Is the practice of haematology evidence based? II. Evidence based morphological diagnosis

Diagnosis of diseases of the blood relies on the interplay of many factors, among which morphology plays a variable part depending on which diseases fall within the differential diagnosis. Haematologists, unlike most other pathologists, are also clinicians; we talk to, touch, examine, investigate, and treat our patients and observe the response to treatment or the natural history of their diseases. It is difficult entirely to divorce these influences from the process of morphological diagnosis without introducing an element of artificiality. Nevertheless, there are a few disorders in which morphological appearances of haemopoietic cells provide some of the important—if not the main—criteria for a diagnosis. Prominent among these are the leukaemias, and they are a useful group to illustrate how far, if at all, the practice of morphological diagnosis is evidence based. By examining the five linked steps involved in the practice of what is currently meant by the term “evidence based medicine” and adapting them to the process of morphological diagnosis we soon notice areas in which we fall short of the highest standards expected by this approach.

How should we make the right diagnosis and what potentially answerable questions can we formulate about the material we see?

Diagnostic criteria which use white cell counts, proportions of blasts in the bone marrow or blood, and other superficially simple measurements assume that we can reliably identify the nature of the cells whose numbers we count. We need to know what particular features of nucleated cells allow us accurately to distinguish blasts from more mature cells and malignant blasts from non-malignant immature cells (in the diagnosis of acute leukaemias) and, if possible, non-malignant from malignant “mature” cells (in the case of chronic leukaemias). Words are clumsy tools with which to record the appearances of cells. It is difficult to formulate answerable questions about which particular morphological appearances or combinations of appearances will most surely guide us to the right answer. In practice the question we usually ask is “do these appearances look like those of other cases whose diagnoses have been made either by ourselves or by others?” Despite the undeniable success of our capacity for pattern recognition, our ability to measure such appearances is unrefined, even if we could agree upon which to focus. In addition, unless there is some incontrovertible, non-morphological gold standard (one possible example being the presence of a Ph chromosome or rearrangement of BCR/ABL, its molecular hallmark, in chronic granulocytic leukaemia), doubts about the validity of the diagnosis may weaken any conclusions we are tempted to draw about the relative discriminatory power of the morphological features we have chosen (or have been taught) to concentrate on.

Do we track down with maximum efficiency the best evidence? Sadly the answer, even if “yes,” should be qualified by an appreciation that the best we can do here, at present, is not very good. The main reason for this lies in the nature of the evidence itself and the data banks in which most of it is stored. The evidence, in the main, consists of images. A very small sample is contained in atlases. Some are accompanied by a few words attempting to describe the subtle and diverse range of visible cytological abnormalities. Lastly, and by far the biggest store, are our collective and individual memories. These images are not always reliably or efficiently accessed, no matter how conscientiously we try.

Do we critically appraise the evidence?

Often the conditioned, pattern recognition process works so swiftly that we reach a diagnosis without knowing how or why. We often then use favourite or fashionable descriptive terms to rationalise the diagnosis. No conscious critical appraisal is involved in such circumstances and we will probably not recognise when we have gone wrong. In the case of acute leukaemias we often make the first jump to “acute leukaemia” very quickly. Then we slow down, and use our critical faculties actively to seek further evidence of the subtype. Some of this further evidence is morphological in nature. But do we, each time we make such a morphological diagnosis, actively appraise and compare what we have found in a particular blood or marrow smear with all the evidence in the data banks? Probably not.

In the case of acute lymphoblastic leukaemia (ALL), confirmed by lack of Sudan black or myeloperoxidase positivity, most of us are only concerned to distinguish the French-American-British (FAB) subtype L3 (Burkitt’s) leukaemia from the other forms of ALL. Similarly, in acute myeloblastic leukaemia (AML) the process of accurately subdividing, by morphology, cases into their respective FAB types tends to be restricted to enthusiasts. M3 AML—acute promyelocytic leukaemia—is, however, being treated somewhat differently from other types of AML since the advent of all-trans retinoic acid. Most haematologists will do their best to detect this subtype of leukaemia morphologically in case cytogenetic or molecular confirmation fails or is unavailable; we will analyse the features of potential cases and compare them with published descriptions, notably those of the FAB group; provided we have already been primed in some way to include M3 AML in the differential diagnosis.

The chronic myeloid leukaemias provide a good example of an opportunity to compare what we see with published evidence. A meticulous differential count and assessment of dysplastic features allows correct assignment of almost all cases to three categories, chronic granulocytic leukaemia (CGL), atypical CML, and chronic myelomonocytic leukaemia. Careful morphologists have an impressive ability correctly to predict presence of a rearranged BCR/ABL, even in those rare cases of Ph negative CGL. After further emphasis of the crucial role of morphological assessment, the diagnostic criteria for these three entities were incorporated into an internationally accepted classification. A prerequisite for morphological diagnosis is that the blood film is conscientiously examined. Nevertheless, this approach still begs some questions; the theoretically impressive...
discriminatory power of carefully carried out differential counts and qualitative examination of the film assumes that the observers accurately identify the circulating nucleated cells and detect dysplastic features if they are present.

Do we apply the results of our appraisal to our practice?
This is relatively straightforward. If the appraisal helps us to make or refute a particular diagnosis then we will surely accept that help, particularly if the patient’s management will be affected. Many haematologists will still go through such a process even if no clinically important therapeutic changes might follow, for example, changing a diagnosis from M2 AML to M4 AML. The limitations to the effectiveness with which we apply the results of our appraisal lie more in the nature and quality of the evidence we have examined—both that present in the case in question and that stored in the data banks—and the efficiency with which we gain access to that evidence than in an intrinsic failure to use the results of appraisal.

Do we evaluate our own performance?
In practice, we do this quite rarely—at least as far as formal evaluation goes. Referral hospital/regional review of diagnoses takes place in a few diseases. Central review of diagnostic material takes place in at least some leukaemia trials. The accuracy of diagnosis of leukaemia now seems impressive and this perception of accuracy underpins many large, expensive trials. For example, only 23 cases were excluded because of misdiagnosis (15 AML, two non-Hodgkin’s lymphoma, one aplastic anaemia, one neuroblastoma, and one typical leukaemia, two B-ALL, two non-Hodgkin’s lymphoma). Many haematologists will still go through such a process even if no clinically important therapeutic changes might follow, for example, changing a diagnosis from M2 AML to M4 AML. The limitations to the effectiveness with which we apply the results of our appraisal lie more in the nature and quality of the evidence we have examined—both that present in the case in question and that stored in the data banks—and the efficiency with which we gain access to that evidence than in an intrinsic failure to use the results of appraisal.

The future
Can we practise truly rigorous, evidence based, morphological diagnosis over a wider haematological arena with the sort of scientific credibility now aspired to by those, say, assessing the efficacy of a new treatment regimen? We would need, first, to amass the evidence, and it would have to be evidence of somewhat better quality than is currently available to us for the vast majority of morphological diagnoses. We would have to mount studies of hypothetically valuable morphological features, each given appropriate gradings—degree of poikilocytosis, nuclear stippling, neutral shape, vacuoles, cytoplasmic inclusions, shape and staining of platelets, marrow cellularity, fibrosis or whatever (with or without some other data like the blood count and indices?), then assess their relative value to diagnosis by examining material where the source (putative disease of interest, superficially similar disease, control tissue) is hidden from the observers. We would then need to analyse the data and determine specificity and sensitivity (of each feature, or particular combinations of features?) when compared with the gold standard (ah yes! another fraught issue—how do we establish the right answer?), perhaps calculate $\kappa$ values if several observers examine the material and then publish the results in a peer reviewed journal. We would then need to repeat the exercise for any new, hypothetically valuable morphological feature with the aim of improving the $\kappa$ value. Then there are the problems of nomenclature and classification—also shifting sands.

As to evaluation of our performance, an extension of the national external quality assurance scheme (NEQAS) could be valuable. We could be asked, for example, to note the five most important abnormalities in a blast cell population and then offer a diagnosis. It might then be possible to determine which criteria most of us use to distinguish, say, vacuolated L1 ALL from L3 (Burkitt’s) ALL, or M3 AML from other forms; in both of these scenarios something close to the right answer could have been determined beforehand by ancillary investigations. Similar studies might compare reactive marrows with low level malignant lymphoblast infiltrates in children, or relatively non-clumping non-haemopoietic tumour with leukaemia cells. In such areas, an extension of the role of NEQAS might both assess our individual performances, albeit in rather artificial circumstances, and provide the data necessary to mount formal prospective studies of the relative strengths of particular morphological features in achieving a consensus diagnosis which was also probably the correct one.

However, the task in its entirety requires so prodigious an effort that it is probably unrealistic except in one or two isolated fields of morphological diagnosis. The challenge to haematologists will therefore be to pick fields that are both important enough and in which we have a good enough chance of success to justify the massive investment of time and labour. In the mean time we should not totally discount the “evidence” currently available to us, nor the ways in which we try to access it or we will face some real practical difficulties!

M M REID
Department of Haematology,
Bishop Auckland Hospitals NHS Trust,
Bishop Auckland, Co Durham, UK

M J GALLOWAY
Department of Haematology,
Bishop Auckland Hospital NHS Trust,
Bishop Auckland, Co Durham, UK

Historically it was well known that arsine could be released in the PVC plastic covers of infant cot mattresses.2

Combination and phosphine from fire retardants and other chemicals resulted from the production of the toxic gases arsine, stibine, and phosphine from fire retardants and other chemicals. The suggestion that 10% of SIDS could be related to these fires has provoked controversy. One lessontoberelearnedisthatthepopularmediaarenotnecessarily the best place for scientific disputes to be aired and resolved.

An unexpected infant death for which there is no obvious explanation leaves the parents and medical staff with major questions to answer. Could the death have been prevented and will it happen again with future children? Was the death from natural causes or could some form of abuse have taken place? This situation is unsatisfactory and it is not surprising that considerable effort has been put into trying to elucidate the cause or causes of SIDS and that some of the work has provoked controversy.

In the past decade sleeping position, body temperature, and parental smoking have been linked to SIDS, and the impact of this knowledge on infant care has led to a major reduction in deaths in many countries. The evidence from which these linkages were made was epidemiological and it is still not known what makes a particular infant susceptible to SIDS. Theories proposed include anatomical abnormalities of the respiratory tract, infections, immunisation, nutritional deficiencies, environmental toxins, and metabolic disorders. The latter two hypotheses are particularly attractive as they lend themselves to experimental testing.

The hypothesis that inherited metabolic disorders, in particular inborn errors of fatty acid β oxidation, were a major contributor to SIDS resulted from necropsy studies showing excess fat in the livers and hearts of SIDS babies.3 The suggestion that 10% of SIDS could be related to these disorders, especially medium chain acylCoA dehydrogenase deficiency (MCAD), was disproved by both metabolic and genetic studies and it is now believed that less than 1% of SIDS cases are caused by these disorders. Other inborn errors suggested as causes of SIDS include glycogen storage disease type I, defects of gluconeogenesis, and mitochondrial DNA mutations. While the presence of an appropriate DNA mutation is almost certainly diagnostic it is very important in metabolic studies to use age matched control data. It is unlikely that inborn errors are a major contributor to SIDS but it is important to diagnose these small numbers for genetic counselling and prevention of future cases.

Toxic environmental factors in SIDS have been the subject of several studies. Chlorinated pesticides, polychlorinated biphenyls, and metals have all been measured in SIDS cases. Particular metals that have been studied are lead, cadmium, mercury, manganese, iron, copper, zinc, antimony, and arsenic. In most reports the data either show no difference between SIDS and controls, or conflicting information is found. Studies of low concentrations of trace elements are complicated by contamination and require a high degree of technical skill and attention to detail.

Two chemists, Sprott in New Zealand and Richardson in the United Kingdom, developed the hypothesis that SIDS resulted from the production of the toxic gases arsine, stibine, and phosphine from fire retardants and other chemicals in the PVC plastic covers of infant cot mattresses.2 Historically it was well known that arsenic could be produced from copper arsenate pigment by the fungus Scopulariopsis brevicaulis and that it could have toxic effects, and this observation was extended to include antimony and phosphorus. These gases can act by inhibiting cholinesterases. It was suggested that the increase in SIDS observed in the United Kingdom after 1950 was caused by the use of these fire retardants. The hypothesis was treated with some scepticism but was testable.

Reports that increased antimony concentrations were found in the blood and tissues of SIDS patients received massive publicity in the United Kingdom through two television programmes in 1994, leading to withdrawal from sale of cot mattresses containing the suspect materials. As a result of the public pressure caused by the programmes several studies were designed to seek evidence for the production of stibine, arsine, and phosphine from cot mattresses. Those that were unsuccessful were criticised for not following Richardson’s methods, but in one study which did do so there was no evidence of the presence of the fungus or the production of gases.1

Many of the basic rules of scientific and medical research were not followed in the development of this story. Little is known about the pharmacology and toxicity of stibine in infants and therefore interpretation of blood and tissue concentrations is difficult. Studies to identify the presence of Scopulariopsis brevicaulis in cot mattresses are inadequate and the production of stibine in this situation remains disputed. More importantly there have been no adequate controlled clinical trials which address the particular problem of the nature of mattress covers.

A recent paper has emphasised the analytical difficulties of accurately determining antimony in urine, blood, and tissues1 and in the March issue of the journal Cullen and colleagues7 began to supply some of the important missing data with a study of the blood and urine antimony concentrations in normal infants under one year of age. Further similar studies, together with clinical trials, should finally prove or disprove the contaminated mattress hypothesis.

One lesson to be relearned is that the popular media are not necessarily the best place for scientific disputes to be aired and resolved.

G M ADDISON
Department of Clinical Biochemistry,
Royal Manchester Children’s Hospital, Pendlebury,
Manchester M27 4HA, UK

References

Sudden infant death syndrome (SIDS): the search for the cause.

G M Addison

J Clin Pathol 1998 51: 419
doi: 10.1136/jcp.51.6.419