Comparison of the prognostic accuracy of scoring systems, cardiopulmonary exercise testing, and plasma biomarkers: a single-centre observational pilot study

S. James 1, S. Jhanji 2, A. Smith 3, G. O’Brien 4, M. Fitzgibbon 4 and R. M. Pearse 1*

1 Barts and the London School of Medicine and Dentistry, Queen Mary’s University of London, UK
2 Royal Marsden NHS Trust, London, UK
3 Barts Health NHS Trust, London, UK
4 Mater Misericordiae University Hospital, Dublin, Ireland

* Corresponding author: Adult Critical Care Unit, Royal London Hospital, London E1 1BB, UK. E-mail: r.pearse@qmul.ac.uk

Background. Current approaches to risk assessment before major surgery have important limitations. The aim of this pilot study was to compare predictive accuracy of preoperative scoring systems, plasma biomarkers, and cardiopulmonary exercise testing (CPET) for complications after major non-cardiac surgery.

Methods. Single-centre, observational study of patients aged ≥ 40 yr undergoing major elective non-cardiac surgery. Before surgery, risk scores were calculated and blood samples collected for measurement of plasma biomarkers. Patients underwent CPET for measurement of anaerobic threshold (AT) and peak oxygen consumption (VO2 peak). After surgery, patients were followed for 28 days to evaluate complications and major adverse cardiac events (MACE). Data are presented as area under the receiver operating characteristic curve (AUROC) with 95% confidence intervals.

Results. A total of 100 patients were recruited between April 2009 and October 2010; 17 of whom did not proceed to surgery. CPET variables suggested good predictive accuracy for MACE [AT: AUROC 0.83 (0.69–0.96); VO2 peak AUROC 0.81 (0.69–0.96)] and poor predictive accuracy for all complications [AT: AUROC 0.64 (0.52–0.77); VO2 peak AUROC 0.64 (0.52–0.77)]. There was a trend towards predictive accuracy of the plasma biomarkers B-type natriuretic peptide and estimated glomerular filtration rate (calculated from serum creatinine) for MACE but not all complications. C-reactive protein, ASA score, and revised cardiac risk index had little or no predictive value.

Conclusions. These pilot data suggest that CPET and plasma biomarkers may improve risk assessment before surgery. Only large clinical studies can confirm this observation and define the optimal use of these tests in clinical practice.

Keywords: exercise test; postoperative complications, diagnosis; predictive value of tests; surgical procedures, operative

Accepted for publication: 5 July 2013

Postoperative complications represent an important cause of avoidable morbidity and mortality. 1–4 Approximately 234 million patients undergo surgery worldwide each year, around 25% of whom will develop complications. 5–6 Estimates of hospital mortality rates for in-patient non-cardiac surgery may be as high as 2–4%. 1–6 Long-term survival is significantly reduced after surgery for those patients who develop complications but survive to leave hospital. 5–7 Importantly, around 80% of deaths occur among a subgroup of patients who can be identified as being at high risk before surgery is performed. 2–3

Typical risk factors for postoperative complications include advanced age, co-morbid disease, and major surgery. However, evidence suggests that most patients who meet these high-risk criteria do not receive additional care to improve the prospect of good postoperative outcomes. 1–3 One reason for this may be the subjective nature of the current approach to clinical assessment of risk which often requires a doctor to make the case for additional care for individual high-risk patients. An alternative approach may be the use of objective measures of risk including scoring systems, 8–9 plasma biomarkers, 10–21 and cardiopulmonary exercise testing (CPET). 22–26 Each of these approaches has strengths and limitations in the assessment of individual patient risk. Issues of feasibility, accuracy, and cost have led to debate over their relative merits and utility in clinical practice.
However, few data are available to provide direct comparisons between the most promising candidate methods. Our aim was to undertake a pilot study to compare the prognostic accuracy of CPET and the plasma biomarkers B-type natriuretic peptide (BNP), estimated glomerular filtration rate (eGFR; calculated from serum creatinine), and C-reactive protein (CRP) to the most commonly used clinical scoring systems to predict complications among patients undergoing elective major non-cardiac surgery.

**Methods**

**Patients**

This was a single-centre observational study of adult patients aged 40 yr and over, undergoing major elective non-cardiac surgery at the Royal London Hospital, London, UK. The study was approved prospectively by the East London and City Research ethics committee (reference 09/H0704/23) and written informed consent was given by all patients before enrolment. Exclusion criteria were unsuitability for CPET based on the American Thoracic Society guidelines, pregnancy, and refusal of consent. Potential participants were screened by local investigators having been identified in a consultant-led preoperative assessment clinic. The decision to offer surgery and allocation of postoperative critical care were made on the basis of established clinical protocols which did not include consideration of CPET data. However, CPET data were made available to anaesthetic staff on specific request. Plasma biomarker data were not made available to clinical staff. ASA physical status score was documented on the anaesthetic chart for all patients. Revised cardiac risk index (RCRI) is not routinely calculated by clinical staff.

**Preoperative data**

Baseline clinical data were collected allowing calculation of RCRI and the Portsmouth Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity (P-POSSUM). The ASA score (assessed by the attending anaesthetist) was also noted. Patients were then asked to lie quietly in the supine position before venepuncture. Within 30 min of collection, blood samples were centrifuged at 4300 rpm for 10 min and stored at −80°C for subsequent analysis. BNP was analysed by immunosay (Architect i2000SR, Abbott Diagnostics, USA). The measuring interval was 10–5000 pg ml⁻¹, the average analytical coefficient of variation was <5%, and the reference interval was <135 pg ml⁻¹. CRP was analysed using an immunoturbidimetric method (Architect c16000, Abbott Diagnostics) with a measuring interval of 0.1–160 mg litre⁻¹. The average analytical coefficient of variation was <2%, and the reference interval was <7 mg litre⁻¹. eGFR was calculated from age and serum creatinine with adjustment for ethnicity using themodification of diet in renal disease equation. Creatinine was analysed using a kinetic Jaffé reaction assay (Roche Diagnostics, Switzerland). On the day of exercise testing, patients were fasted for 2 h and refrained from caffeine in accordance with American Thoracic Society guidelines. They were seated on a computer-controlled, electromagnetically braked cycle ergometer and cycled at a cadence of 55–65 rpm. Subjects wore a tight facemask to permit continuous measurements of ventilation, oxygen consumption (VO₂), and carbon dioxide production (VCO₂) in expired gas. These measurements allowed estimation of peak oxygen consumption (VO₂ peak) and anaerobic threshold (AT) using a modified V-slope method. Arterial pressure, 12-lead ECG, and arterial pulse oximetry readings were taken at rest. During exercise, arterial pressure was measured every 3 min while pulse oximetry and ECG rhythm were monitored continuously. Subjects performed an incremental ramp test to the limit of tolerance. A ramp slope, based on gender, age, and estimated physical fitness of the subject, was set in order to obtain a test of ~10 min duration. The limit of tolerance (maximum) was defined as the point at which the subject could not maintain a pedalling cadence of 55–65 rpm despite encouragement. All tests were supervised by a physician trained in the analysis of these tests. Criteria for stopping or preventing a subject from being tested were also based on American Thoracic Society guidelines.

**Postoperative data**

After surgery, patients were followed up for 28 days by a research assistant who was not present at the exercise test. Data were collected describing pre-defined postoperative complications (Supplementary material) including major adverse cardiac events (MACE) (defined as myocardial infarction, cardiogenic pulmonary oedema, cardiac arrest, or complete heart block), duration of hospital stay, and mortality. The P-POSSUM score was completed for each patient immediately after surgery.

**Statistical analysis**

This was a pilot study to evaluate the feasibility and inform the design of a larger observational study. No formal sample size calculation was made. Instead, we planned to recruit 100 patients undergoing major surgery. Patients who were enrolled but did not subsequently undergo surgery were prospectively excluded from the analysis. Receiver operating characteristic (ROC) curves were constructed by first tabulating and then plotting the sensitivity and specificity of the test result at various cut-offs, then calculating the area under these curves (AUROC) to quantify overall prognostic discrimination for MACE and all complications. Categorical variables were tested with the Fisher’s exact test. AUROC is presented with 95% confidence intervals. Optimal threshold values and associated sensitivities and specificities are presented for variables with AUROC >0.8 for the corresponding outcome (good predictive accuracy). Other data are presented as mean (SD) where normally distributed or median (IQR) where not normally distributed. Statistical analysis was performed using GraphPad Prism version 4.0 (GraphPad Software, USA).

**Results**

One hundred patients were recruited between April 2009 and October 2010. After preoperative assessment, 17 patients did
A total of 83 patients who underwent CPET proceeded to surgery, all of whom were followed up for 28 days. The median operating time was 270 min (198–378). All patients underwent general anaesthesia, 64 (77%) in combination with epidural analgesia. Patient characteristics are presented in Table 1. Forty patients developed complications (48%); nine of whom developed MACE (11%) (Table 2). There were three deaths, all of which followed MACE. The median duration of hospital stay was 12 days (7–22) and three patients died. For those patients alive at day 28, the median number of critical care free days was 25 (21–26). Clinicians made few requests for CPET data, but unfortunately, we did not document the frequency of these requests and did not record any qualitative data on how they influenced patient care.

Univariate analysis of the association with outcomes for the candidate prognostic markers is presented in Table 3. For preoperative scoring systems, both ASA score [AUROC 0.60 (0.48–0.72) for all complications; 0.68 (0.49–0.87) for MACE] and RCRI [0.53 (0.40–0.65) for all complications; 0.68 (0.49–0.86) for MACE] were poor predictors of MACE and not predictive of all complications. For plasma biomarkers, CRP was not predictive of clinical outcome while BNP and eGFR had fair predictive accuracy for MACE but were not predictive of all complications (Fig. 2). The predictive value of CPET-derived variables was greater than that of scoring systems and plasma biomarkers. Both AT (sensitivity 75%, specificity 85% for an optimal threshold value of 10.6 ml min\(^{-1}\) kg\(^{-1}\)) and VO\(_2\) peak (sensitivity 88%, specificity 69% for an optimal threshold value of 14.0 ml min\(^{-1}\) kg\(^{-1}\)) were good predictors of MACE and poor predictors of all complications (Fig. 3).

**Discussion**

The principal findings of this small pilot study are that CPET and plasma biomarkers may have adequate prognostic accuracy in predicting which patients will develop complications after major non-cardiac surgery. CPET and biomarkers appeared to be more effective in the prediction of MACE than for unselected complications. Of the plasma biomarkers, BNP appeared to be the strongest prognostic marker followed by eGFR, while CRP was not predictive of any clinical outcomes. ASA score and RCRI were not predictive of overall complications in this small cohort and only poorly predictive of MACE. Our findings do not confirm the clinical utility of these tests but do suggest that a multi-centre study to confirm and extend these findings would be both feasible and worthwhile. These data may help in the estimation of event rates, and the rates of drop-out before and after surgery.

There is a growing evidence base describing the use of various tests for the prediction of postoperative outcomes. However, in many cases, this consists of small, single-centre case series with methodological limitations. In the case of CPET, there are few reports of the predictive accuracy of this test for complications after surgery. Two systematic reviews have suggested that CPET-derived variables are valid predictors of complications after non-cardiac surgery which may outperform alternative methods. Recent evidence suggests that CPET-derived variables may also be predictive of long-term postoperative mortality. However, there is substantial heterogeneity in the literature. Two studies, both notable for being clinician blinded, provide data describing the accuracy of CPET in predicting postoperative complications. In the first study, AUROC of only 0.59 for AT and 0.66 for VO\(_2\) peak are described, while investigators in the more recent study report AUROC of 0.85 for AT (AUROC for VO\(_2\) peak not reported). Our findings are consistent with those of published studies suggesting a role for BNP in the preoperative assessment of risk. The findings of a systematic review also suggest a role for this biomarker, in particular in the prediction of MACE. There are less published data describing the potential for eGFR in risk assessment before non-cardiac surgery patients but the available data are encouraging, and it is interesting to note that preoperative serum creatinine was an independent risk factor for MACE in the work which defined RCRI. We identified two publications exploring risk assessment with CRP and it is not possible to confirm the utility of this biomarker at the current time. The findings of one study suggest some predictive value of preoperative CRP measurement, while those of a more recent study suggested a weak relationship in keeping with our findings in the current study. Recent data have demonstrated the value of the ASA score in describing the severity of co-morbid disease in a large patient cohort. However, the limited number of

---

**Fig 1** Patient recruitment and follow-up.

| 112 patients screened for participation | 5 patients declined consent |
| 100 patients enrolled | 4 patients did not attend clinic |
| 83 patients underwent surgery | 3 patients not suitable for CPET |
| 83 patients followed up for 28 days | 6 patients decline surgery |
| | 5 patients cancelled by clinician |
| | 5 patients surgery not indicated |
| | 1 patient died before surgery |
| 40 patients developed complications with three deaths |
categories limits the predictive value of the ASA score. This is especially the case because the five ASA categories were collapsed together into two categories to improve prognostic accuracy. We are not aware of any previous evaluation of the predictive value of the ASA score for postoperative complications. The current literature for preoperative prediction of postoperative complications is characterized by a number of common limitations for all the methods described here; most importantly, small sample size, a lack of clinician blinding, and retrospective methodology. None of the published studies has compared all the candidate risk assessment tools in a single large patient cohort.

The strength of the current study is that it is, to our knowledge, the first comparison of CPET, plasma biomarkers, and scoring systems in the evaluation of the risk of postoperative complications before non-cardiac surgery. However, this work also has a number of important limitations. The sample size was small and susceptible to selection bias. Patients were already considered at sufficient risk to require referral to a consultant-led preoperative assessment clinic. Unfortunately, we did not document the frequency of clinician requests to view CPET data or how this may have influenced patient care. The profile of patients who developed MACE appeared to favour prediction of this outcome whereas the poor predictive accuracy for all complications may reflect the wider presence of risk factors for this outcome. Even among those patients at high risk, not every patient will develop complications because this outcome is also influenced by factors which occur during and after surgery. The high event rate for all complications suggests a uniformly high level of risk across all patients which might reduce the opportunity to discriminate between patients in terms of the preoperative characteristics associated with poor postoperative outcome. We did not perform a multivariate analysis because of the small sample size and anticipated low event rates which limit the number of variables which can be included in such models. These problems can only be resolved in a much larger clinical study with a wider distribution of perioperative risk to provide more robust estimates of predictive value and the associated discriminatory thresholds.

**Conclusions**

These pilot data suggest that CPET-derived variables and plasma biomarkers have the potential to predict complications, in particular cardiac complications, within 28 days of surgery.
Table 3  Univariate analysis of preoperative risk assessment markers for all complications and MACE. Data presented as median (IQR), mean (SD), or n (%). RCRI, revised cardiac risk index; BNP, B-type natriuretic peptide; CRP, C-reactive protein; eGFR, estimated glomerular filtration rate; AT, anaerobic threshold.

<table>
<thead>
<tr>
<th></th>
<th>No complications (n = 43)</th>
<th>Complications (n = 40)</th>
<th>P-value</th>
<th>No MACE (n = 74)</th>
<th>MACE (n = 9)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASA I and II</td>
<td>18 (45%)</td>
<td>10 (25%)</td>
<td>0.16</td>
<td>27 (36%)</td>
<td>1 (11%)</td>
<td>0.26</td>
</tr>
<tr>
<td>ASA III and IV</td>
<td>25 (55%)</td>
<td>30 (75%)</td>
<td></td>
<td>47 (56%)</td>
<td>8 (89%)</td>
<td></td>
</tr>
<tr>
<td>RCRI I and II</td>
<td>26 (60%)</td>
<td>23 (58%)</td>
<td>0.83</td>
<td>47 (56%)</td>
<td>2 (33%)</td>
<td>0.03</td>
</tr>
<tr>
<td>RCRI III and IV</td>
<td>17 (40%)</td>
<td>17 (43%)</td>
<td></td>
<td>27 (36%)</td>
<td>7 (78%)</td>
<td></td>
</tr>
<tr>
<td>BNP (pg ml$^{-1}$)</td>
<td>26 (16–71)</td>
<td>48 (21–106)</td>
<td>0.15</td>
<td>28 (17–79)</td>
<td>89 (43–345)</td>
<td>0.0001</td>
</tr>
<tr>
<td>CRP (mg litre$^{-1}$)</td>
<td>4.9 (2.4–12.1)</td>
<td>2.8 (1.7–7.0)</td>
<td>0.28</td>
<td>3.5 (1.9–10.1)</td>
<td>5.0 (2.8–17.1)</td>
<td>0.87</td>
</tr>
<tr>
<td>eGFR (ml min$^{-1}$)</td>
<td>79 (19)</td>
<td>77 (28)</td>
<td>0.59</td>
<td>80 (23)</td>
<td>61 (21)</td>
<td>0.02</td>
</tr>
<tr>
<td>AT (ml min$^{-1}$ kg$^{-1}$)</td>
<td>13.4 (2.9)</td>
<td>12.0 (2.8)</td>
<td>0.03</td>
<td>13.1 (2.9)</td>
<td>10.1 (1.8)</td>
<td>0.006</td>
</tr>
<tr>
<td>VO$_2$ peak (ml min$^{-1}$ kg$^{-1}$)</td>
<td>17.3 (4.5)</td>
<td>14.9 (3.8)</td>
<td>0.02</td>
<td>16.6 (4.3)</td>
<td>12.3 (2.3)</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Fig 2  ROC curves for eGFR and the plasma biomarkers BNP and CRP in the preoperative prediction of all complications and MACE. Area under ROC curves: eGFR 0.58 (0.45–0.71) for complications and 0.72 (0.54–0.90) for MACE, BNP 0.60 (0.48–0.73) for all complications and 0.75 (0.59–0.92) for MACE; CRP 0.61 (0.48–0.74) for all complications and 0.56 (0.36–0.76) for MACE.

Fig 3  ROC curves for the CPET indices AT and VO$_2$ peak in the preoperative prediction of all complications and MACE. Area under ROC curves: AT 0.64 (0.52–0.77) for all complications and 0.83 (0.69–0.96) for MACE; VO$_2$ peak 0.64 (0.52–0.77) for all complications and 0.81 (0.68–0.93) for MACE.
surgery. Large multi-centre studies are now required to confirm the clinical utility of these prognostic tests. This would allow evaluation of each category of prognostic test (scoring systems, plasma biomarkers, and CPET biomarkers) and define how and if multiple tests may be used effectively in an integrated assessment of risk of MACE, overall complications, and postoperative death. Importantly, an integrated approach has the potential for low cost risk screening for all patients allowing both an individual risk assessment for each patient and risk-adjusted evaluation of outcomes using multivariate models, facilitating effective large-scale clinical audit and benchmarking between centres. These pilot data confirm the feasibility of large clinical studies of this type.

Supplementary material
Supplementary material is available at British Journal of Anaesthesia online.

Authors’ contributions
R.M.P. formulated the hypothesis and developed the protocol. Data collection was performed by S.J. and A.S. Biochemical analysis and data interpretation was performed by M.F. and G.O’B. S.J., S. Jh., and R.M.P. performed the data analysis and drafted the manuscript for critical revision by all authors and approval of the final version.

Declaration of interest
None declared.

Funding
This work was supported by departmental research funds. R.M.P. is a National Institute for Health Research (UK) Clinician Scientist and a Royal College of Anaesthetists Career Development Fellow.

References
9 Soklad M. Grading of patients for surgical procedures. Anesthesiology 1991; 2: 281–4
10 Ackland GL, Scollay JM, Parks RW, de Beaux I, Mythen MG. Pre-operative high sensitivity C-reactive protein and postoperative outcome in patients undergoing elective orthopaedic surgery. Anaesthesia 2007; 62: 888–94

27 ATS/ACCP Statement on cardiopulmonary exercise testing. *Am J Respir Crit Care Med* 2003; 167: 211–77


35 Rodseth RN, Padayachee L, Biccard BM. A meta-analysis of the utility of pre-operative brain natriuretic peptide in predicting early and intermediate-term mortality and major adverse cardiac events in vascular surgical patients. *Anaesthesia* 2008; 63: 1226–33


*Handling editor: J. P. Thompson*