	6	*5.3*	0	*0.0*	10	*6.2*	8	*5.1*	7	*12.1*	79	*5.4*			
Benign	11	*13.4*	126	*44.2*	5	*13.9*	58	*14.7*	31	*24.0*	24	*21.2*	9	*20.0*	47	*29.0*	73	*46.2*	11	*19.0*	395	*27.0*			
Estimated blood loss																								
<500ml	54	*65.9*	237	*84.0*	24	*68.6*	331	*86.0*	68	*52.7*	91	*80.5*	35	*77.8*	135	*83.3*	148	*94.9*	39	*67.2*	1,162	*80.3*	140.4	27	0.0000
500ml-1000ml	20	*24.4*	25	*8.9*	8	*22.9*	35	*9.1*	46	*35.7*	14	*12.4*	5	*11.1*	21	*13.0*	4	*2.6*	17	*29.3*	195	*13.5*			
1000ml-2500ml	7	*8.5*	17	*6.0*	3	*8.6*	17	*4.4*	13	*10.1*	8	*7.1*	3	*6.7*	4	*2.5*	4	*2.6*	2	*3.5*	78	*5.4*			
>2500ml	1	*1.2*	3	*1.1*	0	*0.0*	2	*0.5*	2	*1.6*	0	*0.0*	2	*4.4*	2	*1.2*	0	*0.0*	0	*0.0*	12	*0.8*			
	mean	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**mean**	***sd***	**F**	**df***	**pvalue**
Age at surgery	59.6	*12.2*	60.5	*14.4*	62.1	*12.8*	62.8	*12.8*	63.1	*11.7*	63.8	*12.3*	59.1	*11.8*	64.5	*14.8*	62.0	*12.9*	65.4	*13.3*	62.4	*13.3*	2.4	9	0.0107
BMI	27.8	*7.8*	28.8	*6.7*	31.1	*8.8*	28.5	*6.8*	29.7	*6.6*	28.3	*5.5*	30.2	*10.0*	29.2	*6.9*	29.1	*6.3*	28.6	*6.3*	28.9	*6.9*	1.2	9	0.2625
Duration of surgery (hrs)	148.9	*60.6*	110.6	*51.3*	146.1	*91.2*	161.8	*80.8*	110.3	*51.0*	121.3	*55.4*	138.8	*53.4*	122.6	*56.6*	95.1	*42.6*	114.4	*34.8*	128.9	*66.2*	23.7	9	0.0000