comparison of the transverse and longitudinal techniques for ultrasound-guided identification of the cricothyroid membrane in morbidly obese subjects. *Anaesthesia* 2016; **71**: 675–83

20. Airway Management for Anaesthesiologists. *Ultrasound guided marking of the cricothyroid membrane and needle cricothyrotomy*. Available from: https://airwaymanagement.dk/ultrasound_needle_cricothyrotomy. [Accessed 8 September 2020]
21. Oliveira KF, Arzola C, Ye XY, Clivatti J, Siddiqui N, You-Ten KE. Determining the amount of training needed for competency of anesthesia trainees in ultrasonographic identification of the cricothyroid membrane. *BMC Anesthesiol* 2017; **17**: 74
22. Airway Management for Anaesthesiologists. *Transverse "thyroid-airline-cricoid-airline" approach for identification of the cricothyroid membrane*. Available from: http://airwaymanagement.dk/taca. [Accessed 10 September 2020]
23. Airway Management for Anaesthesiologists. *Longitudinal, "string of pearls" approach to identification of the cricothyroid membrane*. Available from: http://airwaymanagement.dk/pearls. [Accessed 10 September 2020]
24. You-Ten KE, Wong DT, Ye XY, Arzola C, Zand A, Siddiqui N. Practice of ultrasound-guided palpation of neck landmarks improves accuracy of external palpation of the cricothyroid membrane. *Anesth Analg* 2018; **127**: 1377–82
25. Athanassoglou V, Hughes-Jones H, Hadjipavlou G, Teoh WH, Kristensen MS, Vanner R. Depth to the airway lumen at the level of the cricothyroid membrane measured by ultrasound. *Acta Anaesthesiol Scand* 2020; **64**: 48–52

British Journal of Anaesthesia, 126 (1): 27–30 (2021)
doi: 10.1016/j.bja.2020.09.015
Advance Access Publication Date: 21 October 2020

Occam's razor at the sharp end: simplified preoperative risk assessment

Simon J. Davies[1] and Gary Minto[2,*]

[1]Department of Anaesthesia, York Teaching Hospitals NHS Foundation Trust, York, UK and [2]Department of Anaesthesia, University Hospitals Plymouth NHS Trust, Plymouth, UK

*Corresponding author. E-mail: gary.minto@nhs.net

This editorial accompanies: A simplified (modified) Duke Activity Status Index (M-DASI) to characterise functional capacity: a secondary analysis of the Measurement of Exercise Tolerance before Surgery (METS) study by Riedel et al., *Br J Anaesth* 2021:126:181–190, doi: 10.1016/j.bja.2020.06.016

Keywords: anaerobic threshold; cardiopulmonary exercise test; Duke Activity Status Index; maximal oxygen uptake; perioperative outcome; preoperative risk assessment

Prediction of outcome after major surgery is a complex process. Clinicians' subjective assessment does not work.[1] An array of scoring systems, biochemical markers, investigations, and tools to estimate functional capacity are now available to attempt to quantify the likelihood of a good recovery so that this information can be used in shared decision making and perioperative planning.

The Duke Activity Status Index (DASI) questionnaire is a 12-part questionnaire that aims to estimate functional status through questions about activity of daily life,[2] and is advocated by the American College of Cardiology and American Heart Association to be incorporated in a decision tree to alter preoperative investigations and management before major or intermediate risk surgery.[3] In the Measurement of Exercise Tolerance before Surgery (METS) study, one of the largest observational studies to date of risk assessment tools in the perioperative setting, the DASI score was a better predictor than clinicians' subjective assessment, serum brain natriuretic peptide, and some functional capacity variables measured with cardiopulmonary exercise testing (CPET) of death or myocardial infarction within 30 days of elective major surgery.[1]

However, the DASI is not straightforward to administer. The DASI is the aggregate score of all self-reported 'yes' answers, weighted from 1.75 to 8.0, with a range of possible scores between 0 and 58.2, and VO_2 peak can then be estimated from an algorithm.[2] This relative complexity may have limited clinical use of DASI. The search continues for the elusive ideal tool that balances accurate risk prediction with clinical utility.

Occam's razor is the rule that, when faced with several solutions for a problem, the simplest one is usually the best. So the work of Riedel and colleagues[4] published in this issue of

DOI of original article: 10.1016/j.bja.2020.06.016.

the *British Journal of Anaesthesia* to produce a suite of simplified, modified versions of the DASI is intriguing. Using elegant regression analyses on data from 1455 participants in the METS study, Riedel and colleagues[4] sought to determine which domains of the long form DASI best discriminate individuals with an oxygen consumption at anaerobic threshold (AT) >11 ml kg^{-1} min^{-1} and at peak (VO_2 peak) of >16 ml kg $^{-1}$ min^{-1}, which are established thresholds associated with better postoperative outcomes.[5]

The M-DASI-5Q is a derived form of the DASI reduced to just five simple questions, each of which has an equal weighting:

- Are you able to climb a flight of stairs or walk up a hill?
- Are you able to do heavy work around the house (*lifting and moving heavy furniture*)?
- Are you able to do yard work *(raking leaves or pushing a power mower)?*
- Are you able to have sexual relations?
- Are you able to participate in strenuous sports *(swimming, singles tennis, football, basketball or skiing)?*

M-DASI-5Q had similar mathematical correlation as the long form DASI to these markers of functional capacity (AUROC-AT: M-DASI-5Q 0.67 *vs* original 12-question DASI 0.66 and AUROC-VO_2 peak: M-DASI-5Q 0.73 *vs* original 12-question DASI 0.71). The authors tested a few further modified forms: removing the potentially sensitive question about sexual activity (M-DASI-4Q) and adding a dynamic component, and ability to increase HR by more than 58 beats min^{-1} in response to exercise (M-DASI-4Q-HR increment) to overcome possible inaccuracy in a purely self-reported measure of exercise capacity. These simplified versions of the DASI all retained the predictive ability of the original questionnaire.

The authors conclude that M-DASI-5Q (and variants) are a simple and effective screening tool to identify patients who have a low probability of being unfit and hence do not need further testing of functional capacity by preoperative investigations such as CPET. Is this conclusion justified? Riedel and colleagues[4] clearly share our optimistic caution. The discussion section of their paper on the modified DASI includes a detailed analysis of the strengths and limitations of the work.

There are plenty of potential difficulties in the practical application of modified DASI scores. The authors have actually shown the following: a reasonably strong mathematical association between a score (M-DASI-5Q) and some markers of functional capacity (VO_2 and AT) in a large heterogenous cohort of surgical patients. This finding is yet to be externally validated in a different cohort.

How would you apply the M-DASI in clinical practice?

In the case of preoperative risk assessment tools, the study of populations that include many low-risk patients can be especially challenging. A tool that successfully identifies almost all low-risk patients may appear to perform well despite misclassifying some high-risk patients.

The discriminatory ability of a predictive test can be expressed in terms of likelihood ratios, odds ratios, sensitivity, and specificity.[6] However, the key to clinical utility is to pinpoint the 'best' predictive cut point of the score (in this case to discriminate AT>11 ml O_2 kg^{-1} min^{-1}, a VO_2 peak >16 ml O_2 kg^{-1} min^{-1}, or both), usually from a receiver operating characteristic curve, and then to report positive and negative predictive values of that threshold value. This allows us to judge how much the M-DASI-5Q score might influence risk prediction in an individual patient.[7] These measures of discrimination are not explicitly reported in the paper, however there are only six possible scores for the M-DASI-5Q, ranging from 0 to 6.

The bar charts in Riedel and colleague's[4] Supplementary Figures 1 and 2 illustrate the probability of achieving the desired oxygen consumption thresholds in relation to the number of positive responses to the M-DASI-5Q. For example, if a patient reports the ability to perform all five tasks, then the likelihood of achieving AT>11 ml O_2 kg^{-1} min^{-1} and VO_2 peak >16 ml O_2 kg^{-1} min^{-1} is about 75% (95% confidence interval [CI] 69–78) and 86% (95% CI 83–89), respectively. However if M-DASI-5Q score was 4 out of a possible 5, then the likelihood of achieving AT>11 ml O_2 kg^{-1} min^{-1} and VO_2 peak >16 ml O_2 kg^{-1} min^{-1} decreased to 59% (95% CI 53–64) and 76% (95% CI 71–81), respectively. So it is likely that the 'best' pragmatic cut point would be a score of 4 or 5. What is the discriminating ability of such a score? Would clinicians and patients be comfortable to skip CPET in all patients whose M-DASI-5Q score is 5, or perhaps even those in whom it is 4? Would we feel comfortable with a simple test that potentially misclassifies a significant number of patients as being 'fit' (in the functional capacity sense) and therefore at low risk of serious postoperative complications?

How do we proceed once an individual reports their M-DASI-5Q score? How might this information be incorporated into a staged screening strategy? Frequentist statistical analysis of data such as this has only two possibilities; that is to accept or to reject the null hypothesis (in this case that the patient is 'fit'). Whether this individual is 'fit' or not, is in reality, not simply a binary answer defined by a single piece of data. Statistical analysis of the M-DASI-5Q comes with associated alpha and beta errors which convey uncertainty in the background. In real-life clinical decision making, we should neither simply ignore uncertainty nor base our decision on a single data point. What we would actually like to know is the *probability* of the individual being fit, and on the basis of that probability we can decide if they need further testing or not. This approach is used in hierarchical testing, founded in the logic underpinning Bayesian statistics.[8] The Bayesian approach is about belief revision, that is we have underlying knowledge before we administer a test (e.g. we know the average fitness for the population being studied [prior distribution]), which reflects the baseline probability that any particular individual is 'fit'. We can then use new evidence, in this case the M-DASI-5Q (likelihood function), and combine the two to provide an updated belief system (posterior distribution) which reflects the probability of an individual being fit or not, and hence whether further testing is justified. The way forward for research into the true clinical utility of the modified DASI likely involves detailed analysis incorporating Bayesian statistics, not only to convey the overall usefulness of the tool itself, taken along with other available information, but also the clinical significance of the M-DASI score for any particular individual. For a fuller exposition of the rather elusive concepts of the Bayesian theory we recommend Ferreira and colleague's[9] recent review published in the *British Journal of Anaesthesia*.

Why predict a predictor rather than an outcome?

In attempting to estimate functional capacity, the original DASI and the modified versions produced by Riedel and colleagues[4] are essentially predicting a predictor. Clinical outcomes are likely to be of more interest to clinicians and patients than isolated measurements of their functional capacity. Endpoints reported in the METS study included 'death or myocardial infarction' by 30 days (the primary outcome) and incidence of moderate and severe complications in hospital which occurred in 2% and 14%, respectively.[1] Only DASI scores were associated with net reclassification of risk of the primary outcome, whereas only CPET predicted in-hospital moderate or severe complications. It appears that these two predictors are measuring something different, though those that died or suffered myocardial infarction clearly also fit the definition for moderate or severe complications. We suggest that incidence of *all* moderate or severe complications is clinically relevant since they all have a significant bearing on quality of life, healthcare cost and resource utilisation, and long-term patient survival.

Clinical outcomes are not systematically reported by Riedel and colleagues,[4] apart from a statement that only a recalibrated form of the long form DASI and the M-DASI-4Q HR response were significantly predictive of 1-year mortality. By implication, the suite of simplified DASI indices were generally not successful predictors of the various clinical endpoints explored in this study.

An important caveat with any analyses based on the METS dataset is that of spectrum bias (i.e. the study cohort does not necessarily match the perioperative population of interest). Only a quarter of eligible patients approached for the study agreed to participate, possibly those with a willingness to exercise, which is reflected in the relatively high mean AT and peak VO_2 compared with other studies. Further, in 3% of patients, an AT could not be identified, so these patients, generally less fit,[10,11] were excluded, as were several participants who did not undergo surgery for reasons that are unclear but may, in some cases, have been because of concerns about their fitness for surgery, informed by CPET. Clinical outcomes may have been further confounded by modification of perioperative pathways for less fit patients prompted by findings on CPET, bringing their clinical outcomes into alignment with their fitter counterparts, so-called confounding by indication. CPET results in METS were clinician-blinded to mitigate this, however, the protocol permitted unblinding for myocardial ischaemia and other adverse events, which occurred in ~2% of cases.

These multiple sources of bias and confounding are likely to have reduced the incidence of adverse postoperative outcomes in the cohort and weakened their statistical association with any of the predictors tested. This point underlines many of the difficulties inherent in the study of predictor tests in the complex environment of major surgery. METS is a landmark study, of a scale and quality unprecedented in CPET research, and represents an extraordinary collaborative achievement by an international perioperative community, yet the study authors fully acknowledge that their findings are tempered by a host of limitations.[1]

Could DASI in any of its forms really be better than CPET?

It is informative to consider how the 12-item DASI was developed: as a simple tool to estimate peak oxygen consumption.[2] The authors of the index used their clinical judgement to lift questions verbatim from existing tools to represent a broad range of cardiovascular stresses and several dimensions of health status. Weighting was based on the known metabolic cost of each activity. Several of the items are rather similar, raising some doubts about the face validity of DASI. The score was tested on 50 cardiology patients and then validated in a further 50, rather than a general surgical population. The calibration was not perfect: although the original paper reported impressive *P*-values for the correlations between DASI and measured VO_2 peak, there was a substantial scatter around the regression lines.[2] This wide distribution of VO_2 per DASI score is mirrored in the METS study.[1] So, in essence Riedel and colleagues[4] have taken an already rather arbitrarily created tool and simplified it further, in the hope of conferring the DASI with greater utility for adoption into clinical use.

Too simplified a view of CPET?

It is possible that alternative CPET variables are better markers of functional capacity than the ones to which the DASI and biomarkers were compared. In the heart failure literature, where CPET use is long established,[12] ventilatory equivalent for CO_2 at AT (VE/VCO_2), which represents the efficiency of CO_2 transfer in the lungs, may be the single best prognostic marker irrespective of aerobic capacity. Recent work in the perioperative setting suggests that VE/VCO_2 and exercise-induced myocardial dysfunction are better predictors than AT or peak VO_2 of adverse outcomes after major abdominal surgery.[13,14] This is perhaps because reduced measurements of AT or peak VO_2 may, in a substantial proportion of patients, represent only deconditioning or a lack of effort during the test, whilst abnormalities in VE/VCO_2 and oxygen pulse (a surrogate for cardiac stroke volume) always indicate pathology.[14]

The way CPET is reported in clinical research does not reflect how the test is used in everyday clinical practice. In 'real life', greater focus is placed on measurements such as peak wattage, myocardial ischaemia, stroke volume response to exercise, HR at baseline, HR recovery, and measures of respiratory efficiency such as VE/VCO_2 and VE/VO_2.[15] An average exercise test may last ~10 min, collecting data at more than 200 time points, with in excess of 20 variables collected at each, giving more than 4000 measurements. If one considers also how these interact with each other and change over time, the richness of the data increases by several further orders of magnitude. Yet, in studies elucidating risk associated with CPET parameters, only two or three of these data points (e.g. VO_2 at AT or peak) are used to characterise functional capacity. Clearly this makes little sense. Better data usage from CPET, through machine learning techniques for example, may well elicit better risk prediction.

Critics of perioperative CPET may argue that the test has been proved inconclusive in the biggest blinded study to date.[1] However, a fundamental problem with much of the CPET literature, including the METS study, is that the studies investigate statistical associations between individual CPET variables and clinical outcomes. What we really want to know is whether the introduction of a preoperative assessment service that includes CPET into a perioperative pathway produces better outcomes for surgical patients.[16,17]

Implications of this work for clinical practice and future research

It has already been shown that clinicians' subjective assessment has no role in preoperative risk prediction and planning of perioperative pathways. The M-DASI-5Q provides a simple estimation of functional capacity, and may have clinical utility. At this point in time, the M-DASI-5Q has been derived in a reasonably large clinical study. A typical next step would be for the score to be validated to characterise functional capacity in a different cohort of patients. However, as we have suggested above, the strength of the mathematical association between this ordinal score and any particular CPET parameter is of little consequence. What matters is whether the incorporation of this score into perioperative planning makes a difference to important clinical outcomes. We share the belief of Riedel and colleagues[4] that this apparently simple question should now be put to the test in well-designed clinical studies.

Authors' contributions

Wrote, edited, and approved the final version: both authors.

Declarations of interest

The authors declare that they have no conflicts of interest.

References

1. Wijeysundera DN, Pearse RM, Shulman MA, et al. Assessment of functional capacity before major non-cardiac surgery: an international, prospective, cohort study. *Lancet* 2018; **391**: 2631–40
2. Hlatky MA, Boineau RE, Higginbotham MB, et al. A brief self administered questionnaire to determine functional capacity (the Duke Activity Status Index). *Am J Cardiol* 1989; **64**: 651–4
3. Fleisher LA, Fleischmann KE, Auerbach AD, et al. 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014; **130**. 2215e45
4. Riedel B, Li M H-G, Lee CHA, et al. A simplified (modified) Duke Activity Status Index (M-DASI) to characterise functional capacity: a secondary analysis of the Measurement of Exercise Tolerance before Surgery (METS) study. *Br J Anaesth* 2021; **126**: 181–90
5. Rose GA, Davies RG, Davison GW, et al. The cardiopulmonary exercise test grey zone; optimizing fitness stratification by application of critical difference. *Br J Anaesth* 2018; **120**. 1187e94
6. Knottnerus JA, van Weel C, Muris JW. Evaluation of diagnostic procedures. *BMJ* 2002; **324**: 477–80
7. Struthers R, Erasmus P, Holmes K, Warman P, Collingwood A, Sneyd JR. Assessing fitness for surgery: a comparison of questionnaire, incremental shuttle walk, and cardiopulmonary exercise testing in general surgical patients. *Br J Anaesth* 2008; **101**: 774–80
8. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian approaches to clinical trials and health-care evaluation*. New York: John Wiley & Sons; 2004
9. Ferreira D, Barthoulot M, Pottecher J, Torp KD, Diemunsch P, Meyer N. Theory and practical use of Bayesian methods in interpreting clinical trial data: a narrative review. *Br J Anaesth* 2020; **125**: 201–7
10. Lai CW, Minto G, Challand CP, 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* 2013; **111**: 607–11
11. Wilson RJT, Davies SD, Yates D, Redman J, Stone M. Impaired functional capacity is associated with all-cause mortality after major elective intra-abdominal surgery. *Br J Anaesth* 2010; **105**: 297e303
12. Gitt AK, Wasserman K, Kilkowski C, et al. Exercise anaerobic threshold and ventilatory efficiency identify heart failure patients for high risk of early death. *Circulation* 2002; **106**: 3079–84
13. Mann J, Williams M, Wilson J, et al. Exercise-induced myocardial dysfunction detected by cardiopulmonary exercise testing is associated with increased risk of mortality in major oncological colorectal surgery. *Br J Anaesth* 2020; **124**: 473–9
14. Wilson RJT, Yates DRA, Walkington JP, Davies SJ. Ventilatory inefficiency adversely affects outcomes and longer-term survival after planned colorectal cancer surgery. *Br J Anaesth* 2019; **123**: 238–45
15. Levett DZH, Jack S, Swart M, et al. Perioperative cardiopulmonary exercise testing (CPET): consensus clinical guidelines on indications, organization, conduct, and physiological interpretation. *Br J Anaesth* 2018; **120**: 484–500
16. Carlisle J, Swart M, Dawe EJ, Chadwick M. Factors associated with survival after resection of colorectal adenocarcinoma in 314 patients. *Br J Anaesth* 2012; **108**: 430–5
17. Davies RG, Tobin S, Moses T, Appadurai IR, Rose G, Bailey DM. Bowel cancer surgery outcomes and preoperative cardiopulmonary exercise testing: insights from real-world data. *Anaesthesia* 2018; **73**: 1445–6