Inter-operator variability in pPOSSUM scores: a note of caution

ABSTRACT

Purpose: Predicting perioperative morbidity and mortality can be achieved by several risk predicting algorithms. In the UK, the National Emergency Laparotomy Audit, mandated for all patients undergoing emergency laparotomy, uses pPOSSUM as its risk prediction tool. However, there is no literature reporting the inter-operator variability in calculating the score. Inter-rater variability was assessed based on 10 real general surgical cases that went on to have an emergency laparotomy.

Methods: Forty clinicians, 10 each of registrars and consultants in anaesthetics and general surgery, were asked to calculate the pPOSSUM based on the clinical information typically available at the time of making the decision to proceed to emergency laparotomy for the same 10 National Emergency Laparotomy Audit cases. All participants were surveyed to assess their understanding and use of the pPOSSUM score.

Results: More than 80% of respondents stated that they use pPOSSUM in daily clinical practice. There was variability in the calculated scores between the groups analysed. Two subgroups were evident: one in which the calculated mean pPOSSUM was similar between participants but did not reflect the true value, and the other which was accurate, but demonstrated high inter-rater variability.

Conclusions: This is the first study to investigate inter-operator variability in pPOSSUM scores. Previous reports on the validity of the tool fail to account for subjective variation. At a time where pPOSSUM has become a routine part of clinical practice, this variability needs to be accounted for and taken into consideration in the decision-making process.

Accurately predicting perioperative mortality and morbidity for patients undergoing major surgery aids clinical decision making and perioperative planning. Multiple risk prediction scores have been validated to help stratify patients' perioperative risk (Oliver et al, 2015). Examples include pPOSSUM (Copeland et al, 1991), APACHE-II (Knaus et al, 1985) and ACS NSQIP (Bilimoria et al, 2013) which are all in use and broadly consider comorbidity, acute physiology and anticipated surgery.

The UK National Emergency Laparotomy Audit first published its results in 2012 (Saunders et al, 2012). It showed that the unadjusted 30-day mortality for all patients undergoing emergency laparotomy was 14.9%. Among multiple service improvement suggestions to attempt to reduce mortality, the National Emergency Laparotomy Audit recommended that all patients undergoing emergency laparotomy have a formal stratification of their perioperative morbidity and mortality to help plan appropriate perioperative care (Saunders et al, 2012).

The Physiological and Operative Severity Score for the enumeration of Mortality and morbidity (POSSUM) score was first described in 1991 (Copeland et al, 1991) and was subsequently modified in Portsmouth (pPOSSUM) to correct for an over-estimation in mortality, particularly in low-risk patients (Whiteley et al, 1996; Prytherch et al, 1998). The National Emergency Laparotomy Audit adopted the pPOSSUM score, although a National Emergency Laparotomy Audit risk predictor is currently under development.

The National Emergency Laparotomy Audit became a mandatory National Quality Improvement Project audit for all patients undergoing emergency laparotomy from 2012 (National Emergency Laparotomy Audit, 2015; Healthcare Quality Improvement Partnership, 2019). As such, all patients planned for emergency laparotomy must comply with the National Emergency Laparotomy Audit and therefore must undergo pPOSSUM scoring. Trusts who fail to comply with these standards can suffer financial penalties (Healthcare Quality Improvement Partnership, 2018). Despite the universal application of pPOSSUM in emergency surgical practice, there is little formal training in or assessment of understanding of the score.

The pPOSSUM score is calculated from 18 separate parameters, broadly divided into physiological and operative factors (*Table 1*). It has been extensively validated as an accurate morbidity and mortality prediction tool in the context of gastrointestinal surgical procedures (Pratt et al, 2008; Merad et al, 2012; Findlay et al, 2014) as well as modifications for vascular surgery (Grant et al, 2012), orthopaedic surgery (van Zeeland et al, 2011) and specific elderly cohorts (Tran Ba Loc et al, 2010). The validation, however, relied on the postoperative identification of the operative findings. In clinical practice this luxury is not afforded and clinicians are expected to calculate the pPOSSUM preoperatively, introducing the effects of inexperience, the unknown, potential operator bias and subsequent inaccuracies.

There is yet to be an assessment of the inter-operator variability of the preoperative calculated score based on a specific clinical scenario. In an era of increasing reliance on the pPOSSUM score for preoperative

Mr Gijsbert I van Boxel, Surgical Registrar, Department of General Surgery, Royal Berkshire Hospital, Reading RG1 5AN

Dr Stewart McLure, Core Trainee in Anaesthetics, Department of Anaesthetics, Royal Berkshire Hospital, Reading

Mr Keaton Jones, Surgical Registrar, Department of General Surgery, Oxford University Hospitals Foundation Trust, Oxford

Mr Gregory Jones, Consultant, Department of General Surgery, Royal Berkshire Hospital, Reading

On behalf of The OxSCAR Collaborative

Correspondence to: Mr GI van Boxel (gijs.vanboxel@gmail.com)

© 2019 MA Healthcare Ltd

decision making, it is vital that inter-operator variability is low and the score accurate. The apparent objectivity of the morbidity and mortality presumed by the decision-making clinician (particularly if not calculating the score in person) may potentially cause unintentional ill-informed practice. This is best illustrated with an example. A fit 60-year-old patient, presenting with an acute abdomen who turns out to require a simple band adhesiolysis without resection, will have a predicted mortality of ~2%. However, if the same patient (with the same preoperative physiological parameters) proves to have an ischaemic small bowel segment secondary to an intussuscepting tumour with mesenteric nodal involvement, requiring several procedures and suffering significant blood loss and contamination during the resection, the predicted mortality would be 85%. It stands to reason that 'predicting' these alternative scenarios is dependent on experience, the information available and the interpretation of this information at the time of making the decision to proceed to surgery.

Owing to the success of the National Emergency Laparotomy Audit and its mandatory implementation in all NHS hospitals performing emergency laparotomies, pPOSSUM is now universally used. As such, it may be increasingly relied on in surgical decision making. This assumes that the score is calculated on objective facts, rather than subjective interpretation of potentially circumstantial evidence and the anticipated surgery. This study investigated the inter-operator variability in the pPOSSUM score obtained for a given clinical scenario based on information commonly available at the time of making the decision to perform surgery. The hypothesis was that depending on specialty (anaesthetics *vs* surgery) and experience (consultant *vs* registrar) there will be significant variation in the calculated scores.

Methods

Participant selection

Consultants and registrars in general surgery and anaesthesia in the Health Education Thames Valley deanery were invited to participate in this study. They were provided with a printed pack containing the case studies, pPOSSUM scoring proformas and a questionnaire. These were collected in person from the participant on completion.

Case selection

Ten emergency laparotomy cases were selected from one of the deanery hospitals over 3 months (June–August 2017). These were cross-checked against the National Emergency Laparotomy Audit entry for the trust. Case notes were reviewed for clinical, biochemical and radiological findings to generate a brief and accurate synopsis to reflect the information available to a senior clinician at the time of decision making. These clinical vignettes were reviewed by a consultant general surgeon who participates on the general on-call rota (GJ).

Survey

The survey consisted of two components: a questionnaire and 10 clinical cases. The questionnaire was designed to establish participants' understanding and practice of the pPOSSUM score (*Table 2*). Participants gave consent in writing to use their answers in future publications.

Ten anonymised real cases were selected to represent a range of pathologies, comorbidity and urgency. Each case was purposely presented to reflect the typical information available at the time of making the decision to perform an emergency laparotomy. A basic history of the presenting complaint, past medical and drug history were provided. The examination findings of the registrar or consultant as documented in the notes were provided, as well as observations and blood results at the time of admission. Since almost all patients that undergo emergency laparotomy have a computed tomography scan performed, the full report of this was provided for each case.

The participants were asked to complete all subjective components of the score. These included two physiological parameters (respiratory and cardiac function) and all six operative parameters (operation type, number of procedures, estimated blood loss, estimated contamination, malignancy status and National Confidential Enquiry into Patient Outcome and Death (NCEPOD) urgency level).

Table 1. Physiological and operative components of the pPOSSUM score

Operative factors	Operation type (minor, moderate, major, complex major)
	Number of procedures (1, 2, >2)
	Operative blood loss (<100, 100–500, 500–1000, >1000 ml)
	Contamination (none, minor, local pus, free bowel content)
	Malignancy status (none, primary, primary with nodal, metastatic)
	Confidential Enquiry into Patient Outcome and Death (CEPOD) urgency level (elective urgent or emergency, emergency <2 hours)
Physiological factors	Age
	Systolic blood pressure
	Heart rate
	Glasgow Coma Score
	White blood cell count
	Urea
	Sodium
	Potassium
	Electrocardiogram
	Haemoglobin
	Respiratory function
	Cardiac function

From Copeland et al (1991)

© 2019 MA Healthcare Ltd

Table 2. Questionnaire completed by all participants

Question	Options
What does pPOSSUM stand for?	
Do you calculate pPOSSUM yourself before an emergency laparotomy?	Yes / no
Do you consider pPOSSUM to be a useful tool when preparing for an emergency laparotomy?	Yes / no
Above what mortality value would you consider significant in a decision not to operate?	20–29% 30–39% 40–49% 50–59% 60–69% 70–79% 80–89% >90%
Do you consider morbidity when deciding to operate?	Yes / no
Have you ever used the pPOSSUM score in a discussion with a patient or relative to explain a decision not to operate?	Yes / no
Which areas do you find pPOSSUM useful?	Decision to offer surgery
	Informed consent with patient
	Planning perioperative care
Which of the following staff groups would you consider competent to calculate a pPOSSUM?	Foundation trainee (FY1/FY2)
	Core surgical trainee
	Anaesthetic trainee
	Surgical registrar
	Anaesthetic registrar
	Consultant surgeon
	Consultant anaesthetist

Participants and response rate

Registrars and consultants in general surgery and anaesthetics across five different hospitals were approached as part of a trainee collaborative (OxSCAR) project. Each participant was asked to complete the two physiological and six operative parameters for all 10 cases to allow calculation of the pPOSSUM score. All participants completed all components of each case.

Data analysis

pPOSSUM scores were calculated using the previously published algorithm (Merad et al, 2012). Statistical analysis was performed using GraphPad Prism V6.0. The 'true' pPOSSUM mortality and morbidity values were calculated using the actual operative findings and clinical outcomes. Variability was determined by calculating the standard deviation. Accuracy among the cohort was determined where the true pPOSSUM score was within the standard error of the mean.

Results

Participants and response rates

A total of 45 clinicians was approached and 40 returned completed documents (91% response rate) across six hospital sites, comprising a total of 400 reviewed cases. All forms were completed in full.

Questionnaire

The questionnaire reflected the increasing use and awareness of the pPOSSUM score. This is likely to be a direct result of the mandatory nature of the score for all patients undergoing emergency laparotomy. Ninety per cent of participants were able to include all of the terms 'operative', 'physiological', 'mortality' and 'prediction' in their definition.

Fifty per cent of consultants, both surgical and anaesthetic, stated that they did not personally calculate the score, instead relying on more junior members of the team to provide an absolute value in terms of mortality and morbidity. In contrast, all surgical registrars questioned calculated the score themselves. The majority of those questioned (>80% for all groups) found the pPOSSUM score a useful adjunct in clinical practice for both the predicted mortality and morbidity.

The range given as a 'cut off' mortality value to offer or decline surgery was very variable. In general, registrars were more generous, with an average of 71% predicted mortality (mode >90%). Consultants were less optimistic and averaged a cut off of 60% predicted mortality (mode 40–49%).

Of each group 70–80% stated that they would use the score in discussion with patients and family members. The score was used by 88% of participants in planning perioperative care, 85% used it in taking informed consent and 63% used the score in making a decision to offer surgery.

Overall, 37% of participants considered it appropriate for foundation doctors to calculate the pPOSSUM score, with 80% considering it appropriate for core surgical or anaesthetic trainees to calculate it, 98% for surgical registrars, 95% for anaesthetic registrars, 93% for consultant surgeons and 93% for consultant anaesthetists.

Clinical cases

Following analysis of the pPOSSUM data, two distinct patterns emerged. First, there were cases where the participants' estimation of the operative findings, and therefore the pPOSSUM score, were consistent between specialty and seniority. Although tightly clustered, in both cases the predicted mortality was significantly different from the true value (cases 7 and 10, *Figure 1*). Closer assessment of these cases identified a common feature which was that the operative findings did not reflect the report of the cross-sectional imaging available. Case 7 was reported as 'small bowel obstruction likely secondary to adhesions' where the operative findings were perforated diverticulitis. Case 10 was reported as 'perforated caecum' but the operative findings revealed diverticulitis without perforation requiring a wash out and drain. These cases are thus characterized

© 2019 MA Healthcare Ltd

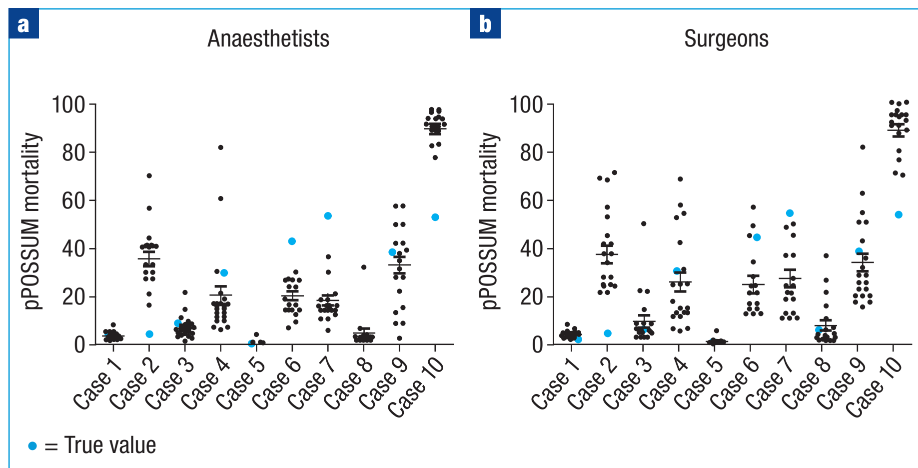


Figure 1. Predicted mortality for cases 1–10 for each participant.

by minimal inter-operator variance, but a mean mortality figure that poorly represents the true value.

In contrast, cases 4 and 9 revealed a different pattern. Both these cases showed a wide variance between predicted mortality in all groups (case 4 range 5.5–82%, case 9 range 8.9–83%). Assessment of the independent parameters showed that the observed inter-operator variation was predominantly driven by differences between estimated blood loss and estimated contamination. Although the calculated mean mortality based on the participants' answers reflected the true value for the actual case, the broad variation meant that these types of cases were at greater risk of significant error in the predicted mortality estimate compared to the true value.

Discussion

To the authors' knowledge this is the first study to assess the inter-operator variability of the pPOSSUM score. Although the internal validity of the score is not addressed or questioned by these findings, these results raise a significant concern regarding the calculation of the score, and its subsequent use in clinical decision making.

The pPOSSUM score is calculated using physiological and operative parameters. Only two physiological parameters (cardiac and respiratory function) could be deemed 'subjective' in what score is assigned to them by the clinician completing the pPOSSUM calculation. The extent to which they can alter the final score is minimal. However, as demonstrated here, the subjective operative components are prone to significant inter-operator variability; they all carry the same weighting (Copeland et al, 1991). This has important implications if the score is used in the clinical decision-making process. Over-estimation of mortality may deny some patients life-saving operative intervention, while, conversely, under-estimation may prevent appropriate perioperative care planning. Assessment of those factors that drive inter-operator variation has shown 'estimated blood loss' and 'predicted contamination' to be the most significant contributors.

This study highlighted two particular groups of patients as noteworthy:

1. Those cases prone to high inter-operator variance in the pPOSSUM calculation, but an accurate mean estimation of the true pPOSSUM
2. Those with minimal inter-operator variance, but an inaccurate mean estimation of the true score.

The second theme identified significant inaccuracy but little variance among respondents. Interestingly, the history, clinical examination and biochemistry results in both cases were in concordance with the ultimate diagnosis, yet the computed tomography report, mistakenly, identified an alternative pathology. These data show that despite ample clinical information there appears to be an over-reliance on the computed tomography findings. Clearly, if decision making was based on the preoperative pPOSSUM then case 10 in particular would have been considered too high risk (predicted mean mortality 91%) when the actual value was 53%. There is a danger that these types of cases may be refused surgery on the grounds of a misleading preoperative pPOSSUM. Equally, but to a less serious degree, is case 7, where surgery could have been offered on the grounds of a reassuring mean predicted mortality (22%), when the true value was 54%.

Thirty-eight per cent of all participants, including 70% of consultant anaesthetists, felt that a foundation doctor was sufficiently competent to calculate a pPOSSUM score and by implication predict the operative findings including contamination, blood loss and complexity regarding resection or anastomosis and laparostomy. This is in the context of 63% of participants stating that they use the score to aid clinical decisions making. Seventy-eight per cent of participants routinely used the score in discussions with patients and relatives in the context of the decision not to operate when the predicted mortality was deemed too high.

The authors certainly do not discourage the use of risk predictors in clinical practice as these have proven very helpful adjuncts in day-to-day practice (e.g. the Rockall score in upper gastrointestinal bleeds (Rockall et al 1996), or the Glasgow Imrie score for acute pancreatitis (Blamey et al, 1984)), but this study illustrates a previously unreported, but clinically significant, caveat to the pPOSSUM score.

The authors would propose that, in the context of preoperative decision making in part on the basis of the pPOSSUM score, the individual operative components, particularly the estimated expected blood loss and contamination, are discussed with the clinician in charge of care in order to decide what the most appropriate predictive score will be. Furthermore, imaging should be discussed specifically and, ideally, images should be reviewed by a senior surgeon (preferably with the reporting radiologist) to corroborate the history, clinical findings, biochemistry, imaging and potential operative approach. This type of multidisciplinary approach is usually the standard of care, but clinicians should be aware that risk predicting scores are no replacement for careful and collaborative clinical practice.

Conclusions

Although the pPOSSUM score is a validated tool in predicting perioperative morbidity and mortality it heavily relies on predicting intraoperative findings and, to a degree, the postoperative course of the

© 2019 MA Healthcare Ltd

patient. This study shows that the inter-operator variability, based on the same clinical information, is wide. Furthermore, an over-reliance on cross-sectional imaging is apparent. In the context of the results of this survey, which show a significant reliance by clinicians on the pPOSSUM score both in their decision to operate and council the patient, the inherent subjectivity of the score needs to be considered. **BJHM**

The authors are grateful to the OxSCAR collaborators for help in data collection: Emanuele Gammeri, YanMei Goh, Perveen Vitish-Sharma, Rishi Pandya, Vasillis Athanassoglou, Gabriel Tong, Jessica Tan, Khalid Akbari. The work reported was presented as a poster at AUGIS 2018, Edinburgh, and the abstract was published by BJS online (https://onlinelibrary.wiley.com/toc/13652168/2018/105/S6)

Conflict of interest: none.

Bilimoria KY, Liu Y, Paruch JL, Zhou L, Kmiecik TE, Ko CY, Cohen ME. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. J Am Coll Surg. 2013 Nov;217(5):833-42.e1-3. https://doi.org/10.1016/j.jamcollsurg.2013.07.385

Blamey SL, Imrie CW, O'Neill J, Gilmour WH, Carter DC. Prognostic factors in acute pancreatitis. Gut. 1984 Dec 01;25(12):1340–1346. https://doi.org/10.1136/gut.25.12.1340

Copeland GP, Jones D, Walters M. POSSUM: A scoring system for surgical audit. Br J Surg. 1991 Mar;78(3):355–360. https://doi.org/10.1002/bjs.1800780327

Findlay JM, Gillies RS, Sgromo B, Marshall REK, Middleton MR, Maynard ND. Individual risk modelling for esophagectomy: a systematic review. J Gastrointest Surg. 2014 Aug;18(8):1532–1542. https://doi.org/10.1007/s11605-014-2524-2

Grant SW, Grayson AD, Mitchell DC, McCollum CN. Evaluation of five risk prediction models for elective abdominal aortic aneurysm repair using the UK National Vascular Database. Br J Surg. 2012 May;99(5):673–679. https://doi.org/10.1002/bjs.8731

Healthcare Quality Improvement Partnership. 2018. Guidance on Quality Accounts, the Quality Accounts List, and Healthcare Service Provider Requirements (accessed 15 May 2019). https://www.hqip.org.uk/wp-content/uploads/2018/05/Guidance-on-Quality-Accounts-January-2018.pdf

Healthcare Quality Improvement Partnership. 2019. National Emergency Laparotomy Audit (NELA) (accessed 13 May 2019). https://www.hqip.org.uk/a-z-of-nca/emergency-laparotomy/#.W9HNvS-Q2Rs

Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. 1985;13(10):818–829. https://doi.org/10.1097/00003246-198510000-00009

Merad F, Baron G, Pasquet B et al. Prospective evaluation of in-hospital mortality with the P-POSSUM scoring system in patients undergoing major digestive surgery. World J Surg. 2012 Oct;36(10):2320–2327. https://doi.org/10.1007/s00268-012-1683-0

NELA project team. 2015. First patient report of the National Emergency Laparotomy Audit. London: Royal College of Anaesthetists.

Oliver CM, Walker E, Giannaris S, Grocott MPW, Moonesinghe SR. Risk assessment tools validated for patients undergoing emergency laparotomy: a systematic review. Br J Anaesth. 2015 Dec;115(6):849–860. https://doi.org/10.1093/bja/aev350

Pratt W, Joseph S, Callery MP, Vollmer CM Jr. POSSUM accurately predicts morbidity for pancreatic resection. Surgery. 2008 Jan;143(1):8–19. https://doi.org/10.1016/j.surg.2007.07.035

Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Br J Surg. 1998 Sep;85(9):1217–1220. https://doi.org/10.1046/j.1365-2168.1998.00840.x

Rockall TA, Logan RF, Devlin HB, Northfield TC. Risk assessment after acute upper gastrointestinal haemorrhage. Gut. 1996 Mar 01;38(3):316–321. https://doi.org/10.1136/gut.38.3.316

Saunders DI, Murray D, Pichel AC, Varley S, Peden CJ; UK Emergency Laparotomy Network. Variations in mortality after emergency laparotomy: the first report of the UK Emergency Laparotomy Network. Br J Anaesth. 2012 Sep;109(3):368–375. https://doi.org/10.1093/bja/aes165

Tran Ba Loc P, Tezenas du Montcel S, Duron JJ et al. Elderly POSSUM, a dedicated score for prediction of mortality and morbidity after major colorectal surgery in older patients. Br J Surg. 2010 Mar;97(3):396–403. https://doi.org/10.1002/bjs.6903

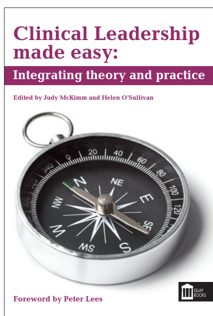
van Zeeland MLP, Genovesi IPO, Mulder JWR, Strating PR, Glas AS, Engel AF. POSSUM predicts hospital mortality and long-term survival in patients with hip fractures. J Trauma Inj Infect Crit Care. 2011 Apr;70(4):E67–E72. https://doi.org/10.1097/TA.0b013e3181edbf7a

Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG. An evaluation of the POSSUM surgical scoring system. Br J Surg. 1996 Jun;83(6):812–815. https://doi.org/10.1002/bjs.1800830628

KEY POINTS

- Interoperator variability in calculating the pPOSSUM score based on the same clinical information is wide.
- The result of cross-sectional imaging is over-relied on when calculating the pPOSSUM score.
- Since the introduction of the National Emergency Laparotomy Audit the majority of clinicians routinely use pPOSSUM in decision making and discussing these decisions with patients and relatives.

Clinical Leadership made easy: Integrating theory and practice



Edited by Judy McKimm and Helen O'Sullivan. Foreword by Peter Lees

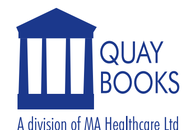
No leadership is 'easy', but this book should help people to understand why and how good and poor leadership exists and how they can improve their own practice, so should be of interest to anyone engaged in or interested in medical leadership.

It is not a 'how to' of leadership activities or development – instead, it poses some key questions about leadership and provide some theoretical foundations for those who wish to develop a deeper understanding of key issues and challenges for medical leaders. The book comprises four sections which look at: Leadership in context, Leadership at all levels, Leadership in practice, Leadership development.

The book aims to debunk some of the myths around leadership theory and show that theories and concepts are helpful, there is evidence about what works and why, and most importantly theory gives us explanatory frameworks through which we can make sense of effective and poor leadership and develop understanding and self-insight about our own leadership behaviours.

Clinical Leadersip made easy: Integrating theory and practice: ISBN: 978-1-85642-431-8; publication: 2016; £24.99

To order please call: 01722 716935 or visit our website **www.quaybooks.co.uk**



© 2019 MA Healthcare Ltd