Guidelines for routine preoperative testing

‘This guidance...was arrived at after careful consideration of the evidence available. Health professionals are expected to take it fully into account when exercising their clinical judgement.’

The National Institute for Clinical Excellence (NICE) released a Guideline in June 2003 for professionals who order routine tests before operations, together with guidance on the subject for patients. In the autumn of 2004, it seems reasonable to ask whether things have changed. Did the NICE Guideline change practice? Has unreasonable variation in preoperative testing now ended? Do junior doctors now always consult the pocket version of the Guideline when uncertain about ordering preoperative tests?

Anaesthetists and other clinicians were deeply involved in producing this Guideline, in contrast with previous efforts to investigate the issue. The Acute Care National Collaborating Centre (NCCAC), within The Royal College of Surgeons worked directly with the Royal Colleges of Anaesthetists, Surgeons, Pathologists, Radiologists, Ophthalmologists, and Obstetricians and Gynaecologists, and consulted all other stakeholders. Patient representatives were involved in later stages of the production process and there was a period of public consultation. At the time of the launch, there was a publicity initiative and materials were distributed across the NHS.

Eleven tests are considered in the NICE Guideline:

(i) Plain chest X-ray.
(ii) Twelve lead resting ECG.
(iii) Full blood count.
(iv) Haemostasis (prothrombin time and activated partial thromboplastin time).
(v) Serum urea, creatinine, and electrolytes.
(vi) Random serum glucose.
(vii) Urine analysis.
(viii) Blood gases.
(ix) Lung function (peak expiratory flow rate, forced vital capacity, and forced expiratory volume).
(x) Pregnancy test.
(xi) Sickle cell haemoglobin test.


NICE had two main motives for becoming involved: variation in the way doctors order preoperative tests, and uncertainty about their value. Preoperative testing varies markedly from hospital to hospital, and clinicians in the same hospital may show extreme variation in their ordering of tests. There is some evidence that these patterns can be modified when they are audited and when local protocols are introduced.2-5

The principal challenge facing anyone attempting a systematic review of preoperative testing is the poor quality of published evidence. In a previous attempt, researchers in Sheffield produced a systematic review focusing on healthy asymptomatic adults,6 and found 70 case series but no controlled trials. In 1997 they concluded that the standard of evidence was too poor to show any good evidence of benefit. Importantly, they also flagged up the equal paucity of evidence of NO benefit. It may be important to anaesthetists weighing up the strengths and weaknesses of the NICE Guideline to know whether, waiting in the wings and just about to report, is the ultimate high-quality randomized multicentre trial of preoperative testing that will, once and for all, supply the missing evidence and render the NICE Guideline obsolete. The Department of Health’s Health Technology Assessment Programme made a valiant attempt in 1998 to commission research to supply the missing evidence. Tenders were invited for a randomized controlled trial (RCT) of preoperative testing strategies. However, it gradually became apparent that such a definitive RCT was unlikely to be carried out.

Difficulties confronting any intending researcher would be formidable including: the large number of tests to be investigated (we narrowed our study down to 11 tests, still a large number of separate questions for a trial to address); the large number of doctors to find who would
agree to alter existing work-up strategies (the design would
have to be a cluster randomized trial, randomized by hospi-
tal, which is organizationally very challenging); and above
all, problems defining and capturing data on outcomes.
Working on this subject for 2 years, we came to believe
that no RCT is going to come rapidly to the rescue.

We did, however, take a second long hard look at the
published literature. We enlarged the scope of the search
for papers to include children, complex surgery, and ASA
grades II and III, but as before, our search for good intelli-
gence was frustrated by the poor quality of published papers.
Sifting through all the previously identified studies and
including 47 more studies than the previous review, we
again encountered many case series but no trials. Perhaps
the most critical failing from the NICE perspective was that
the literature could not help us answer the key question, ‘did
practice change as a result of the test?’ Papers were rarely
explicit about changes of practice, complications, or out-
come of surgery; and few were sufficiently robust or com-
parable to enable worthwhile amalgamation of results.

When the literature evidence base fails us, a next best
option is consensus development, in other words a formal
approach to the harvesting of expert opinion. It is still quite
common for learned medical bodies to publish consensus
statements based on dubious processes, which give consen-
sus development a bad name. For example, a guideline of
sorts can be produced by a group of experts and an audience
meeting for a day or two, listening to part of the evidence and
then just going to press citing references. This is the bottom
end of the guideline scale, which is organizationally very challenging); and above
all, problems defining and capturing data on outcomes.
Working on this subject for 2 years, we came to believe
that no RCT is going to come rapidly to the rescue.

We did, however, take a second long hard look at the
published literature. We enlarged the scope of the search
for papers to include children, complex surgery, and ASA
grades II and III, but as before, our search for good intelli-
gence was frustrated by the poor quality of published papers.
Sifting through all the previously identified studies and
including 47 more studies than the previous review, we
again encountered many case series but no trials. Perhaps
the most critical failing from the NICE perspective was that
the literature could not help us answer the key question, ‘did
practice change as a result of the test?’ Papers were rarely
explicit about changes of practice, complications, or out-
come of surgery; and few were sufficiently robust or com-
parable to enable worthwhile amalgamation of results.

When the literature evidence base fails us, a next best
option is consensus development, in other words a formal
approach to the harvesting of expert opinion. It is still quite
common for learned medical bodies to publish consensus
statements based on dubious processes, which give consen-
sus development a bad name. For example, a guideline of
sorts can be produced by a group of experts and an audience
meeting for a day or two, listening to part of the evidence and
then just going to press citing references. This is the bottom
end of the guideline scale, whilst those present on the
day may passionately believe that they truly considered
everything important and somehow hit on the right answer,
others cannot later follow their intellectual trail, nor is it
clear what emerging knowledge should prompt a review.

Fortunately, more robust approaches to canvassing expert
opinion have been developed over many years in the RAND
organization. We used the RAND modified Nominal
Group Technique, which appears to be a valid approach
to gathering a wide spread of expert opinion. We used two
parallel panels of ‘experts’ in an attempt to increase the
robustness of our findings. In practice, panellists were every-
day practitioners, nominated by colleagues to represent their
speciality professional groups. The panellists were given a
consensus questionnaire and worked in private to rate the
appropriateness of all 11 preoperative tests, using a nine-
point scale. They were prompted for views on appropriaten-
ess of each test for different groups of patients, breaking
down these by age band, three ASA grades (I, II, and III),
presence of three kinds of co-morbidity (cardiovascular,
renal, or pulmonary), and four different grades of surgical
severity (minor to major plus), for which we gave illustrative
examples. In our briefing, we avoided the word ‘routine’ as it
has proved unhelpful, suggesting both habitual repetition
and the absence of co-morbidity.

Our panellists were brought together for a meeting to
explore possible reasons for disagreements and to obtain
an improved consensus. They were shown how the spread
of views across the panels compared with their own response, and then invited to discuss topics for which
there was least consensus. Finally, they again worked in
private to reconsider the appropriateness of tests and revise
ratings if they wished to.

We took the final answers given by panellists as represent-
ing best practice. The NICE guideline therefore shows
the level of consensus that existed within and between
groups, based on each expert’s opinions at the end of all
the panel discussions. It is made up of 38 tables and a flow
diagram that pinpoints the table to look up. To use the flow
diagram, the clinician needs to be armed with only three or
four key facts about the patient in front of them:

(i) age band;
(ii) complexity of intended surgery;
(iii) ASA grade;
(iv) nature of co-morbidity if ASA III.

Consensus to do a test is represented by a green square;
consensus not to do a test is represented by a red square; and
the frequently encountered absence of consensus is repres-
ented by an amber square (implying ‘consider this test’).

Correspondence to NICE has challenged this presenta-
tion, essentially wanting the question further simplified.
However, given the complicated nature of the question
(11 tests) x (four surgical complexities) x (multiple co-mor-
bidity scenarios encompassing disorders in three systems),
we remain unrepentant, believing that one flow chart and 38
tables is a compact presentation of these data.

Other difficulties about this Guideline have been raised,
either by its authors or in correspondence to NICE. First, the
many parts of the tables coloured amber demonstrate clearly
how often our experts could not reach a consensus. When
our experts could not agree on testing in a particular scen-
ario, the Guideline does not provide clinicians with any
decisions regarding that scenario. Second, there are a few
instances where fewer tests are recommended for an ASA
grade II patient than for an ASA grade I patient. These are
inconsistencies but are not errors, and arose because separate
panel meetings considered tests for ASA grade I patients and
tests for ASA grades II and III patients. They had slightly
different compositions and presumably slightly different
beliefs. Third, we would have liked to find more tightly
defined scales of co-morbidity and surgical complexity.
ASA grading, for example is not backed up by any rules
on how to categorize individual clinical scenarios. Fourth,
we presume that just as in other screening programmes, it is
possible that a patient may be harmed as an indirect result of
carrying out a test, but this remains theoretical and was not
feasible to explore given the poor quality of the literature.
Finally, the NICE Guideline does not cover every facet of
everyday routine practice, omitting for example many car-
diac investigations, fortunately recently well covered in a
guideline from the USA.
This first national Guideline summarizes the knowledge that is available now and provides a framework for future improvement. Anaesthesia and surgery departments should ensure that the pocket version features in every induction of new junior staff. Clinicians at the local level may also decide to use the NICE Guideline for wider purposes than supporting individual doctors’ preoperative testing decisions. The Guideline can form the basis of local guidance on pre-operative testing, which may need updating, green boxes representing minimum testing and red boxes representing over-testing. Clinicians may wish to clarify the local level view on amber tests. The Guideline could be used to develop audits aiming at identifying pockets of residual extreme over- or under-testing, at local or national level. Other guidance from the NHS Modernisation Agency on how the over- or under-testing, at local or national level. Other guidance from the NHS Modernisation Agency on how the preoperative process should be organized and how testing fits in, is at www.modern.nhs.uk/theatreprogramme.

J. Carlisle
Department of Anaesthesia
Torbay Hospital
Torquay, UK

J. Langham
Health Services Research Unit
London School of Hygiene and Tropical Medicine
London, UK

G. Thoms*
Department of Anaesthesia
Manchester Royal Infirmary
Manchester, M13 9WL
UK

*Corresponding author. E-mail: gavin.thoms@man.ac.uk

References

Who is at increased risk of pulmonary aspiration?

Since the danger of pulmonary aspiration was recognized in the 1930s in obstetric anaesthesia, and Mendelson established its aetiology in 1946, efforts have been made to reduce its incidence: fasting before anaesthesia, prophylactic medication (such as antacids or H2 antagonists), rapid-sequence induction of anaesthesia with application of cricoid pressure, and the use of a cuffed tracheal tube.

The laryngeal mask airway has gained a firm place in anaesthetic practice since it was made available to clinicians in 1988. The frequency of tracheal intubation has been decreasing, because of routine use of the laryngeal mask airway and several other supraglottic airways (such as the Laryngeal Tube or Airway Management Device). Nevertheless, there has been ongoing concern that avoidance of the use of a cuffed tracheal tube might increase the incidence of pulmonary aspiration. Some consider that spontaneous breathing should be maintained when the laryngeal mask is used, because intermittent positive pressure ventilation may...