Annals of Medicine and Surgery 62 (2021) 1–9

Contents lists available at ScienceDirect

Annals of Medicine and Surgery

journal homepage: www.elsevier.com/locate/amsu

Systematic Review / Meta-analysis

The accuracy of pre-operative (P)-POSSUM scoring and cardiopulmonary exercise testing in predicting morbidity and mortality after pancreatic and liver surgery: A systematic review

J. Dutton [a], M. Zardab [a,*], V.J.F. De Braal [a], D. Hariharan [a], N. MacDonald [b], S. Hallworth [b], R. Hutchins [a], S. Bhattacharya [a], A. Abraham [a], H.M. Kocher [a], V.S. Yip [a]

[a] Barts and London HPB Centre, UK

[b] Department of Anaesthesia, The Royal London Hospital, Barts Health NHS Trust Whitechapel, E1 1BB, UK

ARTICLE INFO

Keywords:
POSSUM
Cardiopulmonary exercise test
Liver and pancreas surgery

ABSTRACT

Background: Cardiopulmonary exercise-testing (CPET) and the (Portsmouth) Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity ((P)-POSSUM) are used as pre-operative risk stratification and audit tools in general surgery, however, both have been demonstrated to have limitations in major hepatopancreatobiliary (HPB) surgery.

Materials and methods: The aim of this review is to determine if CPET and (P)-POSSUM scoring systems accurately predict morbidity and mortality. Eligible articles were identified with an electronic database search. Analysis according to surgery type and tool used was performed.

Results: Twenty-five studies were included in the final review. POSSUM predicted morbidity demonstrated weighted O/E ratios of 0.75(95%CI0.57–0.97) in hepatic surgery and 0.85(95%CI0.8–0.9) in pancreatic surgery. P-POSSUM predicted mortality in pancreatic surgery demonstrated an O/E ratio of 0.75(95%CI0.27–2.13) and 0.94(95%CI0.57–1.55) in hepatic surgery. In both pancreatic and hepatic surgery an anaerobic threshold(AT) of between 9 0.5–11.5 ml/kg/min was predictive of post-operative complications, and in pancreatic surgery ventilatory equivalence of carbon dioxide(VE/VCO2) was predictive of 30-day mortality.

Conclusion: POSSUM demonstrates an overall lack of predictive fit for morbidity, whilst CPET variables provide some predictive power for post-operative outcomes. Development of a new HPB specific risk prediction tool would be beneficial; the combination of parameters from POSSUM and CPET, alongside HPB specific markers could overcome current limitations.

1. Introduction

Hepatopancreaticobiliary (HPB) surgical procedures are often complex providing both a technical challenge to the surgeon and a significant physiological insult to the patient [1]. Despite improvements with both mortality and medical complications, procedure-specific complications remain a significant source of morbidity [1–3].

Appropriate risk-stratification can enable patients to be better informed, improve patient selection and treatment planning; and therefore, overall outcomes. There are, however, limitations to current risk stratification tools.

The application of the POSSUM (Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity) model has been shown to demonstrate a significant lack of fit for predicting both morbidity and mortality after hepatic and pancreatic surgery. Despite adjustments with the logistic regression used in the Portsmouth (P)-POSSUM iteration to better predict mortality previous reviews have recommended further modifications to improve its usefulness [4,5].

Attempts to develop newer risk-stratification have seen some successes with improved prediction of mortality; the surgical outcome risk score (SORT) was modelled on UK national NCEPOD data. A model of 45 risk factors was refined on repeated regression analysis to develop a model comprising six variables. It demonstrated an AUC of 0.91 in predicting 30 day mortality for a general surgical cohort, though there was still some lack of fit when looking only at a HPB cohort (AUC 0.82) [6].

* Corresponding author.
E-mail address: m.zardab@nhs.net (M. Zardab).

https://doi.org/10.1016/j.amsu.2020.12.016
Received 9 November 2020; Received in revised form 8 December 2020; Accepted 10 December 2020
Available online 28 December 2020

More recently, cardiopulmonary exercise testing (CPET) has been proposed as a tool in planning major intra-abdominal surgery [7,8]. CPET provides a global assessment of the cardiopulmonary system's ability to deliver oxygen to tissues under stress, objectively determining the functional reserve [9]. It utilises dynamic pre-operative parameters to provide post-operative predictions; this contrasts the approach used by (P)-POSSUM, which provides post-operative predictions based on static parameters at the pre-operative setting.

The rational to the use of CPET is that patients with a higher cardiopulmonary reserve will be better able to compensate and achieve sufficient tissue oxygen delivery post-operatively, thus will recover quicker with a lower risk of post-operative complications [9,10]. It is currently used as an adjunct to decision making but it's role hasn't been clarified [11].

An up to date systematic review was conducted to assess the power of both (P)-POSSUM and CPET to predict post-operative outcomes in major hepatobiliary and pancreatic surgery. The aim of the review is to evaluate the value of each tool individually in predicting both morbidity and mortality, then identify variables within CPET that demonstrate significant value in predicting post-operative morbidity. This could potentially be integrated into a new scoring system to enable better risk stratification in HPB surgery.

2. Methods

2.1. Search strategy

This systematic review followed PRISMA guidelines (Fig. 1) and AMSTAR (Assessing the methodological quality of systematic reviews) guidelines. Pubmed, Embase and the Cochrane Library were searched without time limits up to 2019 using pre-determined search words. Boolean Operations (AND, OR) combined "POSSUM", "P-POSSUM" or "CPET" with each of the following: "Hepatobiliary", "Hepatic", "Liver", "Gallbladder", "Pancreas", "pancreatic", "Risk", "Morbidity", "Mortality" and "Surgery". Bibliographies of the included papers were searched to identify additional studies.

2.2. Criteria for inclusion and exclusion

Papers were included if they were: retrospective or prospective cohort studies examining either POSSUM, P-POSSUM or CPET in HPB surgery; reported morbidity and/or mortality outcomes; available in English. Studies reviewing (P)-POSSUM had to provide sufficient data with quantification of observed and expected events. For studies reviewing CPET, inclusion required there to be adequate enumeration of the CPET variables of interest, whether this was AT or VE/VCO2. All studies had to provide an adequate quantification of outcomes. Studies in abstract form (n = 1) were included if they contained sufficient data.

Studies were excluded if they were: case reports, review articles or other non-original research; included non-major or transplant surgery; included inadequate data to make reasonable comparison. When there was suspicion of studies containing duplicate data, the most recent data was included.

2.3. Review procedure

Two reviewers (VDB and JD) independently screened titles and abstracts; full-text articles were then assessed for eligibility. Any disagreements were resolved by a third independent reviewer (VSY).

The following information was extracted from the papers: author, year of publication, cohort size, study setting, operation type, CPET and (P)-POSSUM variables, post-operative outcome measures including morbidity, as defined by all definitions, mortality and length of stay, and the statistics used to assess the accuracy of each model including measures of significance.

2.4. Synthesis of results

Due to the heterogeneity of the included studies, the individual risk assessment tools were analysed separately. The primary outcome measures were 30-day mortality rates, and morbidity, based on all definitions. The secondary outcome measure was length of hospital stay (LOS).

The accuracy of POSSUM in predicting morbidity and mortality was assessed with observed event to expected event (O/E) ratios. The

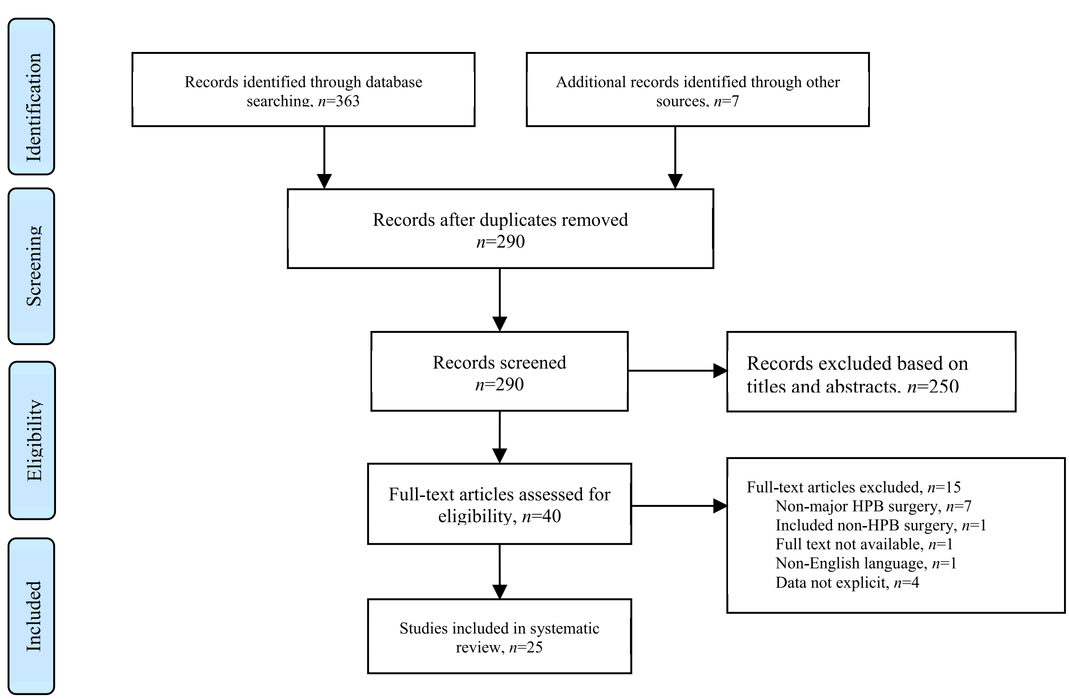

Fig. 1. Prisma flow diagram.

weighted observed to expected (O/E) ratios with 95% confidence interval (CI) were calculated using random effects modelling. All studies were weighted with regards to sample size, regardless of other variables such as definition of morbidity or mortality. Analysis included both the POSSUM and P-POSSUM models. An O/E ratio <1 demonstrates model overprediction whilst >1 implies model underprediction of events. Subanalysis for pancreatic and liver surgery was performed (Supplementary Fig. 1). Statistical analysis was performed using *Review Manager (RevMan) Version 5.3.* (Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014).

There was marked heterogeneity in the CPET studies included and consequently a pooled analysis was not feasible, instead descriptive analysis according to hepatic or pancreatic surgery was performed.

3. Results

Three-hundred and seventy titles were identified using the search strategy outlined above. After the removal of duplicates (80) and screening of titles and full text-articles (290), twenty-five full text articles were included in this systematic review. (Fig. 1).

Sixteen studies reviewed the use of (P)-POSSUM scores in predicting post-operative morbidity (n = 12 studies), and mortality (n = 11 studies) after major hepatopancreatobiliary surgery (Tables 1–3). Four studies reported morbidity based on the original POSSUM definitions [18,19,22,23]. One referred to the International Study Group of Pancreatic Surgery (ISGPS) definitions of complications after pancreatectomy [17]. Four studies identified morbidity according to the Clavien-Dindo (CD) classification [15,16,20,21], and three used an arbitrary list of complications [12–14].

Nine papers reviewed CPET in predicting post-operative morbidity (n = 6) and mortality (n = 5) in HPB surgery; study characteristics are summarized in Table 4. The anaerobic threshold (AT) and ventilatory equivalent of CO_2 (VE/VCO_2) were the most commonly reported variables reviewed with respect to outcome measures including, morbidity, by any definition, mortality, and length of hospital stay (LOS). All studies analysed reported data retrospectively (see Table 5).

3.1. POSSUM

The observed morbidity in major pancreatic surgery ranged from 32.4% to 54.1%, while the POSSUM-predicted morbidity ranged from 36.13% to 76%. The weighted O/E ratio for morbidity for POSSUM in pancreatic surgery was 0.85 (95%CI 0.8–0.9). The observed mortality in major pancreatic surgery ranged from 1.2% to 4%, whilst the POSSUM predicted mortality ranged from 8.7% to 23.7%. The weighted O/E ratio was 0.21 (95%CI 0.09–0.51) (Tables 1a and 2a).

The observed morbidity in major hepatic surgery ranged from 34.7% to 52.0%, while the POSSUM-predicted morbidity ranged from 52.0% to 63.5%. The weighted O/E ratio for morbidity for POSSUM in hepatic surgery was 0.75 (95%CI 0.57–0.97). Whereas the observed mortality in major hepatic surgery ranged from 6.6% to 11.2%, and POSSUM predicted mortality ranged from 9% to 23.7%, producing a weighted O/E ratio of 0.56 (95%CI 0.36–0.88) (Tables 1b and 2b).

3.2. P-POSSUM

The observed mortality for major pancreatic surgery ranged from 1.2% to 7.8%, whilst P-POSSUM predicted mortality ranged from 2.29% to 6.5%. The weighted O/E ratio was 0.75 (95%CI 0.27–2.13); though this was skewed by the larger study by Tamijmarane et al. [12] The observed mortality for major hepatic surgery ranged from 3.95% to 10% and P-POSSUM predicted mortality ranged from 4.2% to 12.9%; the weighted O/E ratio was 0.94 (95%CI 0.57–1.55). Three out of the four papers demonstrated equivocal fit (Tables 3a and 3b).

3.3. CPET in pancreatic surgery

3.3.1. Morbidity

Two studies reported AT as a significant predictor of both morbidity and LOS after PD; pancreatic leaks were analysed in both studies as defined by ISGPS [32,33]. In a cohort of 100 patients undergoing PD or total pancreatectomy (TP), Chandrabalan et al. demonstrated patients with an AT < 10 ml/kg/min had higher incidences of pancreatic fistula, 35.4%, compared to 16% in patients with AT > of 10 $ml^{-1}kg^{-1}min$ (p = 0.028). The same AT was also associated with prolonged LOS; 20 days vs 14 day (p = 0.005), with a hazard ratio of 1.74; [CI: 1.14–2.65] [33].

Ausania and colleagues found an AT ⩽10.1 ml/kg/min to be associated with pancreatic leak [OR of 5.79 (CI 1.62–20.63) (p = 0.007)], with a 45% leak rate compared to 19.2% in patients with an AT of >10.1 ml/kg/min (p = 0.020) [31]. An AT of ⩽ or > 10.1 ml/kg/min also showed a significant difference in predicting any post-operative complication, with 70% compared to 38.5% of patients experiencing a complication (p = 0.013). The same AT was also predictive for length of hospital stay; 29.4 days compared to 17.5 days (p = 0.001) [25]. However, they demonstrated no significant difference in peak VE/VCO_2 between patients who had a pancreatic leak from day 3 post-surgery and those that did not 35.9% vs 37%, (p = 0.409) [32].

3.3.2. Mortality

Junejo et al. demonstrated a VE/VCO_2 of ≥41.0 at AT to be an independent predictor of 30-day mortality in patients undergoing PD [OR 1.35; CI: 1.03–1.77, p = 0.030] and in-hospital mortality [OR 1.26; CI 1.06–1.53] (p = 0.013) [11]. Conversely, when assessing the relationship between AT and mortality, Chandrabalan et al. found no association. [HR 0.77; CI 0.16–3.61] (p = 0.74) [33].

Table 1a
Studies of POSSUM for post-operative morbidity in patients undergoing major Pancreatic surgery.

Study	Year	Country	Patients	Operation	POSSUM		O/E ratio	Comments
					Observed Morbidity (%)	Predicted Morbidity (%)		
Tamijmarane et al. [12]	2008	UK	241	PD	44.8	36.13	1.24	Underpredicts (p < 0.001) *
Khan et al. [13]	2003	UK	50	PD	46	76	0.66	Overpredicts
Dębińska et al. [14]	2011	Poland	65	PD	32.4	64.3	0.5	No association (p = 0.05)[£]
Pratt et al. [15]	2008	US	326	Pancreatic resection	53.1	55.5	0.96	Equivalent (p = 0.206)[$]
Zhang et al. [16]	2009	China	265	PD	39.6	43.8	0.9	Equivalent (p = 0.333)[$]
Knight et al. [17]	2010	UK	99	Pancreatic resection	40.9	47.6	0.86	Poor fit (p = 0.04)*
De Castro et al. [18]	2009	Netherlands	652	PD	50.9	57.8	0.88	Poor fit (P < 0.001)*
Rucket et al. [19]	2014	Germany	697	PD	43.6	58.9	0.74	Overpredicts (p < 0.001)*
Gallacher et al. [20]	2011	UK	81	PD	54.1	63.5	0.86	Overpredicts (p = 0.339)[$]

Table 1b
Studies of POSSUM for post-operative morbidity in patients undergoing major Hepatic surgery.

Study	Year	Country	Patients	Operation	POSSUM		O/E ratio	Comments
					Observed Morbidity (%)	Predicted Morbidity (%)		
Wang et al. [21]	2014	China	100	Cholangiocarcinoma	52	52	1.0	Equivocal (p = 0.488)*
Markus et al. [22]	2005	Germany	190	HPB	34.7	52.1	0.67	Overpredicts (p < 0.01)$
Hellmann et al. [23]	2010	Germany	171	Cholangiocarcinoma	40.9	63.5	0.64	Overpredicts

£ Mann-Whitney *U* test.
$ Chi squared test.
*Goodnes-of-fit analysisx

Table 2a
Studies of POSSUM for post-operative mortality in patients undergoing major Pancreatic surgery.

Study	Year	Mortality	Patients	Operation	POSSUM		O/E ratio	Comments
					Observed Mortality (%)	Predicted Mortality (%)		
Khan et al. [13]	2003	In hospital	50	PD	4	20	0.2	Overpredicts
Pratt et al. [15]	2008	In hospital	326	Pancreatic resection	1.2	16.3	0.07	Overpredicts
Zhang et al. [16]	2009	In hospital	265	PD	3.7	8.7	0.43	Overpredicts (p = 0.018)$
Knight et al. [17]	2010	30-day	99	Pancreatic resection	3	12.5	0.24	Overpredicts (p < 0.0001)$

Table 2b
Studies of POSSUM for post-operative mortality in patients undergoing major Hepatic surgery.

Study	Year	Mortality	Patients	Operation	POSSUM		O/E ratio	Comments
					Observed Mortality (%)	Predicted Mortality (%)		
Wang et al. [21]	2014	In hospital	100	Hilar Cholangio	10	9	1.11	Equivocal (p > 0.05)*
Lam et al. [24]	2004	In hospital	259	Hepatectomy	6.6	14.2	0.46	Overpredicts (p = 0.003)$
Hellmann et al. [23]	2010	In hospital	171	Hilar Cholangio	11.2	23.7	0.47	Overpredicts

£ Mann-Whitney *U* test.
$ Chi squared test.
*Goodnes-of-fit analysis.

Table 3a
Studies of P-POSSUM for post-operative mortality in patients undergoing major Pancreatic surgery.

Study	Year	Mortality	Patients	Operation	P-POSSUM		O/E ratio	Comments
					Observed Mortality (%)	Predicted Mortality (%)		
Tamijmarane et al. [12]	2008	30-day	241	PD	7.8	2.29	3.4	Underpredicts
Khan et al. [13]	2003	In hospital	50	PD	4	6	0.67	Equivocal
Haga et al. [25]	2014	In hospital	231	Pancreatectomy	4.8	6.3	0.76	Overpredicts (p = 0.86)*
Pratt et al. [15]	2008	In hospital	326	Pancreatic resection	1.2	6.5	0.19	Overpredicts
Knight et al. [17]	2010	30-day	99	Pancreatic resection	3	3.8	0.79	Overpredicts (p = 0.09)$

Table 3b
Studies of P-POSSUM for post-operative mortality in patients undergoing major Hepatic surgery.

Study	Year	Mortality	Patients	Operation	P-POSSUM		O/E ratio	Comments
					Observed Mortality (%)	Predicted Mortality (%)		
Wang et al. [21]	2014	In hospital	100	Hilar Cholangio	10	10	1.0	Equivocal
Kocher et al. [26]	2004	In hospital	177	HPB	3.95	4.31	0.92	Equivocal
Lam et al. [24]	2004	In hospital	259	Hepatectomy	6.6	4.2	1.4	Equivocal (p = 0.055)$
Bodea et al. [27]	2018	In hospital	113	HPB	7.09	12.9	0.55	Overpredicts (AUROC 0.61)

£ Mann-Whitney *U* test.
$ Chi squared test.
*Goodnes-of-fit analysis.

3.4. CPET in hepatic surgery

Six studies examined AT, but only one was a randomised controlled trial with the primary outcome measure as surgical morbidities [28]; though it was reported differently amongst them, associations were shown between AT and complications [29,30], length of stay [28,30, 35], and mortality [31,35]. Five of the included papers reported data on VE/VCO2; two papers reported associations between VE/VCO2 and morbidity [31,34].

3.4.1. Morbidity

In this single-centre randomised controlled trial, Dunne et al. found

Table 4
CPET studies.

Study	Year	Country	Patients	Operation	Format	CPET method	Comments
Dunne et al. [28]	2014	UK	197	Hepatectomy	Retrospective	Cycle Ergo	Morbidity, LOS
Kaibori et al. [29]	2013	Japan	61	Hepatectomy, HCC	Retrospective	Cycle Ergo	Mortality
Kasivisvanathan et al. [30]	2015	UK	104	Hepatectomy	Prospective	Cycle Ergo	Morbidity, LOS
Junejo et al. [31]	2012	UK	94	Hepatectomy	Prospective	Cycle Ergo	Mortality, Morbidity
Ausania et al. [32]	2012	UK	124	PD	Prospective	Cycle Ergo	Morbidity, ISGPS fistula
Chandrabalan et al. [33]	2013	UK	100	PD, TP	Retrospective	Cycle Ergo	Mortality, Morbidity, LOS
Junejo et al. [11]	2014	UK	64	PD	Prospective	Cycle Ergo	Mortality
Ulyett et al. [34]	2017	UK	172	Hepatectomy	Prospective	Cycle Ergo	Morbidity (CD)
Snowdon et al. [35]	2013	UK	389	HPB	Prospective	Cycle Ergo	Mortality, LOS

that AT was not independently associated with post-operative CD grade III-IV complications [OR of 1.02 (CI 0.90–1.16, $p = 0.760$)] in 197 patients undergoing major hepatic resection; although it was associated with a reduced LOS, [hazard ratio 2.15 (CI 1.18–3.89) ($p = 0.013$)] [28]. Similarly, Ulyett et al. showed that mean AT was not significant in predicting CD grade following liver resection (12.8 vs 12.5) ($p = 0.84$) [34].

Contrary to this, Kairobi et al. found patients with an AT 11.5 ml/kg/min had a relative risk of 2.73 of complication free survival compared to patients with an AT <11.5 ml/kg/min ($p = 0.0148$) [29] and Kasivisvanathan et al. demonstrated an AT 10.2 ml/kg/min predicts POMS defined morbidity in patients requiring a major hepatic resection [30].

Two studies assessing VE/VCO2 found the measure to be effective in predicting post-operative morbidity. Junejo et al. found that VE/VCO2 of 34·5 or more to be an independent predictor of all post-operative complications (OR 3·97, CI 1·44–10·96, P = 0·008) [31], Ulyett et al. also found the measure to be predictive of CD III–IV complications, OR 1.09 (CI 1.01–1.17, $p = 0.04$) [34].

3.4.2. *Mortality*

Only two papers reviewed mortality reported mortality as one of the outcome measures. Snowden et al. found that an AT < 10 mL/kg/min was predictive of hospital LOS ($\chi 2 = 34.9$; $P < 0.001$) and mortality (OR 0.52, $p = 0.003$); though VE/VCO_2 was not significantly associated ($p = 0.55$) [35]. Similarly, Junejo et al. demonstrated AT <9.9 $ml^{-1}kg^{-1}min$ was predictive for in-patient mortality in patients undergoing hepatectomy (HR1.81, CI 1.04–3.17, $p = 0.038$) [31].

4. Discussion

There is evidence that the POSSUM and its P-POSSUM derivative demonstrate a lack of fit in stratifying patients for HPB surgery [4,5]. Our findings supported that POSSUM is a poor predictor of both morbidity and mortality in pancreatic and hepatic surgical cohorts, whereas the predictive power of P-POSSUM was variable. In hepatic surgery P-POSSUM suggested a good predictive power for assessing mortality risk in hepatic surgery. However, it was less accurate in predicting mortality in pancreatic surgery cohorts. Given this lack of fit they lose their reliability in guiding surgical decision making.

There are a number of reasons why the POSSUM models are inadequate in risk-stratification for HPB surgery. Initially constructed for general surgical populations, they fail to adequately account for the complexity of HPB surgery. The weakness introduced through weighting the degree of surgical complexity is evident in the paper by Tamijmarane et al. the assignment of PD as *major* rather than *major complex* led to a gross underestimation of morbidity as evidenced by the lowest predicted operative score amongst the papers included (13.67 ± 3.42) [12]. Their observed morbidity was well within the accepted range, yet this was the only study to underestimate post-operative morbidity. Furthermore, others have suggested that additional risk factors such as serum bilirubin, and INR are important prognostic factors and should be included in any modification, [18,23]. Yet these are not implemented.

The rationale for reviewing CPET alongside (P)-POSSUM is that, whereas POSSUM includes only static indices, CPET is a dynamic model of a patient's ability to adequately compensate for the physiological stress they may encounter when undergoing major surgery. HPB surgery exposes patients to a significant physiological stress; in addition to the complexity of surgery. The added value provided by CPET, may enable identification of patients who will not tolerate the oxygen supply deficit they are exposed to post-operatively.

This review identified two CPET parameters, the AT and VE/VCO_2, that may provide significant predictive power for post-operative outcomes. In pancreas surgery, three papers looked at post-operative morbidity according to ISGPS defined complications, meaning other non-pancreas surgery specific complications were not accounted for, limiting the meaningfulness of the results. However, an AT of less than 10–10.1 $ml^{-1}kg^{-1}min$ was found to be predictive of grade A-C pancreatic leaks and to be associated with increased hospital length of stay [32,33]. Conversely, Junejo et al. found no association of AT with morbidity [11]. There was significant heterogeneity in the studies reviewing CPET in hepatic surgery, despite this an AT of less than 9.9–11.5 $ml^{-1}kg^{-1}min$ appears likely to be predictive of post-operative morbidity as well increased length of stay. Its predictive value for mortality is less clear. Whilst a higher VE/VCO_2 was demonstrated by two studies to be predictive of morbidity in major hepatic surgery, its value was not supported by others in predicting either morbidity or mortality [28,30,35]. Previous systematic reviews have also reported AT to be a useful parameter for predicting outcomes in non-HPB cohorts, identifying an AT cut-off similar to that reported in this review [36]. The overall conclusion is that VE/VCO_2 is a less reliable predictor of post-operative outcomes.

The heterogeneity of the studies included is a major limitation. For CPET, statistical analysis was inappropriate due to the variability in studies; therefore, only a qualitative approach could be applied. The usefulness of individually reported results is not affected, however, the heterogeneity of the morbidity classification and CPET variables assessed across the available studies severely restricted the ability to draw clear conclusions. To yield future meaningful comparison among studies investigating morbidities and mortalities after HPB procedures, there is a need to standardize complication reporting [37,38]. Another limitation that is difficult to quantify is that of patient selection. CPET is typically done in the pre-operative assessment of patients subjectively deemed to be *high risk,* therefore a population of *low-risk* patients were not included in the CPET studies. It was unclear in the selected studies what threshold was used for these HPB patients, as such this introduces a potential source of bias.

Previous literature has concluded the need for an HPB specific revision of the current risk stratification tools. In other specialties attempts have been made to integrate the use of CPET variables within existing tools; the *Rassi score* is a model that accurately predicts the risk of mortality in Chagas cardiomyopathy, integration of AT increased the accuracy of mortality prediction by 5% [39]. Similarly in major abdominal vascular surgery Thompson et al. demonstrated a significant predictive value of CPET alongside the APACHE II and Detsky scores in predicting both 30-day outcomes and long term survival [40].

This review confirms the limitations of the tools currently being used

Table 5
Summary of the key findings in studies reviewing use of CPET in major HPB surgery.

Study	Year	Country	Patients	Age	Operation	Mortality	Morbidity	LOS	V˙E/V˙CO2 ratio at AT	AT (mL/kg/min)	Notes	
Ausania et al.	2012	UK	124	66 (IQR 37–82)	Pancreaticoduodenectomy	No significant differences observed.	Patients with a lower AT had increased chance of grades A - C pancreatic fistula: AT ⩽10.1 = 45% Vs 19.2% if AT > 10.1 (p = 0.020). [OR = 5.79; CI: 1.62–20.63). (p = 0.007)]. For any post-operative complication, 70% vs 38.5% (p = 0.013).	Patients with a lower AT had increased LOS: AT ⩽10.1 = 29.4 days Vs 17.5 days if AT > 10.1 (p = 0.001).	Peak VeVCO2 not significant for pancreatic leaks. (p = 0.409).	⩽ 10.1 vs > 10.1	Additional factors associated with pancreatic leak were: BMI, jaundice history, pre-operative biliary stent and pancreatic duct size. (p = ⩽ 0.100)	CPET Morbidity & Mortality
Chandrabalan et al.	2013	UK	100	⩽ 65 (n = 47).> 65 (n = 53).	Pancreaticoduodenectomy, Total pancreatectomy	No association with AT [HR 0.77; CI: 0.16–3.61] (p = 0.74)	Greater incidence of ISGPS Grade A-C Pancreatic Fistula when AT < 10: 35.4% v 16% (p = 0.028).Clavien-Dindo grade III- V intra-abdominal abscesses 22.4% vs 7.8% (p = 0.042).	Low AT associated with prolonged LOS: 20 days vs 14 day (p = 0.005). [HR = 1.74; CI: 1.14–2.65].	-	<10 vs ⩾ 10	Patient's less likely to receive adjuvant therapy if low AT. [HR = 6.30; CI: 1.25–31.75] (p = 0.026).	
Junejo et al.	2014	UK	64	64 (IQR 45–80)	Pancreaticoduodenectomy	V˙E/V˙CO2 of ⩾ 41 predicts poor long-term survival [HR 2.05, CI: 1.09–3.86] (p = 0.026), 30 day mortality [OR 1.35; CI: 1.03–1.77]. (p = 0.030) and in-hospital mortality [OR 1.26; CI 1.06–1.53.] (p = 0.013).No significance for AT or VO2 max.	No significant preoperative CPET variable	-	V˙E/V˙CO2 cut off of 41.	-	Neither AT nor V˙E/V˙CO2 at AT were predictive for morbidity or mortality.	
Study	Year	Country	Patients	Age	Operation	Mortality	Morbidity	LOS	V˙E/V˙CO2 ratio at AT	AT (mL/kg/min)	Notes	CPET Morbidityy & Mortality
Dunne et al.	2014	UK	197	70 (64–75)	Hepatectomy	–	HR at AT as predictor of CD 3/4 complication had OR 1.02 (1.0–1.04)	Patients with a higher VO2 L min-1 at AT had increased chances of earlier discharge [hazard ratio 2.15 (CI:	VeVCO2 at AT for all complications OR 1.02 (CI 0.96, 1.08) (p = 0.541) - not significnant	11.5 mean (SD 2.4)VO2 at AT OR 1.02 (CI: .91–1.15) (p = 0.748)	Factor most strongly assocaited with morbidity was performance of major hepatectomy.	

(continued on next page)

Table 5 (*continued*)

Study	Year	Country	Patients	Age	Operation	Mortality	Morbidity	LOS	V˙E/V˙CO2 ratio at AT	AT (mL/kg/min)	Notes
								1.18–3.89), P = 0.013]			
Kaibori et al.	2013	Japan	61	70 (SD = 9)	Hepatectomy, HCC	–	Event free survival had a RR of 2.73 coefficient 1.004 with an SE of 0.412) (p = 0.0148) for an AT of ≥11.5 Vs < 11.5	–	–	<11.5 vs > 11.5	Maintenance of Child-Pugh class between patients with AT VO2 ≥11.5 and < 11.5 ml/min/kg ((p = 0.0464)
Kasivisvanathan et al.	2015	UK	104	65 (IQR 55–70)	Hepatectomy	–	VO2 at AT for predicting morbidity (POMS defined), (OR 1.23, 95% CI 1.02–1.38)	Higher V O2 AT had an increased chance of early discharge [hazard ratio (HR) 1.37, 95% CI 1.13–1.58]	32.4 (29.1–37.2) for POMS >1 (OR 1.02 (.95–1.07)(p = 0.542)		
Junejo et al.	2012	UK	94	71 (24–85)	Hepatectomy	HR 1.81 (CI 1.04–3.17) for mortality in those with AT <9.9 (p = 0.038)	V˙E/V˙CO2 of 34·5 or more at AT to be the only independent predictor (OR 3·97, 95% c.i. 1·44 to 10·96; P = 0·008)	<10.2 then AUC 0.79 (95%CI 0.68–0.86) sensitivity was 83.9% and specificity 52.0%, PPV of 80.6% and NPV of 62.5% for morbidity on day 3 post-opeartive	V˙E/V˙CO2 of >34·5 RR 2·17 (95% c.i. 1·36 to 3·44).	<9.9 Vs > 9.9	
Ulyett et al.	2017	UK	172	69 (22–90)	Hepatectomy	–	VEeqCO2 at AT for developing CD 3/4 OR 1.09 (CI 1.01–1.17) (p = 0.04) (Median VEeqCO2 CD0-II versus CDIII-IV (29.1 vs 31.7) vs 31.7) (p = 0.005)	–	V˙E/V˙CO2 at AT was predictive for CD 3/4 complications. OR 1/09 (CI 1.01–1.17)(p = 0.04).	Mean AT 12.8 (6.4–22.9) versus 12.5 (5.6–23.1) (p = 0.84)	
Snowdon et al.	2013	UK	389	66 (SD = 10.3)	HPB	AT was independent predictor of mortality OR 0.52 (p = 0.003)	–	Patients with an AT < 10 mL/kg/min spent longer in hospital $\chi2$ = 34.9; P < 0.001	V˙E/V˙CO2 at AT was not predictive of mortalitymean 35.4 (6.1)[survivors 35.4 (6.2) versus in-patient mortality 36.3 (4.7)(p = 0.55)]	AT <10 vs > 10 mL/kg/min	

in assessing operative suitability, but it also highlights the need for the integration of dynamic parameters from CPET, such as AT and VE/VCO_2 with indices extracted from (P)-POSSUM, alongside other HPB relevant markers when assessing a patients suitability for major surgery. As clinicians we need to enable patients to make informed decisions regarding their surgical management. Although both tools reviewed are currently in use neither offers sufficiently high predictive power of clinical outcomes to enable a truly informed process. Ultimately, senior clinicians make a decision regarding what they feel is appropriate for an individual patient. Although, this may be informed by years of practice it remains somewhat subjective. A new approach that integrates parameters from both CPET and POSSUM could improve the accuracy of risk stratification and prediction of morbidity and mortality for HPB patients and thus enable a more informed discussion with patients.

5. Conclusion

This review demonstrates the lack of predictive fit of POSSUM and its P-POSSUM derivative alone in major HPB surgery. We also found that the Anaerobic threshold (AT) provided some predictive power for both morbidity and mortality; an AT cut-off value 10–10.1 $ml^{-1}kg^{-1}min$ is likely to be predictive of morbidity after pancreatic surgery, and AT cut-off value of 9.9–11.5 $ml^{-1}kg^{-1}min$ is likely to be predictive of post-hepatic surgery complications. However, there are limitations to both the risk estimation tools evaluated and further prospective research looking at how pre-operative static parameters and dynamic physiological variables can be integrated to enable better risk estimation for a group of patients in whom post-operative morbidity is known to be high.

Ethical approval

N/A.

Sources of funding

None.

Author contribution

The study idea, design and conception as well as clinical expertise and advice was provided by VSY, HMK, AA, SB, SH, NM, DH and RH. VDB and JD performed the screening of titles and abstracts, VSY resolved any disagreements as a third independent reviewer. Writing of the review was performed by JD, MZ, VDB and VSY with additional editing and changes by HMK, AA, SB, SH, NM, DH and RH.

Trial registry number

reviewregistry968.

Guarantor

Dr Johnathan Dutton and Mr Vincent S Yip.

Provenance and peer review

Not commissioned, externally peer-reviewed.
No potential reviewers.

Data statement

All data is accessible on request.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amsu.2020.12.016.

References

[1] C. Ho, Kleeff Jörg, Friess Helmut, Büchler MarkusW, Complications of pancreatic surgery, HPB 7 (2) (2005 Jun) 99–108.
[2] G.A. Gooiker, V.E.P.P. Lemmens, M.G. Besselink, O.R. Busch, B.A. Bonsing, I. Q. Molenaar, et al., Impact of centralization of pancreatic cancer surgery on resection rates and survival: impact of centralization of pancreatic cancer surgery, Br. J. Surg. 101 (8) (2014 Jul) 1000–1005.
[3] K. Takagi, Y. Umeda, R. Yoshida, D. Nobuoka, T. Kuise, T. Fushimi, et al., The outcome of complex hepato-pancreato-biliary surgery for elderly patients: a propensity score matching analysis, Dig. Surg. (2018 Jun 26) 1–8.
[4] H. Wang, T. Chen, H. Wang, Y. Song, X. Li, J. Wang, A systematic review of the Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity and its Portsmouth modification as predictors of post-operative morbidity and mortality in patients undergoing pancreatic surgery, Am. J. Surg. 205 (4) (2013 Apr) 466–472.
[5] T. Chen, H. Wang, H. Wang, Y. Song, X. Li, J. Wang, POSSUM and P-POSSUM as predictors of postoperative morbidity and mortality in patients undergoing hepato-biliary-pancreatic surgery: a meta-analysis, Ann. Surg Oncol. 20 (8) (2013 Aug) 2501–2510.
[6] K.L. Protopapa, J.C. Simpson, N.C.E. Smith, S.R. Moonesinghe, Development and validation of the surgical outcome risk tool (SORT), Br. J. Surg. 101 (13) (2014 Dec) 1774–1783.
[7] C.P. Snowden, J.M. Prentis, H.L. Anderson, D.R. Roberts, D. Randles, M. Renton, et al., Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery, Ann. Surg. 251 (3) (2010 Mar) 535–541.
[8] P.J. Hennis, P.M. Meale, R.A. Hurst, A.F. O'Doherty, J. Otto, M. Kuper, et al., Cardiopulmonary exercise testing predicts postoperative outcome in patients undergoing gastric bypass surgery, Br. J. Anaesth. 109 (4) (2012 Oct) 566–571.
[9] R. Kumar, G. Garcea, Cardiopulmonary exercise testing in hepato-biliary & pancreas cancer surgery – a systematic review: are we any further than walking up a flight of stairs? Int. J. Surg. 52 (2018 Apr) 201–207.
[10] C.W. Lai, G. Minto, C.P. Challand, K.B. Hosie, J.R. Sneyd, S. Creanor, et al., Patients' inability to perform a preoperative cardiopulmonary exercise test or demonstrate an anaerobic threshold is associated with inferior outcomes after major colorectal surgery, Br. J. Anaesth. 111 (4) (2013 Oct) 607–611.
[11] M.A. Junejo, J.M. Mason, A.J. Sheen, A. Bryan, J. Moore, P. Foster, et al., Cardiopulmonary exercise testing for preoperative risk assessment before pancreaticoduodenectomy for cancer, Ann. Surg Oncol. 21 (6) (2014 Jun) 1929–1936.
[12] A. Tamijmarane, C.S. Bhati, D.F. Mirza, S.R. Bramhall, D.A. Mayer, S.J. Wigmore, et al., Application of Portsmouth modification of physiological and operative severity scoring system for enumeration of morbidity and mortality (P-POSSUM) in pancreatic surgery, World J. Surg. Oncol. (2008 Dec) [cited 2019 Feb 21];6(1). Available from, https://wjso.biomedcentral.com/articles/10.1186/1477-7819-6-39.
[13] A.W. Khan, S.R. Shah, A.K. Agarwal, B.R. Davidson, Evaluation of the POSSUM scoring system for comparative audit in pancreatic surgery, Dig. Surg. 20 (6) (2003) 539–545.
[14] I. Dębińska, K. Smolińska, J. Osiniak, P. Paluszkiewicz, The possum scoring system and complete blood count in the prediction of complications after pancreato-duodenal area resections, Pol. Przegl. Chir. 83 (1) (2011 Jan) 10–18.
[15] W. Pratt, S. Joseph, M.P. Callery, C.M. Vollmer, POSSUM accurately predicts morbidity for pancreatic resection, Surgery 143 (1) (2008 Jan) 8–19.
[16] Y. Zhang, L. Fu, Z.-D. Zhang, Z.-J. Li, X.-B. Liu, W.-M. Hu, et al., Evaluation of POSSUM in predicting post-operative morbidity in patients undergoing pancreaticoduodenectomy, J. Int. Med. Res. 37 (6) (2009 Dec) 1859–1867.
[17] B.C. Knight, A. Kausar, M. Manu, B.A. Ammori, D.J. Sherlock, D.A. O'Reilly, Evaluation of surgical outcome scores according to ISGPS definitions in patients undergoing pancreatic resection, Dig. Surg. 27 (5) (2010) 367–374.
[18] S.M.M. de Castro, J.T. Houwert, S.M. Lagarde, J.B. Reitsma, O.R.C. Busch, T.M. van Gulik, et al., Evaluation of POSSUM for patients undergoing pancreatoduodenectomy, World J. Surg. 33 (7) (2009 Jul) 1481–1487.
[19] F. Rückert, M. Kuhn, R. Scharm, H. Endig, S. Kersting, F. Klein, et al., Evaluation of POSSUM for patients undergoing pancreatoduodenectomy, J. Invest. Surg. 27 (6) (2014 Dec) 338–348.
[20] Gallacher, et al., The evaluation of the POSSUM scoring system in the prediction of morbidity following pancreaticoduodenectomy, Br. J. Surg. 98 (S3) (2011 May) 44.
[21] H. Wang, H. Wang, T. Chen, X. Liang, Y. Song, J. Wang, Evaluation of the POSSUM, P-POSSUM and E-PASS scores in the surgical treatment of hilar cholangiocarcinoma, World J. Surg. Oncol. 12 (1) (2014) 191.
[22] P.M. Markus, J. Martell, I. Leister, O. Horstmann, J. Brinker, H. Becker, Predicting postoperative morbidity by clinical assessment, Br. J. Surg. 92 (1) (2005 Jan) 101–106.
[23] S. Hellmann, C. Schafmayer, S. Hinz, B. Schniewind, J. Tepel, D.C. Broering, et al., Evaluation of the POSSUM score in surgical treatment of cholangiocarcinoma, Hepato-Gastroenterology 57 (99–100) (2010 Jun) 403–408.

[24] C.-M. Lam, S.-T. Fan, A.W.-C. Yuen, W.-L. Law, K. Poon, Validation of POSSUM scoring systems for audit of major hepatectomy, Br. J. Surg. 91 (4) (2004 Apr) 450–454.
[25] Y. Haga, Y. Wada, T. Saitoh, H. Takeuchi, K. Ikejiri, M. Ikenaga, Value of general surgical risk models for predicting postoperative morbidity and mortality in pancreatic resections for pancreatobiliary carcinomas, J. Hepato-Biliary-Pancreatic Sci. 21 (8) (2014 Aug) 599–606.
[26] H.M. Kocher, Risk-adjustment in hepatobiliarypancreatic surgery, World J. Gastroenterol. 11 (16) (2005) 2450.
[27] R. Bodea, N.A. Hajjar, A. Bartos, F. Zaharie, F. Graur, C. Iancu, Evaluation of P-POSSUM risk scoring system in prediction of morbidity and mortality after pancreaticoduodenectomy, Chirurgia (Buchar.) 113 (3) (2018) 399.
[28] D.F.J. Dunne, R.P. Jones, D.T. Lythgoe, F.J. Pilkington, D.H. Palmer, H.Z. Malik, et al., Cardiopulmonary exercise testing before liver surgery: exercise Testing before Hepatectomy, J. Surg. Oncol. 110 (4) (2014 Sep) 439–444.
[29] M. Kaibori, M. Ishizaki, K. Matsui, R. Nakatake, T. Sakaguchi, D. Habu, et al., Assessment of preoperative exercise capacity in hepatocellular carcinoma patients with chronic liver injury undergoing hepatectomy [Internet], BMC Gastroenterol. (2013 Dec) [cited 2019 Apr 4];13(1). Available from: https://bmcgastroenterol.biomedcentral.com/articles/10.1186/1471-230X-13-119.
[30] R. Kasivisvanathan, N. Abbassi-Ghadi, A.D.M. McLeod, A. Oliver, R. Rao Baikady, S. Jhanji, et al., Cardiopulmonary exercise testing for predicting postoperative morbidity in patients undergoing hepatic resection surgery, HPB 17 (7) (2015 Jul) 637–643.
[31] M.A. Junejo, J.M. Mason, A.J. Sheen, J. Moore, P. Foster, D. Atkinson, et al., Cardiopulmonary exercise testing for preoperative risk assessment before hepatic resection, Br. J. Surg. 99 (8) (2012 Aug) 1097–1104.
[32] F. Ausania, C.P. Snowden, J.M. Prentis, L.R. Holmes, B.C. Jaques, S.A. White, et al., Effects of low cardiopulmonary reserve on pancreatic leak following pancreaticoduodenectomy, Br. J. Surg. 99 (9) (2012 Sep) 1290–1294.
[33] V.V. Chandrabalan, D.C. McMillan, R. Carter, J. Kinsella, C.J. McKay, C.R. Carter, et al., Pre-operative cardiopulmonary exercise testing predicts adverse post-operative events and non-progression to adjuvant therapy after major pancreatic surgery, HPB 15 (11) (2013 Nov) 899–907.
[34] S. Ulyett, G. Shahtahmassebi, S. Aroori, M.J. Bowles, C.D. Briggs, M.G. Wiggans, et al., Comparison of risk-scoring systems in the prediction of outcome after liver resection [Internet], Perioperat. Med. (2017 Dec) [cited 2019 Apr 4];6(1). Available from: https://perioperativemedicinejournal.biomedcentral.com/articles/10.1186/s13741-017-0073-4.
[35] C.P. Snowden, J. Prentis, B. Jacques, H. Anderson, D. Manas, D. Jones, et al., Cardiorespiratory fitness predicts mortality and hospital length of stay after major elective surgery in older people, Ann. Surg. 257 (6) (2013 Jun) 999–1004.
[36] J. Moran, F. Wilson, E. Guinan, P. McCormick, J. Hussey, J. Moriarty, Role of cardiopulmonary exercise testing as a risk-assessment method in patients undergoing intra-abdominal surgery: a systematic review, Br. J. Anaesth. 116 (2) (2016 Feb) 177–191.
[37] M.E. Egger, J.M. Ohlendorf, C.R. Scoggins, K.M. McMasters, R.C.G. Martin, Assessment of the reporting of quality and outcome measures in hepatic resections: a call for 90-day reporting in all hepatectomy series, HPB 17 (9) (2015 Sep) 839–845.
[38] P.A. Clavien, J. Barkun, M.L. de Oliveira, J.N. Vauthey, D. Dindo, R.D. Schulick, et al., The Clavien-Dindo classification of surgical complications: five-year experience, Ann. Surg. 250 (2) (2009 Aug) 187–196.
[39] RR da Silva, M.S. Reis, B. de B. Pereira, EM do Nascimento, R.C. Pedrosa, Additional value of anaerobic threshold in a general mortality prediction model in a urban patient cohort with Chagas cardiomyopathy, Rev Port Cardiol Engl Ed 36 (12) (2017 Dec) 927–934.
[40] A. Thompson, N. Peters, R. Lovegrove, S. Ledwidge, A. Kitching, T. Magee, et al., Cardiopulmonary exercise testing provides a predictive tool for early and late outcomes in abdominal aortic aneurysm patients, Ann. R. Coll. Surg. Engl. 93 (6) (2011 Sep) 474–481.