Prediction of postoperative mortality in colorectal cancer patients; a comparison of three international scoring systems.

Gaitri Sadadcharam* Niall Fennessy† Aidan T. Manning‡
Kathleen O'Sullivan** Emmet J. Andrews††

*Department of Colorectal Surgery, Cork University Hospital, Wilton, Cork, Ireland, paithiyam@gmail.com
†Department of Colorectal Surgery, Cork University Hospital, Wilton, Cork, Ireland
‡Department of Colorectal Surgery, Cork University Hospital, Wilton, Cork, Ireland
**Department of Mathematical Sciences, University College Cork, Cork, Ireland
††Department of Colorectal Surgery, Cork University Hospital, Wilton, Cork, Ireland
Prediction of postoperative mortality in colorectal cancer patients; a comparison of three international scoring systems.

Gaitri Sadadcharam, Niall Fennessy, Aidan T. Manning, Kathleen O’Sullivan, and Emmet J. Andrews

Abstract

Aims Multiple validated scoring systems exist for the prediction of mortality in patients undergoing surgery for colorectal cancer. This study aims to compare the three most commonly used scoring systems to determine which is most accurate at predicting mortality in colorectal cancer patients undergoing surgery.

Method A retrospective chart review was performed of patients who underwent colorectal cancer surgery at Cork University Hospital to determine mortality 6 months post-operatively. All patients with incomplete data or follow-up were excluded. Three recognised scoring systems were applied to each patient and statistical analysis was performed. The Association Francaise de Chirurgie (AFC), the Colorectal-Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity (Cr-POSSUM) and the Association of Coloproctology of Great Britain and Ireland (ACPGBI) scoring systems were used.

Results A total of 187 patients were assessed. Each of the three scoring systems was predictive of mortality when applied to this cohort (Mann-Whitney U-test p < 0.001) with a significant correlation seen (Spearman’s Correlation co-efficient between 0.6-0.7, p < 0.001). Logistic regression analysis showed that the odds ratios were similar between all three scoring systems (95% CI 1.033-1.132, p=0.002). Backwards regression analysis proved that the AFC score had the highest predictive ability (p=0.001).

Conclusion This study proves that these three scoring systems accurately predict mortality in Irish patients undergoing colorectal cancer surgery therefore allowing for comparisons of mortality rates. In addition, the AFC scoring system is shown to be the most accurate at predicting mortality of the three systems used in this study.

KEYWORDS: Colorectal Cancer; Scoring system; mortality
INTRODUCTION

Clinical audits have for a long time been a standard tool for measuring and improving outcomes\textsuperscript{1,2}. Single outcome measures such as morbidity and mortality are not an accurate reflection of surgical practice as they do not account for patient factors, the complexity of a procedure or the peri-operative support services. With this in mind, a number of scoring systems have been developed to allow for predictions in both specialty and general surgical procedures\textsuperscript{3-13}. These scoring systems not only allow for numerical prediction of mortality and morbidity but also allow for comparative audits with suitable adjustments of patient and operative variables which have a wider application in surgical practice. With regards to prediction of mortality in colorectal cancer resections there are several widely used models including the American Society of Anaesthesiology (ASA) grade, the Acute Physiology and Chronic Health Evaluation II (APACHE II) score, the Physiological & Operative Severity Score for the enUmeration of of Mortality and Morbidity (POSSUM), the Portsmouth POSSUM (P-POSSUM), the colorectal POSSUM (Cr-POSSUM), the Association of Coloproctology of Great Britain and Ireland (ACPGBI) and the Association Française de Chirurgie (AFC) scoring system. The ASA grade is a simple and useful adjunct to informed consent but is an operator dependent tool. The APACHE II system requires invasive monitoring which is not always indicated in colorectal surgery and is too complex for general use. Although the POSSUM model\textsuperscript{3} was specifically designed to predict outcome in surgical patients by taking into account the patients physiology and the nature of the operation it was found to over-predict mortality in the low risk groups and so the P-POSSUM model was developed\textsuperscript{4}. In spite of these changes, both scoring systems were found to either under- or over-estimate the mortality and morbidity figures in colorectal patients\textsuperscript{5-10}. Hence further research was carried out to form the Colorectal POSSUM (Cr-POSSUM) scoring system. This was found to better predict mortality compared to the previous POSSUM systems\textsuperscript{5-7, 11-12} and is a simpler model requiring only six physiological and four operative parameters compared to twelve physiological and six operative parameters in previous POSSUM scoring systems. Table 1 compares the different variables of the Cr-POSSUM, P-POSSUM and the POSSUM scoring systems. The Association of Coloproctology of Great Britain and Ireland developed their own model after the difficulties with the POSSUM derived models became apparent\textsuperscript{13}. This new scoring model required five parameters and like the Cr-POSSUM required a complex mathematical formula (see Figure 1). Finally, Slim et. al. published their own predictive model for both malignant and diverticular diseases. This model has to date been the simplest and most practical requiring only 4 parameters without a pre-requisite mathematical formula as it uses the direct conversion of the presence of ‘parameters’ to a predictive value. Table 2 shows the parameters required for the ACPGBI and AFC models whilst Table 3 presents the conversion table from AFC score to predictive mortality.
These scoring systems have been compared in several studies in different health systems. There has been no previous studies assessing the use of mortality prediction scores in Irish patients or any studies comparing the Cr-POSSUM, ACPGBI and AFC scoring systems.
METHODS

All patients who underwent resection for colorectal adenocarcinoma over a 6 year period (January 2005-December 2010) were retrospectively identified using the hospital inpatient enquiry (HIPE) system. Two hundred and one patients were identified. This included all emergency and elective operations for curative, palliative or diagnostic purposes. Patients for whom mortality scores could not be calculated because of inadequate data or with incomplete follow-up were excluded. The remaining 186 patients were included in the study. Three recognised scoring systems were applied to each patient based on their medical records. These include the ACPGBI, Cr-POSSUM and AFC scoring systems. Parameters for calculating the AFC, Cr-POSSUM and ACPGBI are given in Table 1 and 2. A consensus of all clinical parameters was made by two independent reviewers. The AFC scoring system allowed for direct conversion to a mortality rate (see Table 3). The ACPGBI and Cr-POSSUM scoring systems required complex mathematical equations for their calculation as shown in Figure 1. These scores were calculated for each patient using an online risk prediction calculator (http://www.riskprediction.org.uk). All patients were followed to determine mortality from all causes 6 months post-operatively.

Appropriate statistical analysis was used including the Mann-Whitney U-test, Spearman’s Correlation co-efficient, logistic regression analysis, odds ratios and backwards regression analysis. All statistical analysis was performed on SPSS Version 19.0. A p-value less than 0.05 was accepted for statistical significance.
RESULTS

Following application of the inclusion criteria a total of 186 patients were included in the study. Two patients underwent separate operations for recurrent disease giving a total of 188 operations. There was an almost equal distribution of the sexes with 105 males and 81 females (1.3 males: 1 female). The median age of the population is 72 years. This reflects what is seen with colorectal cancer demographics worldwide. Also in this study there was seven times the elective compared to emergency procedures (i.e. 164 elective and 24 emergency procedures). The distribution of the Duke’s classification seen in this study is detailed in Figure 2.

On examining the patient’s ASA status, the majority (64%) were either healthy participants (ASA I) or suffered only mild disease (ASA II). A third had an ASA III score (61 patients) and the other 6 patients had a score of ASA IV-V. Only eleven patients had moderate to severe heart failure with eight patients suffering with moderate disease. Twenty-seven patients were found to have at least one neurological disorder with the majority presenting with a history of cerebrovascular accidents (71%). Only a small proportion of patients gave a history of greater than 10% weight loss in the six months prior to their surgery (n=42, 22%).

With regards to high risk features at presentation a large number of patients presented with a haemoglobin (Hb) of less than 9.9g/dL (n=46/188, 24.5%). Five patients also presented with severe dehydration with a urea of more than 15mmol/L. Only a small number of patients presented with either a tachycardia of more than 120 beats per minute (bpm) or a bradycardia of less than 40 bpm (n=5/188, 2.6%). In terms of contamination, the vast majority had no contamination but six patients had localised purulent material intra-operatively whilst 11 had gross contamination.

Of the 186 patients, 19 died within the first six months of surgery giving an overall observed mortality rate of 10.2%. It is interesting to note that only one patient in this mortality group had multiple high risk features. Of the patients that died, only two had gross contamination or a Hb ≤9.9g/dL and one had severe dehydration (Ur≥15mmol/L). No patient presented with severe shock. However, eleven patients were over 81 years of age (57.9%) and 12 patients were female (63.2%).

As the data was skewed, non-parametric tests were used including the Mann-Whitney U test, Spearman’s rank correlation co-efficient, odd ratios, logistic regression and backward logistic regression. Upon analysing the data, the Mann-Whitney U test showed that all three scoring systems were capable of predicting mortality in this study population (p=0.002). The Spearman’s rank correlation coefficient was utilised in order to assess the correlation between
the three scoring systems (see Table 4). When employing the Spearman’s rank correlation a low, moderate and high correlation ratio is indicated by values of ≤ 0.3, 0.3-0.7 and ≥ 0.7 respectively. The coefficient range was between 0.6-0.7 for all three scoring models. Although closely correlated, it was found that the Cr-POSSUM was a slight outlier and did not match as closely as the AFC and ACPGBI (Cr-POSSUM 0.671 compared to AFC and ACPGBI co-efficients of 0.627 and 0.639 respectively, p<0.001).

Logistic regression analysis was used to calculate the odds ratio of the different models to determine the strength of association between the scoring systems. The odds ratios and the associated 95% confidence intervals are tabulated in Table 5. Again, all three models were closely related with the AFC and ACPGBI having the most closely matched predictive abilities (p≤0.002). The CR-POSSUM was again different to the other two systems placing it as an outlier.

Finally, backwards regression analysis was used to determine which scoring system has the highest predictive ability. Surprisingly, the AFC model with its easy application and limited parameters was found to be the most efficient followed by the ACPGBI and lastly CR-POSSUM.
DISCUSSION

Scoring systems have been in common use for the prediction of mortality and for comparative audits. This is the first study to compare three specialised colorectal scoring systems in an Irish population. In Ireland it is common practice to determine a predictive score for colorectal cancer patients using the ASA score which is a simple and operator-dependant tool. If the patient requires an intensive care unit (ICU) admission then an APACHE II score is utilised. However, the use of specialised colorectal scoring systems is still infrequent in Ireland.

In terms of colorectal cancer patients, mortality prediction can be estimated by a number of specialised scoring models including the Cr-POSSUM, ACPGBI and AFC scoring systems. These models have been the primary focus of this study where surprisingly the AFC model was found to be the most accurate although it is the simplest of the scoring systems that requires only four parameters and no complex mathematical calculations. In this study, the patient at greatest risk of death is the female patient over 81 years of age which contradicts findings made by Tan et. al\textsuperscript{14} which show that octogenarians undergoing major colorectal resection had an acceptable peri-operative survival / mortality rate. Also although many patients presented with high risk features particularly a low Hb (n=46/188, 24.5%) only one patient in the mortality group had multiple co-morbidities.

All the models assessed were predictive of mortality but only showed a moderate correlation between themselves. This would be expected considering the variant development of each system and the different parameters amongst them. There are only two common factors measured, age and operative urgency. The ACPGBI and Cr-POSSUM include more physiological parameters whilst the Cr-POSSUM also accounts for peritoneal soiling which is not examined by the other two models. Although all the models performed in a similar manner, throughout the study the Cr-POSSUM was an outlier in its predictive ability in spite of its more rounded approach. This includes its performance in logistic regression analysis which is considered by many to be the gold standard in predicting clinical outcome especially in risk of death predictions\textsuperscript{4,15-21}. The higher predictive ability of the AFC against the ACPGBI was determined using backward regression analysis.

On reviewing the literature it is apparent that there are conflicting views regarding the different models. The AFC was more recently developed and validated in 2006 by Slim\textsuperscript{22} for both colonic diverticular and malignant disease after he determined that the POSSUM model overestimated post-operative deaths in all cases. Alves then re-validated the model for both diseases the following year\textsuperscript{23}. It must be noted however that although Slim followed mortality for a four month period, the main outcome measure of the AFC
model was post-operative in-hospital mortality and that both studies used the P-POSSUM as a comparative measure as it had been validated as a good predictive measure except in elective surgery where it overestimated death twofold. In their study, the AFC was as predictive as the P-POSSUM but had the added advantage of being a simpler clinical tool. Since these two studies there has been little in the literature regarding the AFC particularly research comparing it to other mortality models.

The POSSUM systems as already indicated by Slim had a reputation for over-predicting mortality. Many other researchers have found that the POSSUM scoring system over-predicted mortality and its derivates were also unreliable predictors\cite{4,7,9,11,22,24}. In addition the P-POSSUM was also found to over-predict mortality in all studies except Ferjani\cite{6} where it underestimated the mortality rate. Many of the same researchers\cite{5,6,11,24} found that the Cr-POSSUM gave the closest prediction of mortality compared to the POSSUM and the P-POSSUM. However other studies found that the Cr-POSSUM still over-predicted mortality\cite{7,9} whilst Ramkumar\cite{12} found no difference in the predictive ability of these three related models. Yan et. al\cite{25} found that both the Cr-POSSUM and the ACPGBI were the best discriminators between survivors and non-survivors compared to the POSSUM, the P-POSSUM and the APACHE II. There are a number of hypotheses for this discrepancy in the predictive ability of the POSSUM systems. The first is the lack of calibration of the POSSUM and P-POSSUM systems at the extremes of age and high emergency workload\cite{10}. This has important implication in clinical practice, as consultants with a high emergency workload may seem to underperform when these scoring systems are applied. The second is in the application of the different models. When Copeland\cite{3} first formulated the original POSSUM scoring system he excluded all patients who were admitted for less than 24 hours and all patients in this pioneering study had full bloods and ECG tracings for accurate measurement of the required physiological variables. This is not representative of the true spectrum of colorectal practice as many patients that present acutely may be at higher risk of mortality and performing full bloods and ECG is not in keeping with many hospital policies. This means that many low risk patients that do not require such tests would be given a normal figure for physiological variables in all the POSSUM-derived scores which would exaggerate the mortality rate\cite{26}. Another issue is the time at which the POSSUM and its derivative are applied. The original score required application at time of presentation. Hence, any pre-operative intervention or ambiguity of this kind will distort the results\cite{27,28}. Finally, Teeuwen et. al.\cite{29} found that the POSSUM scores predicted outcomes in patients with diverticular disease or other indications more accurately than for malignancy. He also found that the ACPGBI was superior to the various POSSUM scores in patients who had colorectal malignancy. One other difficulty with the POSSUM scores is that they cannot be applied in the pre-operative period due
to the operative variables and hence are not useful in patient selection particularly in the high-risk groups.

On reviewing our own study a number of limitations were present. As this is a retrospective study we were unable to prospective control for a variety of biases and documentation errors which may result in type I or type II errors. Two independent clinicians reviewed all the data to counter for clinical variability. There is also an apparent selection bias where there was seven times the number of elective cases which means that data from this study may not extrapolate accurately in the emergency cases. This may mean that in the emergency setting or in patients with a particular presentation e.g. peritoneal soiling that the other more complex models e.g. Cr-POSSUM may predict mortality more accurately. As well as this the study had small numbers due to the exclusion of charts with incomplete data. Finally, we did not examine which variable predicted mortality most accurately as this was not one of the end-points of the study. Nonetheless, in this study the AFC not only compared favourably to the other two models but was also shown to have a greater predictive value. This alongside with its easy application and it’s ability to be used in the pre-operative setting makes it an ideal tool in patient selection, informed consent as well as in performing comparative audits.

ACKNOWLEDGEMENTS
None
REFERENCES


