British Journal of Anaesthesia **112** (3): 491–7 (2014)
Advance Access publication 22 October 2013 · doi:10.1093/bja/aet346

BJA

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

Editor's key points

- The ability of cardiopulmonary exercise testing (CPET) and plasma biomarkers to predict complications after major surgery is uncertain.
- This small pilot study found that CPET and some biomarkers might be useful in predicting major adverse cardiac events within 28 days.
- However, the study was not powered to detect differences between the scoring systems, biomarkers, or other measures used.
- Larger studies are required.

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 (VO_2 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); VO_2 peak AUROC 0.81 (0.69–0.96)] and poor predictive accuracy for all complications [AT: AUROC 0.64 (0.52–0.77); VO_2 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.[4–6] 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,[27] 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).[8 28] 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 immunoassay (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 the modification of diet in renal disease equation.[29] Creatinine was analysed using a kinetic Jaffe 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.[27] 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 (VO_2), and carbon dioxide production (VCO_2) in expired gas. These measurements allowed estimation of peak oxygen consumption (VO_2 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.[27]

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),[8] 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