A VALIDATION STUDY WITH CONSEQUENCES

G.M. Woerlee[1]
J.M. Grefkens[2]
M.J.S. Bakker[3]
J.A. Jansen[4]

[1] Anesthesiologist retired, formerly attached to the Alrijne Hospital group, Leiderdorp (principal author)
[2] Anesthesiologist, Alrijne Hospital group, Leiderdorp
[3] Anesthesiologist in training, Leiden University Medical Centre, Leiden
[4] Orthopedic surgeon, Alrijne Hospital group, Leiderdorp

Contact information:
G.M. Woerlee
Email: mortalminds@gmail.com

Financial support:
None.

Conflicts of interest:
G.M. Woerlee is a retired anesthesiologist, and the developer and owner of the computer program PANSURAS® (www.pansuras.com)

Abstract
During 2018, a validation study was performed of a new computerized outcome-driven machine-learning perioperative assessment and quality control system PANSURAS (Perioperative ANesthesia & SURgical Assessment System). It revealed significant differences in postoperative outcomes between patients who underwent planned oncological colorectal surgery versus those who underwent elective total hip and knee arthroplasty ($p < 0.0001$), despite these two groups having similar preoperative non-oncological Charlson comorbidity score profiles ($p > 0.05$).

This article discusses possible causes for these differences. Furthermore, it discusses the implications of this study for current preoperative assessment algorithms, as well as the implications for the current and future profile of the anesthesiologist as medical specialist in preoperative assessment, and for perioperative management of patients requiring operative management.

Samenvatting
In 2018 werd een validatiestudie uitgevoerd van een nieuw geautomatiseerd resultaat-gestuurd machine-learning perioperatief evaluatie- en kwaliteitscontrolesysteem PANSURAS (Perioperative ANesthesia & SURgical Assessment System). Resultaten lieten significante verschillen zien in postoperatieve uitkomsten tussen patiënten die geplande oncologische colorectale chirurgie ondergingen versus patiënten die electieve totale heup- of kniearthroplastie ondergingen ($p <0{,}0001$), ondanks vergelijkbare preoperatieve niet-oncologische Charlson comorbiditeitsscore-profielen tussen deze twee groepen ($p> 0{,}05$).

Mogelijke oorzaken voor deze verschillen worden besproken. Verder worden de implicaties van deze studie voor huidige preoperatieve beoordelingsalgoritmes besproken, evenals de implicaties voor het huidige en toekomstige profiel van de anesthesist als medisch specialist in preoperatieve beoordeling en voor de perioperatieve behandeling van patiënten die operatieve behandelingen nodig hebben.

Keywords
Colorectal surgery, hip arthroplasty, knee arthroplasty, PANSURAS, retrospective study, postoperative outcomes, profile anesthesiology, perioperative surgical home.

Introduction
During the latter half of 2018, the authors performed a short validation study of the statistical and machine-learning modules of a new program designed for preoperative assessment, audit, and quality control. The results of this validation study for the program entitled PANSURAS (Perioperative ANesthesia & SURgical Assessment System) were such as to raise questions about current anesthesia practice.

The motivation behind the above computer system was based upon a consideration of the current anesthesiology practice of preoperative assessment, and its place in the totality of perioperative patient management. Traditionally, preoperative assessment has been based upon the implicit premise that more pre-existing comorbidity plus increased surgical severity always correlates with an increased chance of postoperative morbidity[1]. This is the basic premise underlying the ASA classification when first described during 1941[2]. Nonetheless, studies of ASA score rating consistency have long been known to be subject to reveal considerable variation (Owens 1978)[3], but lack of any other usable surgical risk scoring system ensured its continued use. Use of the ASA risk classification also encourages another generally unspoken error, that poor physical condition always correlates with increased perioperative risk regardless of the nature of the operation. For example, that a person with an ASA 3 risk category has an elevated risk of perioperative problems undergoing a subcutaneous lipectomy as well as an esophagectomy. Yet all anesthesiologists and surgeons know from clinical experience that individual postoperative consequences of operative procedures depend upon both the level and type of surgical procedure modified by the physiological condition of the individual patient.

This latter reality is addressed by preoperative assessment systems such as the POSSUM algorithm[4], its variant the P-POSSUM[5], the POSPOM system[6], the American College of Surgeons NSQIP calculator[7] and many others. All these systems include an assessment of the magnitude of surgical trauma, as well as of the underlying physiological condition. The NSQIP calculator further specifies the exact type of surgical stress by using specific operative procedure codes. These systems all reveal that minimal surgical trauma has practically no effect upon even very unhealthy persons.

Nonetheless, these and similar systems all suffer from one or more deficiencies. For example, the POSSUM, P-POSSUM, and POSPOM algorithms suffer from a lack of any structured update[4,5,6] to compensate for differing and changing socioeconomic and medical circumstances, or absence of specificity for a particular disorder or procedure[4,5,6], or they are really only valid for the populations in which they are used[4,5,6,7]. All these factors render these algorithms of dubious accuracy when employed outside the populations and time frames from which they are derived.

The above points raise other generally unspoken aspects of the processes of modern anesthesia practice. Preoperative assessment by anesthesiologists using the current clinical qualitative assessment systems are nearly always based upon data acquired in other countries, often from different operations and perioperative management protocols, and from populations with very different socioeconomic profiles. The anesthesiologist and/or surgeon makes qualitative assessments and predictions based upon such data and personal experience. Moreover, anesthesiologists do not merely make an estimate of the safety of anesthesia, but actually make an estimate of the safety of the planned operation plus anesthesia, together with predictions of possible complications. Curiously, anesthesiologists very seldom study the postoperative courses of patients for feedback related to their estimations of safety and predictions, and almost never request structured feedback on postoperative problems. It is the surgeon/operative specialist who manages postoperative problems as they occur, and he almost never provides anesthesiologists any feedback as to the occurrence of postoperative morbidity. This is the reality of current anesthesia and surgical practice. Operative and anesthesia specialties function in differing "silos", even though these physicians all aim to provide optimal treatment for the same patients.

This realization, together with the aim of providing a fundament for any future "Perioperative Surgical Home", was the stimulus for the development of a new machine-learning pre-, and perioperative surgical risk assessment, quality control and audit system called PANSURAS. PANSURAS is developed and designed as a preoperative predictive system, as well as a perioperative quality control and audit system, for anesthesiologists and surgeons using both expert systems and transparent postoperative outcome feedback driven machine-learning algorithms. This enables continual adaptation of the predictive algorithms to the socioeconomic situation of each location where used, as well as to continually evolving changes in medical practice; to finally achieve real cooperation between anesthesiologists and surgeons, and ultimately achieve the ideal expressed by Copeland et al during 1991:

Obviously no regression equation for assessment of risk of morbidity and mortality should remain static over time; hopefully improvements in surgical management will occur. Logistic regression analysis of the observed mortality and morbidity rates on a hospital, district or regional basis would allow the 10-90 per cent ranges to be updated at regular intervals. The extrapolated score values (both physiological and operative severity scores) of 50 per cent risk of mortality and morbidity may potentially allow comparison between units or hospitals. Indeed comparison of physiological and operative severity scores of patients undergoing similar procedures in different units may be of benefit by highlighting different operative and management practices, and also differing modes of presentation[4].

As mentioned earlier, the purpose of the 2018 study was to test the validity of the statistical and machine-learning modules built into PANSURAS. This was planned as a limited validation study designed to include only two patient groups which literature study revealed had a significant chance of postoperative morbidity.

Methods

Institutional approval was obtained from the Alrijne Hospital, in Leiderdorp, the Netherlands to perform retrospective analyses on data extracted from the electronic health record (EHR) system (Chipsoft Hix®) for patients who had been discharged from postoperative check-ups.

Two patient groups were selected for study. One group consisted of patients undergoing planned laparoscopic assisted oncological colorectal surgery for which international studies revealed a frequency of 17.7% to 24.3% postoperative morbidity[8,9]. The other patient group underwent elective total hip and knee arthroplasty for which international studies showed a frequency of 4.9% to 5.8% postoperative morbidity[10,11].

Data were extracted from the EHR and manually entered into a spreadsheet for a total of 207 patients who underwent laparoscopic assisted oncological colorectal surgery, and from 108 patients who underwent elective hip and knee arthroplasty. Criterium for scoring all postoperative outcomes was any mention of a postoperative outcome in the EHR, as these records did not permit of any refinement of the nature of the outcome. The following outcomes were scored: Intensive Care Unit (ICU) admission during postoperative period, length of stay, sepsis / SIRS, postoperative respiratory complications, postoperative cardiac complications, postoperative renal complications (defined as a postoperative eGFR lower than preoperative eGFR), postoperative wound infection, postoperative wound dehiscence, bowel anastomosis leak. Readmission was defined as readmission within 8 days after discharge.

The ASA score given by the anesthesiologist was noted, and the Charlson comorbidity score[12] was used to calculate a more quantitative totality of comorbidity. Subsequently all patient identifying features in the final raw data were anonymized prior to further analysis and importing into the PANSURAS database for validation.

For analyses of these two different groups of patients with uncertain differences in statistical distributions, we analyzed categorical variables and nominal data with the Pearson's Chi-square test, and continuous data were tested with the Moods test. The reason for this type of analysis in this observational study was the independence of these tests of differences in variance and statistical distribution.

	Laparoscopic colorectal surgery	Hip & knee arthroplasty	p
Demography	N = 207	N = 108	
Gender (female)	103 (49.76%)	75 (69.44%)	P < 0.0001
Age (yr) (mean, sd, median)	69.08(+/-12.37), 70	69.5(+/-8.73), 70	P < 0.05
Anesthesia	N = 205	N = 107	P << 0.0001
- General	191	3	
- General + neuraxial	14	1	
- neuraxial	0	103	
Postoperative in-hospital morbidity	N = 207	N = 108	P << 0.0001
Intensive care admission	26 (12.6%)	1 (0.9%)	
Respiratory problems	24 (11.6%)	0 (0%)	
Cardiac problems	13 (6.3%)	0 (0%)	
Renal problems	18 (8.7%)	13 (12%)	
Sepsis / SIRS	8 (3.9%)	0 (0%)	
Wound infection	15 (7.2%)	0 (0%)	
Wound dehiscence	4 (1.9%)	0 (0%)	
Length of Stay (days)			
-Women (N, mean, sd, median)	100, 6.61(+/-5.14), 5	75, 2.61(+/-1.64), 2	
-Men (N, mean, sd, median)	102, 8.32(+/-7.42), 6	33, 2.39(+/-1.58), 2	
P	P >0.05	P <0.05	
Readmission <8 days post discharge	N = 2	N = 0	

Table 1. *Patient populations and demographic profiles, together with types of surgery, anesthesia and in-hospital outcomes.*

Significance level for baseline variables and multivariable regression analysis were initially tested at $p<0.05$.

Results

Table 1 shows the patient populations studied, their demographic profiles, outcomes and management. Postoperative in-hospital morbidity only included morbidity that was common to both groups, excluding problems unique to the type of surgery such as bowel anastomotic leak, etc.Table 2 shows more detail regarding the age profiles of the two patient groups. This table is included because the differences in postoperative outcome may be a function of age. This table reveals that patients who underwent hip and knee arthroplasty had a narrower age spectrum than those who underwent laparoscopic colorectal surgery ($p = 0.44$).

The preoperative Charlson comorbidity score profiles were defined according to the criteria set out in table 3 of Charlson 1987[12]. Table 3a shows the original Charlson score profiles for the two patient groups. The Charlson score adds a score of "2" for non-disseminated cancer (table 3 in Charlson 1987)[12], so when the score profile for oncological colorectal surgery was moved up two rows to adjust the score for non-disseminated cancer "2" from the total Charlson comorbidity score (see table 3b), we found the comorbidity profiles of the colorectal and arthroplasty patients to be very similar (Chi-square $p = 0.94$).

Table 1 revealed that there were no really significant male-female differences within each of the two patients groups, except that the proportion of females undergoing hip or knee arthroplasty was significantly greater (69.44%) than in the colorectal surgery group (49.76%).

The comparable in-hospital postoperative outcomes for both patient groups as shown in table 1 differed significantly ($p << 0.0001$),

Age range	Colorectal surgery	Hip & knee arthroplasty
20-29	1	0
30-39	5	0
40-49	10	2
50-59	25	13
60-69	56	36
70-79	64	43
80-89	42	14
90-99	4	0
N	**207**	**108**

Table 2.
Age profiles of the two patient groups.

Charlson score	Colorectal surgery	Hip & knee arthroplasty
0	0	71
1	0	20
2	132	11
3	34	6
4	23	0
5	12	0
6	3	0
7	2	0
N	**206**	**108**

Table 3a.
Charlson comorbidity score profiles for both patient groups

Charlson score	Colorectal Surgery	Hip & knee Arthroplasty
0	132	71
1	34	20
2	23	11
3	12	6
4	3	0
5	2	0
6		
7		
N	**206**	**108**

Table 3b.
Very similar non-oncological Charlson comorbidity score profiles in both patient groups after adjusting Charlson scores for non-disseminated cancer in the colorectal surgery group ($p = 0.94$)

despite very similar non-oncological preoperative Charlson comorbidity score profiles ($p > 0.94$).

Discussion of validation study results

The observations revealed by this validation study of outcome differences between patients with equivalent comorbidity profiles undergoing colorectal surgery and major joint arthroplasty display significant differences between the postoperative consequences of different types of surgery. To our knowledge, this is the first time such differences have been so clearly demonstrated for two different procedures for patients with similar preoperative comorbidity profiles. This raises the question how such differences could arise.

The differences may simply be related to the length of stay, because admission duration for the orthopedic patients was much shorter than that for the patients who underwent colorectal surgery. This study therefore cannot answer the question whether the orthopedic patients experienced medical problems (except for ICU admission, SIRS/sepsis) with the same frequency as the colorectal group. Severity of postoperative complications, (if any), experienced at home was certainly not reflected in the readmission rates for the orthopedic patients (see Table 1).

Another possible cause is the gender difference. Most orthopedic patients were female. More study is required to determine any possible gender differences, and whether these manifest in other postoperative outcomes.

A possible significant contributory reason for these differences is the enhanced recovery after surgery (ERAS) policy of the orthopedic surgeons in the Alrijne Hospital. Important factors of the orthopedic ERAS protocol were, and are, an opiate sparing multimodal analgesic policy with paracetamol and naproxen, use of high volume wound infiltration analgesia (150cc ropivacaine 0.2%), and pre-operative administration of tranexamic acid and dexamethasone in order to achieve decreased post-operative hematoma with lower pain scores and less nausea. No wound drains or urinary catheters were used, and all patients were mobilized out of bed on the day of surgery by the physiotherapist.

Differences in physiological responses to surgery under neuraxial versus general anesthesia may have influenced the postoperative outcomes[13]. However, studies reveal that these differences do not translate to less postoperative morbidity, as was clearly illustrated by a review of 191,576 laparoscopic colorectal surgery cases in the USA in which epidural analgesia was used in 4,102 patients[14]. This large review revealed that the use of epidural anesthesia was associated with a higher incidence of postoperative urinary tract infections and a longer hospital stay, but there was no observed reduction in the incidence of postoperative respiratory failure[14]. Patients undergoing hip or knee arthroplasty were operated using very short-acting spinal anesthesia, after which they were mobilized on the same day after undergoing surgery.

It is unclear from this short validation study why the difference between these two types of surgery should occur. Nonetheless, whatever the cause of such type of surgery-related outcome differences, they do imply the possibility of significant potential improvements for future management of preoperative assessment and planning, and these consequences are made very relevant by the current Covid-19 pandemic.

General discussion of implications

The current Covid-19 pandemic not only stimulates new thinking on preoperative screening, but forces changes in the manner of preoperative screening by the introduction of web-based questionnaires, telephone, and video consults. Knowledge that different types of procedures have different consequences for postoperative morbidity is implicit in the current selection criteria of patients suited for ambulant operative procedures[15,16,17,18]. Experience shows that such systems are generally safe and efficient for certain types of patients and patient categories[18]. They also save the hospital money, because fewer personnel is required for preoperative assessment. Nonetheless, telephone, internet, or web-based questionnaire systems introduce no fundamental changes in the methodology, relevance, safety, or quality of preoperative assessment. Patients who require physician preoperative assessment are still assessed in the old clinical methodology, while those who do not need to see a physician for preoperative assessment are assessed with an expert system, which may be enhanced by a machine-learning algorithm. No postoperative outcome-driven machine-learning is applied. Such a reduction in patient - anesthesiologist contacts actually increases the distance between the silos in which anesthesiologists and operative specialists function. The added value is financial savings, but not of economy together with improved quality of care.

Further development of machine-learning preoperative risk assessment systems independent of EHR, basing predictions upon ICD diagnostic and procedural codes with outcome-driven feedback, would enable preoperative risk predictions specifically tailored to specific clinics/hospitals. In the future, such a preoperative assessment system may even be employed to inform specific patients and surgeons of the real local risks of surgery when a surgeon determines an indication for surgery. PANSURAS is designed as such a system, and has an inbuilt system for rapid outcome-driven feedback of complications, quality control, and audit, with outcome-driven machine-learning to regularly adjust predictive algorithms. Such systems enable a true predictive – reactive – proactive feedback loop system[19], eventually enabling optimization of perioperative management to the great benefit of patients, physicians, and cost effectiveness of medical care systems[20]. By so doing, such systems add value to the perioperative process for all stakeholders.

This short validation study and associated discussion reveals several important points for anesthesia practice, as well as for the position of the anesthesiologist as medical specialist in the Netherlands, and also other parts of the world.

- The generally accepted belief among anesthesiologists and operators that more preoperative comorbidity is always associated with more postoperative morbidity[1], requires some rethinking and nuance.

- The results of this study indicate that different operative procedures result in different spectrums of possible postoperative morbidities. This has implications for preoperative assessment and management systems. Continual observation and monitoring with outcome-driven machine-learning preoperative assessment will enable perioperative management and anesthesia protocols to be optimized to take advantage of these differences.

- Anesthesiologists and operative specialists currently function in "silos" using inappropriate data to form opinions, and generally fail to employ structured outcome-driven feedback systems. The short validation study reveals by implication how knowledge of outcomes for specific types of surgery can be used to optimize perioperative protocols.

- The current drive to reduce the number of preoperative visits to anesthesiologist managed preoperative assessment clinics, encourages further separation of the "silos" in which anesthesiologists and operative specialists function. This saves the hospital costs of preoperative assessment, but does not fundamentally change or improve the quality of perioperative medical practice for patients, surgeons or anesthesiologists.

- There is a necessity for structured continual observational systems for clinical practice with outcome-driven machine-learning to formulate money-saving, efficient, and safer evidence-based preoperative assessment systems adding true value to perioperative patient management.

References

1. Wolters U, Wolf T, Stützer H, Schroder T. ASA classification and perioperative variables as predictors of postoperative outcome. British Journal of Anaesthesia 1996; 77: 217-222.
2. Saklad M. Grading of patients for surgical procedures. Anesthesiology 1941; 2: 281-284.
3. Owens WD, Felts JA, Spitznagel EL. ASA Physical Status Classifications: A Study of Consistency of Ratings. Anesthesiology, 1978; 49:239-243.
4. Copeland GP, Jones D, Walters M. POSSUM: A scoring system for surgical audit. British Journal of Surgery 1991; 78: 356-360.
5. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. British Journal of Surgery 1998; 85:1217-1220.
6. Manach Y le, Collins G, Rodseth R, Bihan-Benjamin C le, Biccard B, Riou B, Devereaux PJ, Landais P. Preoperative Score to Predict Postoperative Mortality (POSPOM): Derivation and Validation. Anesthesiology 2016; 124: 570-579.
7. 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. Journal of the American College of Surgeons 2013; 217: 833-842.
8. Kolfschoten NE, Marang-van de Mheen PJ, Wouters MW, Eddes EH, Tollenaar RA, Stijnen T, Kievit J. A combined measure of procedural volume and outcome to assess hospital quality of colorectal cancer surgery, a secondary analysis of clinical audit data. PLoS One. 2014; 9: e88737. doi:10.1371/journal.pone.0088737.
9. Scheer A, Martel G, Moloo H, Sabri E, Poulin EC, Mamazza J, Boushey RP. Laparoscopic colon surgery: does operative time matter? Diseases of the Colon and Rectum 2009; 52: 1746-1752.
10. Belmont PJ, Goodman GP, Waterman BR, Bader JO, Schoenfeld AJ. Thirty-Day Postoperative Complications and Mortality Following Total Knee Arthroplasty: Incidence and Risk Factors Among a National Sample of 15,321 Patients. The Journal of Bone & Joint Surgery 2014; 96: 20-26.
11. Vries LM de, Sturkenboom MC, Verhaar JA, Kingma JH, Stricker BH. Complications after hip arthroplasty and the association with hospital procedure volume. Acta Orthopaedica 2011; 82: 545-552.
12. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. Journal of Chronic Diseases 1987; 40: 373-383.
13. Atia AM, Abdel-Rahman KA. Combined Thoracic Epidural with General Anesthesia vs. General Anesthesia Alone for Major Abdominal Surgery: Anesthetic Requirements and Stress Response. Journal of Anesthesia & Clinical Research 2016; 7:4. DOI: 10.4172/2155-6148.1000616.
14. Halabi WJ, Kang CY, Nguyen VQ, Carmichael JC, Mills S, Stamos MJ, Pigazzi A. Epidural Analgesia in Laparoscopic Colorectal Surgery: A Nationwide Analysis of Use and Outcomes. JAMA Surgery 2014 ;149: 130-136.
15. Chung F, Mezei G, Tong D. Pre-existing medical conditions as predictors of adverse events in day-case surgery. British Journal of Anaesthesia 1999; 83: 262-270.
16. Mathis MR, Naughton NN, Shanks AM, Freundlich RE, Pannucci CJ, Chu Y, Haus J, Morris M, Kheterpal S. Patient selection for day case-eligible surgery: identifying those at high risk for major complications. Anesthesiology 2013; 119: 1310-1321.
17. Moon TS, Joshi GS. Are morbidly obese patients suitable for ambulatory surgery? Current Opinion in Anesthesiology 2016; 29: 141-145.
18. Olson RP, Dhakal IB. Day of surgery cancellation rate after preoperative telephone nurse screening or comprehensive optimization visit. Perioperative Medicine 2015; 4: 12. DOI 10.1186/s13741-015-0022-z.
19. Gambus PL, Jaramillo S. Machine-learning in anaesthesia: reactive, proactive ... predictive! British Journal of Anaesthesia 2019; 123: 401-403.
20. Dean HF, Carter F, Francis NK. Modern perioperative medicine – past, present, and future. Innovative Surgical Sciences 2019; 4: 123-131.