Seminars in Fetal & Neonatal Medicine (2005) **10**, 83—90

ELSEVIER

Seminars in FETAL & NEONATAL MEDICINE

www.elsevierhealth.com/journals/siny

Towards safer diagnosis in clinical practice — understanding and using diagnostic tests in the neonatal unit more appropriately

Peter W. Fowlie*

Neonatal Intensive Care Unit, Ninewells Hospital and Medical School, Dundee DD1 9SY, UK

KEYWORDS
Decision making;
Diagnosis;
Diagnostic test;
Evidence-based medicine;
Infant;
Laboratory techniques and procedures;
Newborn;
Quality control

Summary The diagnostic process is a complex task that is more often than not done inherently by clinicians. However, it is in fact based around quantitative risk assessment and, as a result, when done intuitively is open to a significant risk of bias. By adopting a more structured and quantitative approach to diagnosis, clinicians might be in a position to make better diagnostic decisions. To achieve this, explicit recognition about the uncertainty surrounding diagnosis and knowledge about the basic properties of diagnostic tests, including disease incidence and predictive values, is necessary, as well as some consideration of newer concepts such as 'action thresholds'. Examples from everyday neonatal practice illustrate the potential clinical risks associated with the inappropriate use and interpretation of diagnostic tests and the potential benefits of approaching diagnosis in a more robust manner. A number of tools are now readily available to help clinicians move towards more 'evidence-based' diagnosis.

Introduction

Clinicians have been using 'tests' to aid the diagnostic process for centuries. This has in the past contributed to the perceived art of medicine but is a tacit acknowledgement that it is rare to be absolutely certain of the diagnosis. The use of diagnostic tests is clearly not a perfect science however (clinicians still get it wrong) nor is it risk-free. A better understanding of the diagnostic process and how diagnostic tests 'work' might reduce the chances of inappropriate decisions being made.

The diagnostic process

When clinicians first come into contact with a patient, for example a sick newborn infant, they are rarely able to diagnose the problem with absolute certainty straight away. Certain facts are immediately available, however, and on the basis of these facts, clinicians develop a list of potential diagnoses — the differential diagnoses. By identifying and collating further new information they

* Tel.: +44 1382 660111x33218; fax: +44 1382 633921.
E-mail address: peter.w.fowlie@tuht.scot.nhs.uk

1744-165X/$ - see front matter
doi:10.1016/j.siny.2004.09.002

refine this list until a diagnosis (or diagnoses) is reached with as much certainty as possible. Although under certain circumstances diagnosis can be approached differently, this is perhaps the most common method, and one that most clinicians will use inherently.[1]

Every disease or condition has diagnostic criteria or a diagnostic definition. This might be based on symptoms and signs alone, or on a pathognomonic test or series of tests. If this 'gold standard' definition is achieved easily, then clinicians will quite rightly attempt to reach it although, for a variety of reasons, this might not be possible. Achieving the gold standard might involve invasive testing or could have potential serious side effects. It might be technically difficult or be prohibitively expensive. Ultimately, it might be possible to determine the gold standard only at post-mortem!

If absolute diagnostic certainty cannot be achieved for whatever reason, clinicians will inherently consider how likely or unlikely any condition of interest might be given the information that is available. By then considering further 'tests' they will try to make the diagnosis more or less certain, depending on the results. By examining this process objectively and considering quantitative estimates it is possible to move up and down a 'line of uncertainty' (Fig. 1), thus illustrating how useful a diagnostic test might be.[2]

By approaching the diagnostic process in this way, diagnostic 'tests' do not need to be limited to hospital laboratories and radiology departments. Items of the clinical history and clinical examination can be regarded as diagnostic tests, each contributing quantifiably to reaching a diagnosis.[3]

Accuracy

The accuracy of a test can be defined as how well it predicts the condition of interest. To assess this characteristic objectively, diagnostic tests are often measured in some way against an accepted 'gold standard'. This can be illustrated by considering the contribution that measuring C-reactive protein (CRP) makes towards diagnosing bacterial sepsis in sick neonates.

In a group of unwell neonates in whom bacterial sepsis is seriously being considered, the CRP of an infant is compared to the eventual gold standard. Accepting the inevitable difficulties of defining true sepsis, the 'gold standard' might be defined as a positive culture from a normally sterile body fluid, for example cerebrospinal fluid, blood or urine. For simplicity, if a dichotomous test result is used − raised versus normal CRP − a 2 × 2 table can be constructed (Table 1). (Note: studies examining diagnostic tests frequently have design weaknesses that threaten the validity of the results. The following data are taken from a small but methodologically rigorous study.[4])

From the table, the following objective measures can be described:

- The incidence of bacterial infection is the proportion of the population under consideration that is truly infected according to the gold standard [$(a + c)/(a + b + c + d) = 30\%$].
- The positive predictive value of CRP testing quantifies the chances of an infant with a raised CRP truly being infected [$a/(a + b) = 38\%$].
- The negative predictive value of CRP testing quantifies the chance of a sick infant with a normal CRP being free of infection [$c/(c + d) = 85\%$].
- The sensitivity of CRP testing illustrates the proportion of infected infants who have a raised CRP [$a/(a + c) = 83\%$].
- The specificity of CRP testing illustrates the proportion of non-infected infants who have a normal CRP [$d/(b + d) = 41\%$].
- The likelihood ratio associated with a raised CRP is the relative risk of a raised CRP (positive result) in infected infants compared to non-infected infants [sensitivity/(1 − specificity) = 1.42].

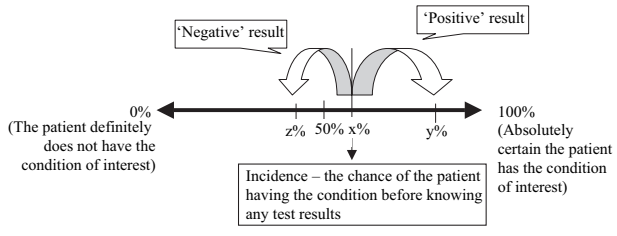

Figure 1 Diagnostic 'line of uncertainty'. From a given starting point (the incidence of the condition), tests allow the clinician to become more or less certain the patient has the condition of interest depending on the result and on how 'good' the test is.

Table 1 A 2 × 2 table that enables the description of the incidence, positive predictive value, negative predictive value, sensitivity, specificity and likelihood ratio of CRP testing

	Gold standard (proven sepsis on culture)	
	Infected	Not infected
Raised CRP (a positive 'test')	25 (a)	41 (b)
Normal CRP (a negative 'test')	5 (c)	29 (d)

See the text for calculations using (a)–(d).

- The likelihood ratio associated with a normal CRP is the relative risk of a normal CRP (negative result) in infected versus non-infected infants [(1 − sensitivity)/specificity = 0.40].

Which of these measures is of most practical value to the neonatologist at the cot side? The question that the clinician is likely to be asking is 'If the CRP is raised (or normal), what are the chances that this infant has sepsis (or not)?' This is in fact what the predictive values indicate and it is the predictive values that will usually be the most useful measure of diagnostic accuracy to clinicians, particularly if the values are derived from a population very similar to the one under scrutiny. In this case, an unwell infant with a raised CRP has a 38% chance of being truly infected. If the CRP is not raised, the infant only has a 15% chance of being infected. This can be visualised on the 'line of uncertainty' as in Fig. 2:

These data suggest that a single measure of CRP at the onset of illness can be relatively unhelpful. In fact, the use of acute-phase proteins and other potential markers of infection, such as white cell counts and ratios, has been and continues to be investigated extensively.[5–7] In general, single tests and even series of tests are of limited value, although it has been suggested that some strategies might be clinically useful.[8]

Historically, the performance of diagnostic tests is frequently reported in terms of sensitivity and specificity. Is it ever clinically useful to consider these measures, and when might the likelihood ratio be useful? In general, when a highly sensitive test is negative it will make the diagnosis fairly unlikely, whereas a positive result in a highly specific test will tend to make the diagnosis much more likely. Although this is clearly not as explicit as using predictive values might be, predictive values vary when a population with a different incidence of the condition of interest is considered. For example, the incidence of patent ductus arteriosus (PDA) will be higher when considering preterm infants in the first few weeks of life on the neonatal unit than when following-up such infants in the out patient clinic. Clinical signs ('tests') that might point towards a diagnosis of PDA might well perform differently in terms of predictive value in different settings. Although there are exceptions, test sensitivity and specificity, however, are thought to be more stable across different populations. As the likelihood ratio can be regarded as a function of sensitivity and specificity, it too is less susceptible to variation in incidence across populations and can be used to calculate the chances of an infant having a given condition.[3,9] It is, however, arguably only necessary to consider sensitivity/specificity and likelihood ratios if the data describing predictive values are derived from a population with a very different incidence of the condition of interest to the one under scrutiny.

Some tests can be reported as continuous variable, e.g. thyroid stimulating hormone (TSH) and white cell count (WCC), and the accuracy of these tests can also be assessed. Receiver operating characteristic (ROC) curves and Hosemer–Lemeshow goodness-of-fit tests[10,11] are well accepted statistical contortions that can help in interpreting diagnostic tests, for example by determining appropriate cut off values; however, they are of little value to the clinician faced with a sick neonate. To get clinically meaningful information — predictive values (or likelihood ratios) — continuous variables need to be collapsed into ranges.

Whereas it is useful to consider the use of single diagnostic tests, in reality, clinicians will often use a series of different tests to refine a diagnosis. This can be regarded as taking multiple leaps along the line of uncertainty and is again intuitively how most clinicians think. The more tests that are 'positive' (or 'negative'), the more (or less) likely the diagnosis. With appropriate attention to issues relating to multiple testing and independence of tests, objective values and risks can be calculated.[1]

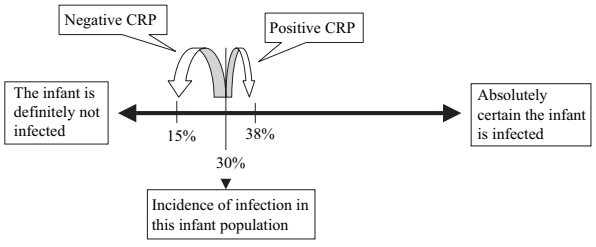

Figure 2 Diagnostic 'line of uncertainty' for CRP in sick infants. From knowing there is an incidence of 30% infection in this type of infant, the CRP (depending on the result) will allow the clinician to be more or less certain the infant is infected as illustrated.

Reliability

As well as being accurate, diagnostic tests must be reliable. That is, the measurement or estimate they are providing must be repeatable. For common laboratory tests this is incorporated into 'quality control' as a matter of routine but it should be considered more widely, especially when tests involving subjectivity are under scrutiny. Various statistics, such as the kappa statistic and inter-rater and intra-rater correlation coefficients, can be used to describe reliability.[12]

The interpretation of cranial ultrasound images is an example in neonatal care when reliability is an issue. Clinicians looking at the same cranial ultrasound will disagree on the findings up to 20% of the time, with agreement beyond chance perhaps no greater than 50%.[13–16] This clearly has implications for how ultrasound should be reported and what actions are taken on the basis of results.

Near-patient testing

Near-patient testing introduces different problems in terms of assessing accuracy and reliability. Near-patient testing commonly involves measuring something — either while in direct contact with the patient or while physically close to the patient — that would otherwise necessitate sending a sample, usually blood, from the patient to the laboratory. There are many reasons why this might be attractive; these are primarily to do with aspects of efficiency, although this cannot be at the expense of accuracy and reliability. Common near-patient testing in neonatology includes the estimation of serum bilirubin using hand-help spectrophotometers[17–20] and the estimation of blood glucose using a reagent strip method.[21–24] These 'tests' purport to measure the same thing as the 'gold standard' — serum bilirubin or blood glucose — but in a technically different way. Under these circumstances, the performance of the near-patient equipment should always be compared directly with the 'gold standard' laboratory estimates. Although such 'quality control' is frequently undertaken by laboratory staff without involving clinicians directly, it is important for clinicians to be aware of how these tests perform, as this will have an impact directly on decision making.

For example, given sufficient accuracy and reliability, it would seem possible that hand-held spectrophotometers used to assess the degree of neonatal jaundice might reduce stress to babies and families, as well as saving time and money. The risk, however, is that falsely low estimates might expose infants to potential bilirubin toxicity, which could be avoided, and that falsely high estimates might expose them to interventions with potential side effects. Several studies have compared bilirubin estimates from such hand-held spectrophotometers with estimates from blood taken at the same time and assayed in the laboratory (Fig. 3). Results from such studies vary but tend to suggest that even the most advanced hand-held device might not be accurate or reliable enough to avoid blood letting for laboratory confirmation once a given estimate of serum bilirubin is reported by the hand-held device.[17–20]

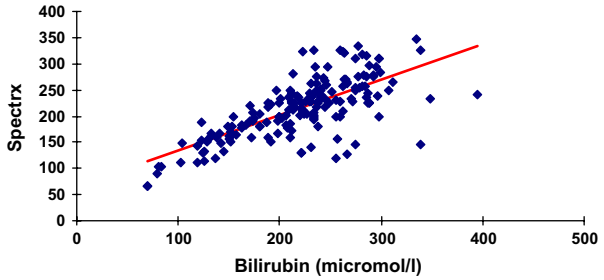

Figure 3 Scatter plot showing correlation between serum bilirubin measured in the laboratory versus estimate from Spectrx® hand-held spectrophotometer (175 matched data, correlation coefficient = 0.71, Spectrx® = 0.675, bilirubin + 66.594; from personal data, 2002).

Action thresholds

Simply knowing how likely it is that a patient has a given condition is of limited value in itself. Having carried out a 'test', how a clinician acts will depend largely on what action(s) is(are) being considered. For example, there is a very low threshold in sick preterm infants for starting antibiotics if the infant is showing signs of possible infection, even if it is only until the results of cultures (the gold standard?) become available 24 h or so later. In such symptomatic infants, what level of risk would a clinician tolerate in order to withhold antibiotics — a 10% risk of sepsis? 5%? 1%? Such thresholds have been described as 'action' thresholds and little empirical evidence is available to describe how clinicians arrive at such thresholds.[25] They contribute to a complex clinical decision making process that involves many factors computed internally by the clinician. It can be argued that if individual tests or series of tests do not allow such action thresholds to be crossed, there is little reason for carrying out the test(s). In the example above, if negative results can never move the clinician to a point where the risk of sepsis is so low that antibiotics might be withheld, is there any point in undertaking the tests in the first place? Will antibiotics not be indicated regardless of results?

To use diagnostic tests to their full potential, explicit consideration of action thresholds needs to be added into the decision making process. There are many examples of action thresholds that neonatologists consider on a daily basis. How certain can clinicians be when considering intracranial bleeding and the risk of subsequent adverse outcome? Cranial ultrasound might never be sufficiently reliable and accurate to justify withdrawal of active treatment based solely on the finding of apparent severe intracranial bleeding,

but the threshold for intimating a worrying prognosis to parents might well be lower. What degree of certainty about a PDA is needed before it would be considered appropriate to transfer a sick neonate for cardiac opinion? Does clinical examination on its own reach this threshold or are other tests required? How certain do clinicians have to be that a neonate has necrotising enterocolitis before withholding milk feeds? In a sick preterm infant with a distended abdomen, would a 'normal' abdominal X-ray reassure clinicians sufficiently to allow continued feeding?

Diagnostic tests and clinical risk

When considering diagnostic tests, several aspects of risk can be considered: risk to the patient, risk to the clinician and risk to the healthcare provider or organisation.

Any test used in clinical practice must be sufficiently accurate and reliable for the purpose intended. If it is not, there is clearly the potential for acting inappropriately. The wrong diagnosis might be made and the patient started on an unnecessary treatment with the very real risk of associated adverse side effects, e.g. indometacin for a PDA that is actually closed. Parents might be given misleading information and an inaccurate prognosis, e.g. misinterpretation of cranial ultrasound images. Infants might miss out on effective treatment with potentially catastrophic consequences, e.g. missed hyperbilirubinaemia (for which there is an effective treatment) on the basis of inaccurate near-patient testing perhaps leading to the need for a longer period of phototherapy, avoidable exchange transfusion or even the development of encephalopathy. Babies might also be subjected to unnecessary medical intervention; these might not be associated with side effects in the same way as drug treatments but they might none the less impact in a negative way on the baby and the family. For example, incorrect diagnosis of neonatal hypoglycaemia using near-patient testing might lead to admission to the neonatal unit, further painful investigation, changes to feeding regimes and possible separation of mother and baby.

By explicitly considering the predictive values of tests along with action thresholds, clinicians can reduce the chances of exposing patients and themselves to inappropriate actions or outcomes. Although it is quite correct to place significant resource towards training staff in the practical aspects of undertaking diagnostic tests and procedures, it is equally important that attention is given to teaching how diagnostic tests work and contribute to the diagnostic process. How many clinicians know, even approximately, the incidence of even the more common conditions they come across, let alone the predictive values associated with tests they might use on a daily basis?

The use of diagnostic tests also comes with associated risk and potential 'cost' for healthcare providers and organisations. The basic financial cost of buying 'diagnostic tests', i.e. equipment and reagents, can be high and even tests that are relatively cheap can incur significant expenditure if they are done frequently. The costs of training staff are often forgotten but cannot be ignored in terms of human resource and organisational time. In the case of screening, programmes can be enormously complicated and expensive to introduce and it might be difficult to show cost benefit and cost effectiveness, especially at local level. In addition, if tests are being used inappropriately, organisations (and individuals) will clearly be less efficient and might even expose themselves to an increased risk of medical complaint and litigation.

The safety and risk management in relation to near-patient testing (i.e. any ward-based diagnostic equipment) but especially, in this context, blood gas analysers and blood glucose analysers, perhaps merits special mention. In the UK, the Association of Clinical Biochemists provides guidelines for near-patient testing[26] in which issues surrounding risk management, training and operation, health and safety and quality control are highlighted. The purchase and subsequent management of any equipment should be undertaken only in collaboration with the appropriate laboratory/diagnostic department. Near-patient testing equipment must be checked regularly by appropriate internal quality control and external quality assessment, as well as being maintained under a formal service-level agreement between users and the appropriate laboratory/department. Every potential user must receive training based around an agreed and established written 'standard operating procedure' and 'named operator list'. Use of any equipment should be restricted to those on such lists, so that any given test result can be traced back to the particular individual who carried out that test. Furthermore, in the UK, statutory bodies such as the Medicines and Healthcare Products Regulatory Agency[27] and the Incident Reporting and Investigation Centre (IRIC)[28] have responsibilities that include monitoring the use of near-patient testing equipment and issuing safety/hazard notices. Healthcare organisations must incorporate ways of assessing and disseminating such information into local policies for near-patient testing and strategies for risk management.

All the issues associated with inappropriate use of tests — incorrect diagnoses, unnecessary further investigation and/or treatments, delays in reaching a diagnosis, unsafe use of near-patient testing — therefore have the potential to reduce efficiency and can result in adverse outcomes. Individual practitioners and healthcare providers and organisations must become more aware of this and develop appropriate strategies for minimising the particular type of risk.

Towards evidence-based diagnosis

Individual clinicians can contribute to a more robust, and ultimately safer, use of tests by beginning to think more critically about diagnosis. They should acknowledge the important properties of diagnostic tests — accuracy and reliability — and their impact on action thresholds. They should begin to consider in detail at least those tests they use commonly in day-to-day practice and look for robust information to support their practice. Assessments of diagnostic tests are published in the medical literature and strategies are available for the efficient searching of the more common electronic databases.[29–31] As with all clinical research, however, the validity of any results presented will depend heavily on how rigorously the study has been conducted. Critical appraisal of such studies is important to exclude the chances of significant bias, as this might have an impact on the subsequent interpretation of the results.[32]

Having confirmed the validity of any particular test, clinicians should then look critically at whether the same test could be carried out in the local clinical setting. Do those undertaking the test have sufficient skills? Is any equipment that might be used the same or sufficiently similar? Are resources available for appropriate quality assurance measures? Generalising results from the research setting might not be possible.[33,34] Even then, clinicians must also begin to introduce explicit and objective action thresholds into their decision making. What actions are being considered on the basis of any potential result(s)? What level of certainty needs to be reached before taking that action? And can that level of certainty ever be reached using the proposed diagnostic strategy?[34]

Research opportunities

Over the last decade, the Cochrane Collaboration has demonstrated the value of rigorously assessing the potential benefit (and harm) of interventions used in medicine.[35] A similar process now needs to be undertaken for diagnostic tests, so that they are not introduced inappropriately into routine practice.[36,37] Systematic reviews of existing studies, as well as robust new studies examining the properties of existing and novel diagnostic tests, are needed. Action thresholds and their impact on decision making need to be explored in greater detail. Ultimately, the ability of diagnostic tests to alter clinically important outcomes — the utility of a diagnostic test — should be tested in randomised controlled trials of different management strategies.[38]

Conclusion

Diagnostic tests are used extensively throughout medicine, including on the neonatal unit; they are often used inappropriately and are frequently misinterpreted. Clinicians, healthcare provider organisations and researchers must begin to consider the diagnostic process more critically and review their management strategies accordingly. Perhaps then the risks inherent in using all diagnostic tests will be reduced.

Practice points

- The diagnostic process is complex and frequently misunderstood by clinicians.
- The use and interpretation of 'tests' extends beyond the laboratory and radiology department and includes the interpretation of items of the clinical history and examination.
- Clinicians need to have some knowledge about the incidence of disease seen in their practice and some knowledge about test performance in their population before diagnostic tests can be used and interpreted appropriately.
- Many commonly used tests in neonatal practice have limited accuracy and limited reliability.
- Clinicians need to think explicitly about action thresholds when considering whether or not a diagnostic test might be useful in any given situation.
- As well as being associated with risks to individual patients, the inappropriate use of diagnostic tests may expose individual clinicians and health care provider organisations to risk in a variety of ways.

Research directions

- Improved training for all clinicians in how to interpret and use diagnostic tests is needed — evidence-based medicine.
- Clinical studies examining common diagnostic tests need to be carried out with greater methodological rigour.
- Systematic reviews and data synthesis of existing studies examining diagnostic test properties are needed.
- Action threshold and decision analysis needs to be explored in greater detail.
- Randomised clinical trials examining the impact of different diagnostic strategies on important clinical outcomes should be conducted before incorporating 'tests' into routine clinical practice.

References

1. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology; a basic science for clinical medicine*. 2nd ed. Lippincott Williams and Wilkins; 1991. [chapters 4—5].
2. Richardson WS, Wilson M, Guyatt G. The process of diagnosis. In: Guyatt G, Rennie D, editors. *Users guides to the medical literature*. AMA Press; 2002. p. 155—67.
3. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB. *Evidence-based medicine: how to practice and teach EBM*. 2nd ed. Churchill Livingstone; 2000. p. P67—93.
4. Hiew TM, Tan AM, Cheng HK. Clinical features and haematological indices of bacterial infections in young infants. *Singapore Med J* 1992;**33**(2):125—30.
5. Fowlie PW, Schmidt B. Diagnostic tests for bacterial infection from birth to 90 days — a systematic review. *Arch Dis Child Fetal Neonatal Ed* 1998;**78**(2):F92—8.
6. Benitz WE, Han MY, Madan A, Ramachandra P. Serial serum C-reactive protein levels in the diagnosis of neonatal infection. *Pediatrics* 1998;**102**(4):E41.
7. Franz AR, Steinbach G, Kron M, Pohlandt F. Reduction of unnecessary antibiotic therapy in newborn infants using interleukin-8 and C-reactive protein as markers of bacterial infections. *Pediatrics* 1999;**104**(3 Pt 1):447—53.
8. Franz AR, Kron M, Pohlandt F, Steinbach G. Comparison of procalcitonin with interleukin 8, C-reactive protein and differential white blood cell count for the early diagnosis of bacterial infections in newborn infants. *Pediatr Infect Dis J* 1999;**18**(8):666—71.
9. Fagan TJ. Nomogram for Bayes's theorem. *N Engl J Med* 1975;**293**:257.
10. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29.
11. Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons; 1989. p.106—18.
12. Streiner D, Norman G. Health Measurement Scales — a practical guide to their development and use. 3rd ed. Oxford University Press [chapter 8].
13. Pinto J, Paneth N, Kazam E, et al. Interobserver variability in neonatal cranial ultrasonography. *Paediatr Perinat Epidemiol* 1988;**2**(1):43—58.
14. Corbett SS, Rosenfeld CR, Laptook AR, et al. Intraobserver and interobserver reliability in assessment of neonatal cranial ultrasounds. *Early Hum Dev* 1991;**27**(1—2):9—17.
15. Pinto-Martin J, Paneth N, Witomski T, et al. The central New Jersey neonatal brain haemorrhage study: design of the study and reliability of ultrasound diagnosis. *Paediatr Perinat Epidemiol* 1992;**6**(2):273—84.
16. O'Shea TM, Volberg F, Dillard RG. Reliability of interpretation of cranial ultrasound examinations of very low-birthweight neonates. *Dev Med Child Neurol* 1993;**35**(2):97—101.
17. Briscoe L, Clark S, Yoxall CW. Can transcutaneous bilirubinometry reduce the need for blood tests in jaundiced full term babies? *Arch Dis Child Fetal Neonatal Ed* 2002;**86**(3):F190—2.
18. Rubaltelli FF, Gourley GR, Loskamp N, et al. Transcutaneous bilirubin measurement: a multicenter evaluation of a new device. *Pediatrics* 2001;**107**(6):1264—71.
19. Bhutani VK, Gourley GR, Adler S, Kreamer B, Dalin C, Johnson LH. Noninvasive measurement of total serum bilirubin in a multiracial predischarge newborn population to assess the risk of severe hyperbilirubinemia. *Pediatrics* 2000;**106**(2):E17.
20. Wong CM, van Dijk PJ, Laing IA. A comparison of transcutaneous bilirubinometers: SpectRx BiliCheck versus Minolta AirShields. *Arch Dis Child Fetal Neonatal Ed* 2002;**87**(2):F137—40.
21. Papp M, Sharief N. Comparison of two strip test methods of whole blood glucose measurement in the neonatal period. *Acta Paediatr* 2001;**90**(9):1042—6.
22. Peet AC, Kennedy DM, Hocking MD, Ewer AK. Near-patient testing of blood glucose using the Bayer Rapidlab 860 analyser in a regional neonatal unit. *Ann Clin Biochem* 2002;**39**(Pt 5):502—8.
23. Newman JD, Pecache NS, Barfield CP, Balazs ND. Point-of-care testing of blood glucose in the neonatal unit using the AVL Omni 9 analyser. *Ann Clin Biochem* 2002;**39**(Pt 5):509—12.
24. McNamara PJ, Sharief N. Comparison of EML 105 and advantage analysers measuring capillary versus venous whole blood glucose in neonates. *Acta Paediatr* 2001;**90**(9):1033—41.
25. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;**302**(20):1109—17.
26. Association of Clinical Biochemists, <http://www.acb.org.uk/>. [accessed 19 May 2004].
27. Medicines and Healthcare Products Regulatory Agency, <http://www.mhra.gov.uk/>. [accessed 19 May 2004].
28. Incident Reporting and Investigation Centre, <http://www.show.scot.nhs.uk/shs/>. [accessed 19 May 2004].
29. Bachmann LM, Coray R, Estermann P, Ter Riet G. Identifying diagnostic studies in MEDLINE: reducing the number needed to read. *J Am Med Inform Assoc* 2002;**9**(6):653—8.
30. Bachmann LM, Estermann P, Kronenberg C, Ter Riet G. Identifying diagnostic accuracy studies in EMBASE. *J Med Libr Assoc* 2003;**91**(3):341—6.
31. <http://www.ncbi.nlm.nih.gov/entrez/query/static/clinical.html>. [accessed 31 March 2004].
32. Jaeschke R, Guyatt G, Sackett. DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994;**271**(5):389—91.
33. Barratt A, Irwig L, Glasziou P, et al. Users' guides to the medical literature: XVII. how to use guidelines and recommendations about screening Evidence-Based Medicine Working Group. *JAMA* 1999;**281**(21):2029—34.
34. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about

a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;**271**(9):703—7.
35. Cochrane organisation, <http://www.cochrane.org>. [accessed 31 March 2004].
36. Cochrane organisation, <http://www.cochrane.org/cochrane/sadtdoc1.htm>. [accessed 31 March 2004].
37. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;**323**(7305):157—62.
38. Merelle ME, Nagelkerke AF, Lees CM, Dezateux C. *Newborn screening for cystic fibrosis (Cochrane Review)*. The Cochrane library, issue 2. Chichester, UK: John Wiley & Sons, Ltd; 2004.