Assessing the effectiveness of new technologies

Adrian K Dixon

Introduction

Rapid changes in the way in which healthcare is delivered continually necessitate different methods of assessing new technology. Economical issues have become paramount. Perhaps even more than mere fiscal assessment comes the need to avoid hospitalisation and reduce the number of patient contacts with expensive secondary care. Whether these financial imperatives are perceived by the patient as optimal often remains unanswered, even in these days of supposed 'patient choice'. The main methods of new technology assessment are discussed in this article, along with possible future assessments in the light of recent changes in the way in which diagnostics are being introduced into the United Kingdom (UK).

New technology assessment

The original hierarchical five level method of assessing new technology within the imaging field is now well established with only minor variations[1,2]. These are

- Technical performance;
- Diagnostic performance;
- Diagnostic impact;
- Therapeutic impact; and
- Impact on health.

Technical performance

The first level is that of technical performance, which assesses whether the new equipment or technique does deliver what it is expected to do on the technical front. For a new piece of diagnostic imaging equipment, this might assess whether or not the new machine yields anatomical images of spatial (and/or contrast) resolution equal to or better than existing equipment[3]. For nuclear medicine and other functional imaging techniques this might assess the additional physiological data that is obtained. For a new interventional stent, it might be a more mechanical assessment about tensile strength and biocompatibility.

For the average radiology department, there is little involvement in such assessments as manufacturers will not bring novel techniques or

> "There should be no let up in the quest for even better techniques for identifying women at risk from this dreaded disease (breast cancer)."

> Whilst the overall accuracy is very high (NHSBSP), this is based on the very large numbers of true negative findings in normal women.

technologies to the market place without all such technical performance information available. However, research departments become involved with assessments of prototypes, and some of these may have been developed by their staff members.

Diagnostic performance

Diagnostic performance, namely how well the new technique fares with regards to making the diagnosis, is often regarded as the be all and end all of technology assessment. It is often regarded, erroneously, as being synonymous with the diagnostic accuracy of the new technique. It is now realised that studies describing overall accuracy are significantly influenced by the prevalence of disease in the population under scrutiny. In fact, the markers of most importance are the sensitivity and specificity of the new investigation and these are often combined to produce a receiver operator curve for the new investigation[4]. Again, many prestigious research departments become involved in such early assessments. But of course, the prevalence of various disease processes in these institutions may be far removed from the real world of a typical general hospital. Hence, there is need for early, large, multicentre studies of new methodologies, which will include a spectrum of different practices.

The data now available from the very high quality UK National Health Service Breast Screening Programme (NHSBSP) provides interesting conclusions: whilst the overall accuracy is very high[5], this is based mainly on the very large numbers of true negative findings in normal women. The predictive value of a positive result is not all that high and this means that a fair number of normal women still have to undergo a traumatic biopsy. The sensitivity for some of the more aggressive lesions is rather lower and MRI may be better, especially in younger women[6]. Consequently, despite the considerable advances which have arisen as a result of the NHSBSP, there should be no let up in the quest for even better techniques for identifying women at risk from this dreaded disease.

The assessment of diagnostic performance is relatively straightforward in breast screening where a final diagnosis (cancer or no cancer) is definitively established. Many workers have pointed out how difficult this becomes when the diagnosis is elusive and the patient may not necessarily undergo early biopsy or surgery – viz magnetic resonance imaging (MRI) in the diagnosis of multiple sclerosis[7]. Even for something relatively straightforward such as MR of the knee, the fact that only a selection of patients undergo arthroscopy hinders the assessment of diagnostic performance[8].

Diagnostic impact

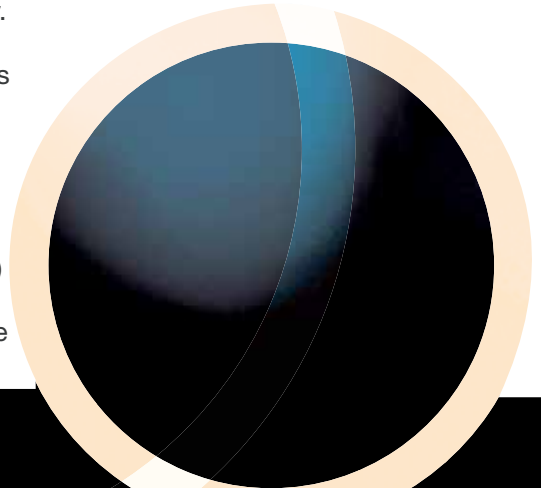
If a new technique successfully passes through the above two stages of assessment, it should be possible to prove that it helps make an impact on the clinician's diagnosis, either by providing a new, unexpected diagnosis or by improving the clinician's confidence in their working clinical diagnosis. Such information is extremely difficult to obtain unless the confidence in the working

diagnosis is established before the first imaging investigation. Historically, diagnostic confidence information has been obtained in patients who were referred for VQ scintigrams for possible pulmonary embolus (PE) where the *a priori* clinical probability was necessary in order to provide a definitive report. This made it relatively simple to compare and prove the beneficial diagnostic impact of computed tomography (CT) pulmonary angiography[9] and has led to general acceptance in the UK that nuclear medicine (NM) referrals for possible PE should now be limited to those patients with a relatively normal chest radiograph, with all other patients directed towards CT.

Relatively few clinicians state their possible differential diagnosis in general referrals for imaging examinations, let alone their confidence in the leading diagnosis. Only when this information is demanded at the outset can the true diagnostic impact of a new investigation be measured. Some might argue that all imaging is aimed at reducing diagnostic uncertainty[10] and that any investigation which improves diagnostic confidence is a 'good thing'. But an investigation which merely confirms the clinician's diagnosis may be an unnecessary luxury. Alternatively, an investigation which completely alters the clinician's diagnosis (say from benign disease to probable cancer) may completely change the patient's clinical course – and save the clinician the embarrassment of going up the wrong diagnostic alley. And such changes in diagnosis (diagnostic impact) must be measured as radiologists and clinicians need to justify the considerable expense of diagnostic certainty.

Therapeutic impact

Now that the investigation under scrutiny has passed the first three levels of the technology assessment hierarchy, it must be shown to have therapeutic value. If, after an investigation has been performed, the clinician ends up doing what he or she would have done anyway, it could be argued that the test was unnecessary. In the days when radiologists had to scrutinise requests assiduously to avoid too many referrals and a departmental overspend, they frequently asked 'will the result of this investigation change your management?'. Indeed, this is still a very valid question, particularly when radiation exposure is at stake[11,12]. Again, clinicians must be encouraged to state before an investigation details about their proposed management; and they must also be asked what their management will be once the result of the investigation is apparent. Only by this method can the true therapeutic impact be measured. This can be achieved by means of a prospective observational study[13] but a purer study is where the new investigation is assessed alongside a conventional investigation on the same patient population by means of a randomised trial[14,15]. However, randomised trials are notoriously difficult to perform in diagnostic radiology and ethical review boards may refuse to sanction a study where one arm of patients is 'denied' access to the diagnostic test under scrutiny, no matter how new or experimental.



Impact on health

It should follow that any investigation of proven technical and diagnostic performance with positive clinical and therapeutic impact would be associated with a beneficial impact on health. But this may be difficult to prove: the side effects of surgery may mask the health gain in the immediate post-operative period and sometimes the diagnosis may have such an unfavourable outcome (for example, pancreatic cancer) that the beneficial influence of the diagnostic test is masked by the natural history of the disease. Even for benign disease, it is difficult to prove the benefit of high technology diagnosis in terms of improvement in quality of life[16]. It is much easier to prove the beneficial influence of interventional radiological techniques, where great advances have been made recently[17]. As a result, several surrogate markers have been used to assess impact of new techniques on health. Clearly, the avoidance of ionising radiation is one marker; and another may be the avoidance of potentially dangerous investigations. The use of MRI instead of Endoscopic Retrograde Cholangiopancreatography (ERCP) is one example; the development of fluid sensitive techniques allowed the introduction of MR cholangiopancreatography (MRCP)[18], which has virtually replaced diagnostic ERCP. Interestingly, this change in practice has come about with no full health technology assessment – merely common sense! In these days of patient choice, it may be that we should ask the patients which test they would prefer – all other information being equal; the presumed preference for MR over myelography was never

formally assessed but can be assumed. However, the preference for MR over some other techniques may not be as apparent as expected; for example, not all patients prefer conventional MR of the shoulder over conventional arthrography – despite the perceived invasiveness of the latter[19].

Societal impact as a technology assessment measure

Because of the difficulties in proving the benefits of high technology investigations, health economists and others responsible for planning and purchasing healthcare have looked to yet other surrogates. Increasingly, health economists point out that the really expensive bits of healthcare relate to secondary care and in-patient stays. If it can be shown that the judicious use of imaging can make secondary care more efficient and shorten hospital admissions, then the case for the greater use and increased expenditure on imaging is made. And this is very much behind the recent UK NHS initiatives in providing increased access to imaging. Even with the recent expansion in imaging capacity, first on the back of various cancer initiatives and now from additional independent sector provision, the UK still performs far fewer CT and MR examinations than other developed nations. For example, the rate for CT examinations in the USA is now around 250 per 1000 of the population per annum; many times the rate in the UK. Numerous experts have tried to decide on the optimal level of provision for all such examinations, with scant evidence available. However, it is worth considering some of the societal benefits of increased imaging capacity which are now being addressed:

The patient. Patients do not like waiting for investigations and would prefer to avoid unnecessary and additional visits to clinics, hospitals, etc. So, if there is adequate capacity to offer an investigation (eg Chest CT) on the day of the clinic attendance, the patient is saved an unnecessary second visit for the diagnostic imaging required, and the whole investigation is cheaper – no bookings or letters, less car parking, etc. There is considerable evidence that MR is a better investigation for lumbar spine problems than plain radiography yet many patients are still referred for conventional lumbar spine radiographs[20]. Because of this a small number of patients suffer a delay in the diagnosis of serious disease (metastatic deposits, disc space infection, major disc herniation, etc). When multiplied, the cost of such delays in diagnosis may justify the increased expenditure.

The referring clinician. Only recently have clinicians started to accept that the objective findings of imaging are, in many situations, superior to their subjective clinical examination, even for something simple such as the presence or absence of an abdominal mass[14]. Furthermore, the newer generation of clinicians relies much more on the results of imaging to guide their management decisions and will frequently insist on high technology imaging before offering a final clinical diagnosis. Additionally, most modern clinicians prefer the newer investigations to the old[21]. Indeed, there is good evidence that modern imaging can optimise the surgical approach in many conditions, such as rectal carcinoma[22]. Such advantages need to be quantified in many more clinical situations. Interestingly, defence organisations have started to realise the importance of preoperative imaging and there are now some cases coming through where surgery is regarded as inappropriate in the light of the imaging findings (or the absence thereof).

The community. CT and MR were both developed at times when healthcare costs were under very close scrutiny and both came to be regarded (erroneously) as expensive investigations. Of course, when these machines could only handle one patient an hour they were expensive and many of the original cost-effectiveness studies were based on very high costs per procedure. However, both techniques can now offer very rapid, high volume functionality for most routine referrals. The costs of CT and MR are often lower than the alternatives they have replaced[23,24]. For example unenhanced CT of the abdomen is cheaper than even a short intravenous urogram (IVU); the cost of an MR of the lumbar spine pales into insignificance compared with the cost of a myelogram.

"If it can be shown that the judicious use of imaging can make secondary care more efficient and shorten hospital admissions, then the case for the greater use and increased expenditure on imaging is made."

But the real gain for high technology comes with reduced hospital admissions. It has been shown in numerous studies that early and judicious use of a single, high technology investigation can provide full diagnostic information (thereby avoiding a lengthy sequence of other tests) which, in turn, is related to shorter hospital stays[14,15]. Indeed, for abdominal conditions, a prompt CT examination may assist the emergency physician to decide not to admit the patient. Again, much work is still required to prove that overall costs can be reduced by increasing the availability of and access to appropriate imaging.

Conclusion

Radiologists, radiographers and others allied to imaging have, hitherto, been satisfied merely in showing the marvellous images produced by the increasingly sophisticated imaging devices now available and basking in the collective, reflected glory. Although these new techniques obviously assisted the referring clinician and often saved the patient more invasive tests, there has still been relatively little effort made to prove that they contribute to the totality of healthcare. Only by proving the effectiveness of pounds/dollars/euros spent on imaging will we be able to obtain the real and sustainable growth in imaging which many of us consider necessary. And, we will have to be on very firm ground, because increased expenditure on imaging will, almost certainly, mean cuts elsewhere.

Adrian Dixon is an honorary consultant at the Department of Radiology, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust and professor of radiology at the University of Cambridge.

References

1. Fineberg HV, Wittenberg J, Ferrucci JT Jr, Mueller PR, Simeone JF, Goldman J. The clinical value of body computed tomography over time and technologic change. AJR 1983;141:1067-72.
2. Mackenzie R, Dixon AK. Measuring the effects of imaging: an evaluative framework. Clin Radiol 1995;50:513-8
3. Mackenzie R, Logan BM, Shah NJ, Keene GS, Dixon AK. Direct anatomical-MRI correlation: the knee. Surg Radiol Anat 1994;16:183-192
4. Doubilet PM. Statistical techniques for medical decision making: applications to diagnostic radiology. AJR 1988; 150:745-50.
5. Department of Health. Breast Cancer Screening. www.dh.gov.uk/PublicationsAndStatistics/Statistics/StatisticalWorkAreas/StatisticalHealthCare/StatisticalHealthCareArticle/fs/en?CONTENT_ID=4104034&chk=mjJ%2Bof
6. Leach MO, Boggis CR, Dixon AK, Warren RM, et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). Lancet 2005; 365: 1769-78.
7. Harbord R, Main C, Deeks, JJ et al. Accuracy of magnetic

resonance imaging for the diagnosis of multiple sclerosis: systematic review. BMJ 2006; 332:875-84.
8. Mackenzie R, Palmer CR, Lomas DJ, Dixon AK. Magnetic resonance imaging of the knee: diagnostic performance statistics. Clin Radiol 1996;51:251-7.
9. Cross JJL, Kemp PM, Walsh CG, Flower CDR, Dixon AK. A randomized trial of spiral CT and ventilation perfusion scintigraphy for the diagnosis of pulmonary embolism. Clin Radiol 1998;53:177-82.
10. Hobby JL, Tom BD, Todd C, Bearcroft PW, Dixon AK. Communication of doubt and certainty in radiological reports. Br J Radiol. 2000; 73:999-1001.
11. Royal College of Radiologists. Making the best use of a department of clinical radiology: guidelines for doctors. 2007 RCR, London.
12. IRMER The Ionising Radiation (Medical Exposure) Regulations 2000 www.dh.gov.uk/PublicationsAndStatistics/Publications/PublicationsPolicyAndGuidance/PublicationsPolicyAndGuidanceArticle/fs/en?CONTENT_ID=4007957&chk=FkG/YB
13. Mackenzie R, Dixon AK, Keene GS, Hollingworth W, Lomas DJ, Villar RN. Magnetic resonance imaging of the knee: assessment of effectiveness. Clin Radiol 1996;51:245-50.
14. Dixon AK, Kelsey Fry I, Kingham JGC, McLean AM, White FE. Computed tomography in patients with an abdominal mass; effective and efficient? Lancet 1981;i:1199-1202.
15. Ng CS, Watson CJ, Palmer CR, See TC, Beharry NA, Housden BA, Bradley JA, Dixon AK. Evaluation of early abdominopelvic computed tomography in patients with acute abdominal pain of unknown cause: prospective randomised study. British Medical Journal 2002; 14:325 (7377): 1387
16. Mackenzie R, Hollingworth W, Dixon AK. Quality of life assessment in the evaluation of magnetic resonance imaging. Qual Life Res 1994;3:29-37.
17. Molyneux AJ, Kerr RS, Clarke M, et al. International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. Lancet. 2005;366:809-17.
18. Bearcroft PW, Lomas DJ. Magnetic resonance cholangiopancreatography. Gut 1997; 41: 135-7
19. Blanchard TK, Bearcroft PWP, Dixon AK, Lomas DJ, Teale A, Constant CR, Hazleman BL. Magnetic resonance imaging or arthrography of the shoulder: which do patients prefer? Br J Radiol 1997;70:786-90.
20. Hollingworth W, Todd CJ, King H, Males T, Dixon AK, Karia KR, Kinmonth AL. Primary care referrals for lumbar spine radiography: diagnostic yield and clinical guidelines. British Journal of General Practice 2002; 52: 475-80.
21. Southern JP, Teale A, Dixon AK, Freer CEL, Rubenstein D, Wilkinson IMS, Hall LD, Sims C, Williams A. An audit of the clinical use of magnetic resonance imaging of the head and spine. Health Trends 1991;23:75-9.
22. Brown G, Richards CJ, Newcombe RG, et al. Rectal carcinoma: thin-section MR imaging for staging in 28 patients. Radiology 1999; 211:215-22.
23. Moore AT, Dixon AK, Rubenstein D, Wheeler T. Cost-benefit evaluation of body computed tomography. Health Trends 1987;19:8-12.
24. Hollingworth W, Bell MI, Dixon AK, Antoun NM, Moffat DA, Todd CJ. Measuring the effects of medical imaging in patients with possible cerebellopontine angle lesions: a four-center study. Acad Radiol 1998;5 (Suppl 2):306-309.