Education and debate

Should we screen for gestational diabetes?

R J Jarrett wrote to us arguing that gestational diabetes was a muddled concept that provided uncertain benefits for mother and infant. We invited Richard Beard and colleagues to put the other side of the argument and make the case for screening all pregnant women for gestational diabetes

"The concept of gestational diabetes was popularised before considerations of evidence based medicine came on the scene" [1]

R J Jarrett

45 Bishopsthorpe Road, London SE26 4PA
R J Jarrett
emeritus professor of clinical epidemiology

BMJ 1997;315:736–7

Much confusion surrounds the topic of screening for glucose intolerance-hyperglycaemia during pregnancy in terms of who should be screened, how to screen, and the management of those with positive results.[2-4] Confusion arises from lack of or poor quality evidence, compounded in this instance by a concept (gestational diabetes mellitus) founded on risk of subsequent non-insulin dependent diabetes mellitus rather than outcome of the index pregnancy.[5] In addition the criteria for gestational diabetes prescribe a minimum, but not a maximum, level of glucose intolerance, so that any group of women labelled as having gestational diabetes might contain some with glycaemia in the range that qualifies for a diagnosis of non-insulin dependent diabetes, rendering comparisons of different series impossible. Coustan, whose comment is reproduced in my title, suggested four questions which required answers to achieve resolution[1]:

(1) How severe must maternal hyperglycaemia be to measurably worsen pregnancy outcome?
(2) Can we intervene to prevent adverse outcomes?
(3) Is such intervention cost effective?
(4) If so, what is the most appropriate way of screening and detecting the problem?

Severity of maternal hyperglycaemia

Women with pre-existing diabetes, either insulin dependent or non-insulin dependent, undoubtedly have an increased risk of bearing a child with a congenital abnormality and this risk is related to the degree of hyperglycaemia during embryogenesis.[6 7] However, this is not relevant to screening at typical booking times as embryogenesis is complete by week seven of gestation. By contrast, gestational diabetes as defined[5] is not associated with risk of congenital abnormalities,[8] despite the presence of some women with glucose intolerance sufficient to qualify them for a diagnosis of non-insulin dependent diabetes.

The only commonly (though not absolutely consistently) reported "complication" of gestational diabetes is macrosomia, a rather emotive description of a newborn infant with a birth weight in the upper centiles (variously defined) of the distribution. To what extent birth weight is determined by maternal glycaemia is debated, but the relation is confounded by maternal fatness.[9] A very large baby is more likely to give rise to obstetric problems and to acquire a birth injury, but one estimate suggested that about 4% of women with untreated gestational diabetes would deliver infants weighing 4500 g or more compared with about 2% of the general obstetric population.[10] While there is some evidence that treatment can reduce fetal weight,[11] this cannot be automatically assumed to be justified given the data showing an inverse association between birth weight and the incidence of disorders in later life.[12]

Are there any adverse effects of the diagnosis of gestational diabetes? Women with gestational diabetes are more likely to be delivered by caesarean section. This has been attributed to their higher proportion of large babies, but in a recent study the section rate was higher even though the proportion of large babies was not,[1] supporting the view that the diagnostic label sensitises obstetricians. The gestational diabetes label also leads to the necessity of self monitoring of blood glucose and possibly insulin injections. The possible distress due to screening and treatment in someone who previously thought herself to be healthy has not been investigated.

Intervention

There is only one clinical trial of any merit.[11] In this 66 women, mostly Hispanic and including an appreciable (though unstated) number with undiagnosed non-insulin dependent diabetes, were treated at random with either a more or a less intensive insulin regimen. Birth weights were, on average, nearly 400 g less in the intensively treated group, but caesarean section rates were not significantly different. Whether treatment influences any outcome of pregnancy in women discovered to have non-insulin dependent diabetes during pregnancy has not otherwise been subject to clinical trial. Indeed, pregnancy associated with non-insulin dependent diabetes has attracted little research interest.

Screening tests and cost effectiveness

Screening for hyperglycaemia is bedevilled by the lack of a suitable screening test.[13] If sensitivity is important then some kind of glucose tolerance test is essential to identify gestational impaired glucose tolerance. Oral glucose tolerance tests are, however, tedious to perform and poorly reproducible. Single blood tests, such as glycated haemoglobin and fructosamine, cannot even identify the lower range of non-insulin dependent diabetes glycaemia, let alone gestational impaired glucose tolerance, though they could identify more florid hyperglycaemia.

"Screening is bedevilled by the lack of a suitable test."

The only relevant data available on cost effectiveness concern the yield of screening for non-insulin dependent diabetes using the World Health Organisation's "epidemiological" criterion—a plasma glucose value over 11.0 mmol/l two hours after a 75 g oral glucose load. Two studies provide minimal estimates of incidence of 4/10 000 for Europid women[14] and 18/10 000 for south Asian women.[15] As stated earlier, there is no good evidence of any undoubted benefit for the index pregnancy, though there are putative (but uncosted) benefits in having one's diabetes diagnosed early.[13]

Conclusions

The ethics of screening require the screener to show the likelihood of benefit from screening. No clear benefit has been shown from screening for glucose intolerance-hyperglycaemia (at least for the woman being screened) during pregnancy, and there are disadvantages, which include the acquisition of disease status and an increased risk of caesarean section. It is argued that screening to identify someone at risk of subsequent non-insulin dependent diabetes or with undiagnosed non-insulin dependent diabetes is a good thing. If so, it should be available to all adults, not only pregnant women. However, the most recent review of population screening for non-insulin dependent diabetes[13] advocates extensive and varied further research on all aspects of the question. This was in the context of non-pregnant adults, but the same requirements apply to screening in pregnancy before it can be regarded as justified.

Since submitting this article I have noted two American groups which do not recommend screening for gestational diabetes. The American College of Obstetricians and Gynecologists, which in 1986 recommended selective screening, in 1994 noted the absence of data to support screening and did not make a specific recommendation.[16] The US Preventive Services Task Force cites insufficient evidence for or against screening for gestational diabetes.[17] In contrast, an expert committee of the American Diabetes Association continued to recommend screening, though no longer without some degree of selection.[18]

1 Coustan DR. Management of gestational diabetes mellitus: a self-fulfilling prophecy? *JAMA* 1996;275:1199-1200.
2 Jarrett RJ. Gestational diabetes: a non-entity? *BMJ* 1993;306:37-8.
3 Gabbe SG, Landon MB. Management of diabetes in pregnancy: survey of materno-fetal specialists in the United States. In: Sutherland HW, Stowers JM, Pearson DWM, eds. *Carbohydrate metabolism in pregnancy and the newborn IV*. London: Springer,1989: 309-17.
4 Nelson-Piercy C, Gale EAM. Do we know how to screen for gestational diabetes? Current practice in one regional health authority. *Diabetic Med* 1994;11:493-8.
5 O'Sullivan JB, Mahan C. Criteria for the oral glucose tolerance test in pregnancy. *Diabetes* 1964;13:278-85.
6 Hanson U, Persson B, Thunell S. The relation between HbA_{1c} in early diabetic pregnancy and the occurrence of spontaneous abortion and malformation in Sweden. *Diabetologia* 1990;33:100-4.
7 Towner D, Kjos SL, Leung B, Montoro MM, Xiang A, Mestman JH, et al. Congenital malformations in pregnancies complicated by NIDDM. *Diabetes Care* 1995;18:1446-51.
8 Molsted-Pedersen L, Damon P, Buschard K. Diabetes diagnosed during pregnancy: follow-up studies. In: In: Sutherland HW, Stowers JM, Pearson DWM, eds. *Carbohydrate metabolism in pregnancy and the newborn IV*. London: Springer,1989: 277-86.
9 Goldenburg RL, Tamura TT. Pre-pregnancy weight and pregnancy outcome. *JAMA* 1996;275:1127-8.
10 Ales KL, Santini DL. Should all pregnant women be screened for gestational glucose intolerance? *Lancet* 1989;i:1187-91.
11 de Veciana M, Major CA, Morgan MA, Asrat T, Toohey JS, Lien JM, et al. Postprandial versus preprandial blood glucose monitoring in women with gestational diabetes mellitus requiring insulin therapy. *N Engl J Med* 1995;333:1237-41.
12 Barker DJP. *Fetal and infant origins of adult disease*. London: BMJ Publishing Group, 1992.
13 Engelgau MM, Aubert RE, Thompson TJ, Herman WH. Screening for NIDDM in nonpregnant adults: a review of principles, screening tests, and recommendations. *Diabetes Care* 1995;18:1606-17.
14 Roberts RN, Moohan JM, Foo RLK, Harley JMG, Traub AI, Hadden DR. Fetal outcome in mothers with IGT in pregnancy. *Diabetic Med* 1993;10:438-43.
15 Samanta A, Burden ML, Burden AC, Jones GR. Glucose tolerance during pregnancy in Asian women. *Diabetes Res Clin Prac* 1989;7:127-35.
16 American College of Obstetricians and Gynecologists. *Management of diabetes mellitus in pregnancy*. Washington: ACOG, 1994.
17 US Preventive Services Task Force. *Guide to clinical preventive services*. 2nd ed. Baltimore: Williams and Wilkins, 1996.
18 Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 1997;7:1183-97.

(Accepted 4 March 1997)

The case for screening for gestational diabetes

Jacquiline de A C Soares, Anne Dornhorstt, Richard W Beard

Screening for gestational diabetes is imperative but we need to refocus away from short term improvements in obstetric outcomes to more important medium and long term health benefits. Today 5% of United Kingdom[1] and 12% of United States[2] total healthcare expenditure is spent on diabetes and its complications. There is a global epidemic of non-insulin dependent diabetes, and radical preventive measures are required if morbidity and mortality from diabetes are to fall. We would ask whether we can afford *not* to screen for gestational diabetes.

Identifying future diabetics

Detecting gestational diabetes identifies women at risk of future non-insulin dependent diabetes.[3 4] The success of treating non-insulin dependent diabetes is severely hampered by the high percentage of diabetic complications present at diagnosis,[5] so earlier diagnosis is important in preventing complications.[6] Half of all people with non-inuslin dependent diabetes are clinically undiagnosed, and diagnosis takes, on average, seven years from onset of the disorder.[7]

Department of Obstetrics and Gynaecology, Imperial College School of Medicine at St Mary's Hospital, London W2 NY1

continued over

BMJ 1997;315:737–9

Jacqueline de A C Soares, *research assistant*
Richard W Beard, *professor*
Department of Endocrinology, Imperial College School of Medicine at Hammersmith Hospital, London W12 0NN
Anne Dornhorstt *honorary senior lecturer*
Correspondence to: Professor Beard

The rate of progression to non-insulin dependent diabetes mellitus after a pregnancy where the woman has had gestational diabetes depends predominantly on ethnicity and the degree of glucose intolerance both in pregnancy and immediately afterwards.[8] Other contributing factors are weight during pregnancy and subsequent weight gain, age, parity, and family history.[8 9] In high risk populations, such as Hispanic American women, about 40% of women with gestational diabetes develop diabetes within six years, which rises to 70% among those with impaired glucose tolerance (by World Health Organisation criteria) after birth.[10] In white Europeans the rate of progression to diabetes is slower—20-40% within 20 years.[11] Identifying women who are at risk while they are still young provides an opportunity to identify the disease in subsequent pregnancies and to modify the natural history of non-insulin diabetes mellitus.

The onset of non-insulin dependent diabetes in women can be delayed by weight control and exercise, the benefits being greatest among obese women with a family history of diabetes.[12] Whether further benefit can be obtained with a change of lifestyle or drugs is currently being evaluated in the United States. If this study does show that intervention can delay diabetes in these women the long term benefits of screening for gestational diabetes will be further established.

Effects on the child

The consequences of gestational diabetes on the health of the child have until recently received little attention. The recent explosion of studies underlining the importance of the intrauterine environment for future adult chronic diseases[13] has challenged the sceptics' view that in pregnancy lesser degrees of maternal glucose intolerance, not justifying treatment of the non-pregnant woman, are irrelevant.[14] The concept that diabetes begets diabetes through an intrauterine effect on the fetal pancreas,[15] additional to any genetic effect, is strongly supported by animal and human epidemiological studies.[16 17]

In support of this hypothesis are studies within the Pima population, where children of a diabetic mother are at greater risk of diabetes and childhood obesity than older siblings born before their mother become diabetic.[18] Maternal carbohydrate metabolism may also influence future human fetal insulin secretion and function as suggested by the studies on black and white American adolescents. Those born to diabetic mothers have greater insulin resistance and are more likely to be glucose intolerant during puberty.[19] The literature supports an effect of maternal hyperglycaemia on an infant's future susceptibility to abnormalities of carbohydrate metabolism.[16] However, the critical threshold of hyperglycaemia is currently not known.

Poor maternal diabetic control is associated with an increased risk of large for gestational age infants.[20] In certain ethnic groups up to half of all pregnancies where gestational diabetes is present have evidence of accelerated fetal growth.[21] The apparent paradox of low birth weight associated with an increased risk of future diabetes and high birth weight with a decreased risk[22] has distracted attention from the knowledge of the potential harm in later life of accelerated intrauterine fetal growth associated with gestational diabetes. In a high risk population with a 5% prevalence of gestational diabetes up to half the infants above the 90th birthweight percentile theoretically could be from a diabetic pregnancy. In such populations both high and low birth weights are associated with an increased risk of diabetes in later life.[23] By contrast, when the prevalence of gestational diabetes is low (0.5-1%) at most only 5-10% of all infants above the 90th birthweight percentile could be the result of maternal diabetes. Thus in low risk populations this argument for universal screening is not so strong.

Demonstrating benefit

The short term benefits of screening and treating gestational diabetes have focused on pregnancy outcome. In high risk populations, with a high background prevalence of diabetes combined with limited access to medical and perinatal care, perinatal mortality can be seen to improve after screening for and treatment of gestational diabetes.[24] Retrospective studies suggest a benefit on stillbirth rates after the introduction of screening and treating gestational diabetes in low risk populations,[25] but demonstrating a benefit on perinatal mortality in prospective trials has proved more difficult. In Western populations, with a low prevalence of diabetes, good access to medical care, and low perinatal mortality and morbidity rates, there are ethical constraints in mounting randomised trials with sufficient power to test whether treating gestational diabetes reduces perinatal morbidity. Prospective studies in these populations have therefore assessed pregnancy outcome using surrogate markers of diabetic control. These include macrosomia, need for caesarean section, and fetal hypoglycaemia. None of these end points are specific for diabetes and many are influenced by the practice of individual obstetricians, maternal obesity, age, and parity.

These difficulties should not, however, detract from the fact that maternal hyperglycaemia is the cause of a diabetic fetopathy syndrome of Pedersen. With the knowledge that a baby showing evidence of this syndrome may well develop diabetes in later life, there

ANITA CORBIN/IMPACT

Screening for gestational diabetes: should we do it?

is a good case for early detection and treatment of the mother. Evidence of increased visceral fat and enlargement of the liver, spleen, and heart may be apparent on ultrasound as early as 28 weeks' gestation. These accelerated growth patterns, associated with gestational diabetes, can be corrected with diet or insulin, which results in fewer large for gestational age infants and fewer operative deliveries.[26] To what extent, if any, abnormal fetal growth patterns due to hyperinsulinaemia reflect aberrant fetal cell programming and what influence this may have on the future insulin sensitivity and adult health remains speculative. To ignore such a link in the face of mounting animal evidence would, however, be short sighted.

Who should be screened?

Should all pregnant women be screened or only those at risk? The answer needs to reflect the ethnicity of the population, the availability of health care, and the economic and medicolegal expectations of the country. Once the decision has been made to screen a reproducible screening test needs to be chosen that is sensitive, specific, and easily applied.

The most universally researched screening test is the O'Sullivan test, which involves a one hour timed blood glucose sample after a 50 g oral glucose load, a value $\geq$7.8 mmol/l being positive.[27] This test has a $\approx$95% sensitivity and $\approx$85% specificity for detecting pregnancy induced glucose intolerance that occurs at 20-28 weeks' gestation. A first trimester test is advisable in high risk populations, in which more women will have gestational diabetes before 20 weeks. This can be done with either the 50 g oral glucose load [28] or a timed plasma glucose value, which will identify women with glucose intolerance likely to require insulin treatment—namely, a fasting plasma glucose concentration $>$6 mmol or a 2 hour postprandial value $>$9 mmol/l.[20] Other screening tests—which include random glucose values, glucosuria, fructosamine, diurnal glucose profiles, and glucose responses to mixed meals—have been less extensively evaluated in pregnancy than the O'Sullivan test, which remains the gold standard. The sensitivity of purely clinical risk factors is poor, $<$70%, especially in multiethnic populations, since they do not include ethnicity.[27 29]

The confirmational diagnostic test for gestational diabetes remains controversial. Gestational diabetes is usually diagnosed on the basis of an oral glucose tolerance test. However, the exact load administered (50, 75, or 100 g) varies between centres.[30] The need for one test and one set of diagnostic criteria is recognised. Epidemiologically the 75 g oral glucose tolerance test has the advantage that it is internationally used outside pregnancy. However, the diagnostic limits at which treatment is required still need to be defined.

A dogmatic stand against screening for gestational diabetes not only ignores the proved benefits of treatment on perinatal outcome but also denies affected mothers the possibility to reduce their own and their babies' risk of later diabetes.

1 Laing W, Williams R. *Diabetes, a model for health care management.* London: Office of Health Economics, 1989.
2 American Diabetes Association. *Diabetes 1996: vital statistics.* Alexandria, VA: ADA, 1996: 67-9.
3 O'Sullivan JB. Body weight and subsequent diabetes mellitus. *JAMA* 1982;248:949-52.
4 Dornhorst A, Bailey PC, Anyaoku V, Elkeles RS, Johnston DG, Beard RW. Abnormalities of glucose tolerance following gestational diabetes. *Q J Med* 1990;284:1219-28.
5 UK Prospective Diabetes Study VI. Complications in newly diagnosed Type 2 diabetic patients and their association with different clinical and biochemical risk factors. *Diabetes Res* 1990;13:1-11.
6 Eastman RC, Silverman R. Harris M, Jaui HJC, Chiang VP, Gorden P. Lessening the burden of diabetes. *Diabetes Care* 1993;16:1095-1102.
7 King H, Rewers M. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. *Diabetes Care* 1993;16:157-77.
8 Metzger BE, Cho NH, Roston SM, Radvany R. Prepregnancy weight and antepartum insulin secretion predict glucose tolerance five years after gestational diabetes mellitus. *Diabetes Care* 1993;16:1598- 1605.
9 Peters RK, Kjos SL, Xiang A, Buchanan TA. Long-term diabetogenic effect of single pregnancy in women with previous gestational diabetes mellitus. *Lancet* 1996;347:227-30.
10 Kjos SL, Peters RK, Xiang A, Henry OA, Montoro M, Buchanon TA. Predicting future diabetes in Latino women with gestational diabetes. *Diabetes* 1995;44:586-91.
11 Henry OA, Beischer NA, Sheedy MT, Walstab JE. Gestational diabetes and follow-up among immigrant Vietnam-born women. *Aust N Z J Obstet Gynaecol* 1993;33:109-14.
12 Manson JE, Rimm EB, Stampfer MJ, Colditz GA, Willett WC, Krolewski AS, et al. Physical activity and incidence of NIDDM women. *Lancet* 1991;338:774-8.
13 Barker DJP. Fetal growth and adult disease. *Br J Obstet Gynaecol* 1992;99:275-6.
14 Jarrett RJ. Gestational diabetes: a non-entity? *BMJ* 1993;306:37-8.
15 Freinkel N. Banting Lecture 1980: of pregnancy and progeny. *Diabetes* 1980;29:1023-35.
16 Aerts L, Sodoyez-Goffaux F, Sodoyez JC, Malaisse WJ, Van Assche FA. The diabetic intrauterine milieu has a long lasting effect on insulin secretion by B cells and on insulin uptake by target tissues. *Am J Obstet Gynecol* 1988;159:1287-92.
17 Alcolado JC, Alcolado R. Importance of maternal history of non-insulin-dependent diabetic patients. *BMJ* 1991;302:1178-80.
18 Pettitt D, Aleck K, Baird H, Carraher M, Bennet B, Knowler W, et al. Congenital susceptibility to NIDDM: role of intrauterine environment. *Diabetes* 1988;37:622-28.
19 Silverman B, Metzger B, Cho N, Loeb C. Impaired glucose tolerance in adolescent offspring of diabetic mothers. Relationship to fetal hyperinsulinism. *Diabetes Care* 1995;18:611-7.
20 Maresh M, Beard RW, Bray CS, Elkeles RS, Wadsworth J. Factors predisposing to and outcome of gestational diabetes. *Obstet Gynecol* 1989;74:342-6.
21 Dornhorst A, Nicholls JSD, Welch A, Ali K, Chan SP, Beard RW. Correcting for ethnicity when defining large-for-gestational-age infants in diabetic pregnancies. *Diabetic Med* 1996;13:226-31.
22 Barker DJP, Hales CN, Fall CHD, Osmond C, Phipps K, Clark PMS. Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X: relation to reduced fetal growth). *Diabetologia* 1993;36:62-7.
23 McCance DR, Pettitt DJ, Hanson RL, Jacobsson LTH, Knowler WC, Bennett PH. Birth weight and non-insulin dependent diabetes: thrifty genotype, thrifty phenotype, or surviving small baby genotype? *BMJ* 1994;308:942-5.
24 Huddle K, England M, Nagar A. Outcome of pregnancy in diabetic women in Soweto, South Africa. *Diabetic Med* 1993;10:290-4.
25 O'Sullivan JB, Mahan CM, Charles D, Dandrow RV. Gestational diabetes and perinatal mortality rate. *Am J Obstet Gynecol* 1973;116: 901-4.
26 Buchanan TA, Kjos SL, Montoro MN, Wu PV, Madilejo NG, Gonzalez M, et al. Use of fetal ultrasound to select metabolic therapy for pregnancies complicated by mild gestational diabetes. *Diabetes Care* 1994;17:275-83.
27 O'Sullivan JB, Mahon CM, Charles D, Dandrow RV. Screening criteria for high-risk gestational diabetic patients. *Am J Obset Gynecol* 1973;116:894-5.
28 Super DM, Edelberg SC, Philipson E H, Hertz RH, Kalhan SC. Diagnosis of gestational diabetes in early pregnancy. *Diabetes Care* 1991;14:288-94.
29 Weeks JW, Major CA, deVeciana M, Morgan MA. Gestational diabetes: does the presence of risk factors influence perinatal outcome? *Am J Obstet Gynecol* 1994;171:1003-7.
30 Nelson-Piercy C, Gale EAM. Do we know how to screen for gestational diabetes? Current practice in one regional health authority. *Diabetic Med* 1993;11:493-8.

Endpiece

Dr Johnson doubts the logic of physicians

It is incident to physicians, I am afraid, beyond all other men, to mistake subsequence for consequence.

From a review of Dr Lucas's *Essay on Waters* (1734)

How to read a paper

Papers that go beyond numbers (qualitative research)

Trisha Greenhalgh, Rod Taylor

This is the last in a series of 10 articles introducing non-experts to finding medical articles and assessing their value

Unit for Evidence-Based Practice and Policy, Department of Primary Care and Population Sciences, University College London Medical School/ Royal Free Hospital School of Medicine, Whittington Hospital, London N19 5NF

Trisha Greenhalgh, *senior lecturer*

Exeter and Devon Research and Development Support Unit, Postgraduate Medical School, Wonford, Exeter EX2 5EQ

Rod Taylor, *senior lecturer*

Correspondence to: Dr Greenhalgh p.greenhalgh@ucl.ac.uk

BMJ 1997;315:740–3

What is qualitative research?

Epidemiologist Nick Black has argued that a finding or a result is more likely to be accepted as a fact if it is quantified (expressed in numbers) than if it is not.[1] There is little or no scientific evidence, for example, to support the well known "facts" that one couple in 10 is infertile, or that one man in 10 is homosexual. Yet, observes Black, most of us are happy to accept uncritically such simplified, reductionist, and blatantly incorrect statements so long as they contain at least one number.

Researchers who use qualitative methods seek a deeper truth. They aim to "study things in their natural setting, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them,"[2] and they use "a holistic perspective which preserves the complexities of human behaviour."[1]

Questions such as "How many parents would consult their general practitioner when their child has a mild temperature?" or "What proportion of smokers have tried to give up?" clearly need answering through quantitative methods. But questions like "Why do parents worry so much about their children's temperature?" and "What stops people giving up smoking?" cannot and should not be answered by leaping in and measuring the first aspect of the problem that we (the outsiders) think might be important. Rather, we need to listen to what people have to say, and we should explore the ideas and concerns which the subjects themselves come up with. After a while, we may notice a pattern emerging, which may prompt us to make our observations in a different way. We may start with one of the methods shown in box 1, and go on to use a selection of others.

Box 2 summarises (indeed, overstates) the differences between the qualitative and quantitative approaches to research. In reality, there is a great deal of overlap between them, the importance of which is increasingly being recognised.[4]

Quantitative research should begin with an idea (usually articulated as a hypothesis), which then, through measurement, generates data and, by deduction, allows a conclusion to be drawn. Qualitative research, in contrast, begins with an intention to explore a particular area, collects "data" (observations and interviews), and generates ideas and hypotheses from these data largely through what is known as inductive reasoning.[3] The strength of the quantitative approach lies in its reliability (repeatability)—that is, the same measurements should yield the same results time after time. The strength of qualitative research lies in validity (closeness to the truth)—that is, good qualitative research, using a selection of data collection methods, really should touch the core of what is going on rather than just skimming the surface. The validity of qualitative methods is greatly improved by using a combination of research methods, a process known as triangulation, and by independent analysis of the data by more than one researcher.

The so called iterative approach (altering the research methods and the hypothesis as the study progresses, in the light of information gleaned along the way) used by qualitative researchers shows a commendable sensitivity to the richness and variability of the subject matter. Failure to recognise the legitimacy of this approach has, in the past, led critics to accuse qualitative researchers of continually moving their own goalposts. Though these criticisms are often misguided, there is, as Nicky Britten and colleagues have observed, a real danger "that the flexibility [of the iterative approach] will slide into sloppiness as the researcher ceases to be clear about what it is (s)he is investigating."[5] These authors warn that qualitative researchers must, therefore, allow periods away from their fieldwork for reflection, planning, and consultation with colleagues.

Summary points

- Qualitative methods aim to make sense of, or interpret, phenomena in terms of the meanings people bring to them
- Qualitative research may define preliminary questions which can then be addressed in quantitative studies
- A good qualitative study will address a clinical problem through a clearly formulated question and using more than one research method (triangulation)
- Analysis of qualitative data can and should be done using explicit, systematic, and reproducible methods

Box 1
Examples of qualitative research methods

Documents—Study of documentary accounts of events, such as meetings

Passive observation—Systematic watching of behaviour and talk in natural occurring settings

Participant observation—Observation in which the researcher also occupies a role or part in the setting, in addition to observing

In depth interviews—Face to face conversation with the purpose of exploring issues or topics in detail. Does not use preset questions, but is shaped by a defined set of topics

Focus groups—Method of group interview which explicitly includes and uses the group interaction to generate data

Evaluating papers that describe qualitative research

By its very nature, qualitative research is non-standard, unconfined, and dependent on the subjective experience of both the researcher and the researched. It explores what needs to be explored and cuts its cloth accordingly. It is debatable, therefore, whether an all-encompassing critical appraisal checklist along the lines of the Users' Guides to the Medical Literature[6-19] could ever be developed. Our own view, and that of a number of individuals who have attempted, or are currently working on, this very task,[3 5] is that such a checklist may not be as exhaustive or as universally applicable as the various guides for appraising quantitative research, but that it is certainly possible to set some ground rules. The list which follows has been distilled from the published work cited earlier,[2 3 5] and also from our own research and teaching experiences. You should note, however, that there is a great deal of disagreement and debate about the appropriate criteria for critical appraisal of qualitative research, and the ones given here are likely to be modified in the future.

Question 1: Did the paper describe an important clinical problem addressed via a clearly formulated question?

A previous article in this series explained that one of the first things you should look for in any research paper is a statement of why the research was done and what specific question it addressed.[20] Qualitative papers are no exception to this rule: there is absolutely no scientific value in interviewing or observing people just for the sake of it. Papers that cannot define their topic of research more closely than "We decided to interview 20 patients with epilepsy" inspire little confidence that the researchers really knew what they were studying or why.

You might be more inclined to read on if the paper stated in its introduction something like, "Epilepsy is a common and potentially disabling condition, and up to 20% of patients do not remain free of fits while taking medication. Antiepileptic medication is known to have unpleasant side effects, and several studies have shown that a high proportion of patients do not take their tablets regularly. We therefore decided to explore patients' beliefs about epilepsy and their perceived reasons for not taking their medication."

Question 2: Was a qualitative approach appropriate?

If the objective of the research was to explore, interpret, or obtain a deeper understanding of a particular clinical issue, qualitative methods were almost certainly the most appropriate ones to use. If, however, the research aimed to achieve some other goal (such as determining the incidence of a disease or the frequency of an adverse drug reaction, testing a cause and effect hypothesis, or showing that one drug has a better risk-benefit ratio than another), a case-control study, cohort study, or randomised trial may have been better suited to the research question.[19]

Question 3: How were the setting and the subjects selected?

The second box contrasts the statistical sampling methods of quantitative research with theoretical methods of qualitative research. In quantitative research, it is vital to ensure that a truly random sample of subjects is recruited so that the results reflect, on average, the condition of the population from which that sample was drawn.

Box 2
Qualitative versus quantitative research—the overstated dichotomy

	Qualitative	Quantitative
Social theory	Action	Structure
Methods	Observation, interview	Experiment, survey
Question	What is X? (classification)	How many Xs? (enumeration)
Reasoning	Inductive	Deductive
Sampling method	Theoretical	Statistical
Strength	Validity	Reliability

Reproduced with permission from Mays and Pope, *Qualitative Research in Health Care*[3]

In qualitative research, however, we are not interested in an "on average" view of a patient population. We want to gain an in depth understanding of the experience of particular individuals or groups; we should therefore deliberately seek out individuals or groups who fit the bill. If, for example, we wished to study the experience of non-English speaking British Punjabi women when they gave birth in hospital (with a view to tailoring the interpreting or advocacy service more closely to the needs of this patient group), we would be perfectly justified in going out of our way to find women who had had a range of different birth experiences—an induced delivery, an emergency caesarean section, a delivery by a medical student, a late miscarriage, and so on—rather than a "random" sample of British Punjabi mothers.

Question 4: What was the researcher's perspective, and has this been taken into account?

It is important to recognise that there is no way of abolishing, or fully controlling for, observer bias in qualitative research. This is most obviously the case when participant observation is used, but it is also true for other forms of data collection and of data analysis. If, for example, the research concerns the experience of asthmatic adults living in damp and overcrowded housing and the perceived effect of these surroundings on their health, the data generated by techniques

PETER BROWN

such as focus groups or semistructured interviews are likely to be heavily influenced by what the interviewer believes about this subject and by whether he or she is employed by the hospital chest clinic, the social work department of the local authority, or an environmental pressure group. But since it is inconceivable that the interviews could have been conducted by someone with no views at all and no ideological or cultural perspective, the most that can be required of the researchers is that they describe in detail where they are coming from so that the results can be interpreted accordingly.

Question 5: What methods did the researcher use for collecting data—and are these described in enough detail?

I once spent two years doing highly quantitative, laboratory based experimental research in which around 15 hours of every week were spent filling or emptying test tubes. There was a standard way to fill the test tubes, a standard way to spin them in the centrifuge, and even a standard way to wash them up. When I finally published my research, some 900 hours of drudgery was summed up in a single sentence: "Patients' serum rhubarb levels were measured according to the method described by Bloggs et al [reference to Bloggs et al's published paper]."

The methods section of a qualitative paper often cannot be written in shorthand or dismissed by reference to someone else's research techniques. It may have to be lengthy and discursive since it is telling a unique story without which the results cannot be interpreted. As with the sampling strategy, there are no hard and fast rules about exactly what details should be included in this section of the paper. You should simply ask, "have I been given enough information about the methods used?", and, if you have, use your common sense to assess, "are these methods a sensible and adequate way of addressing the research question?"

Question 6: What methods did the researcher use to analyse the data—and what quality control measures were implemented?

The data analysis section of a qualitative research paper is where sense can most readily be distinguished from nonsense. Having amassed a thick pile of completed interview transcripts or field notes, the genuine qualitative researcher has hardly begun. It is simply not good enough to flick through the text looking for "interesting quotes" which support a particular theory. The researcher must find a systematic way of analysing his or her data, and, in particular, must seek examples of cases which appear to contradict or challenge the theories derived from the majority.

One way of doing this is by content analysis: drawing up a list of coded categories and "cutting and pasting" each segment of transcribed data into one of these categories. This can be done either manually or, if large amounts of data are to be analysed, via a tailor-made computer database. The statements made by all the subjects on a particular topic can then be compared with one another, and more sophisticated comparisons can be made such as "did people who made statement A also tend to make statement B?"

In theory, the paper will show evidence of "quality control"—that is, the data (or at least, a sample of them) will have been analysed by more than one researcher to confirm that they are both assigning the same meaning to them, although in practice this is often difficult to achieve. Indeed, when researching this article, we could find no data on the interobserver reliability of any qualitative study to illustrate this point.

Question 7: Are the results credible, and if so, are they clinically important?

We obviously cannot assess the credibility of qualitative results through the precision and accuracy of measuring devices, nor their significance via confidence intervals and numbers needed to treat. It usually takes little more than plain common sense to determine whether the results are sensible and believable, and whether they matter in practice.

One important aspect of the results section to check is whether the authors cite actual data. Claims such as "general practitioners did not usually recognise the value of audit" would be infinitely more credible if one or two verbatim quotes from the interviewees were reproduced to illustrate them. The results should be independently and objectively verifiable—after all, a subject either made a particular statement or (s)he did not—and all quotes and examples should be indexed so that they can be traced back to an identifiable subject and setting.

Question 8: What conclusions were drawn, and are they justified by the results?

A quantitative research paper should clearly distinguish the study's results (usually a set of numbers) from the interpretation of those results (the discussion). The reader should have no difficulty separating what the researchers *found* from what they think it *means*. In qualitative research, however, such a distinction is rarely possible, since the results are by definition an interpretation of the data.

It is therefore necessary, when assessing the validity of qualitative research, to ask whether the interpretation placed on the data accords with common sense and is relatively untainted with personal or cultural perspective. This can be a difficult exercise, because the language we use to describe things tends to impugn meanings and motives which the subjects themselves may not share. Compare, for example, the two statements, "three women went to the well to get water" and "three women met at the well and each was carrying a pitcher."

It is becoming a cliché that the conclusions of qualitative studies, like those of all research, should be "grounded in evidence"—that is, that they should flow from what the researchers found in the field. Mays and Pope suggest three useful questions for determining whether the conclusions of a qualitative study are valid:

- how well does this analysis explain why people behave in the way they do?;
- how comprehensible would this explanation be to a thoughtful participant in the setting?; and
- how well does the explanation cohere with what we already know?[3]

Question 9: Are the findings of the study transferable to other clinical settings?

One of the commonest criticisms of qualitative research is that the findings of any qualitative study pertain only to the limited setting in which they were obtained. In

fact, this is not necessarily any truer of qualitative research than of quantitative research. Look back at the example of British Punjabi women described above. You should be able to see that the use of a true *theoretical* sampling frame greatly increases the transferability of the results over a "convenience" sample.

Conclusion

Doctors have traditionally placed high value on numerical data, which may in reality be misleading, reductionist, and irrelevant to the real issues. The increasing popularity of qualitative research in the biomedical sciences has arisen largely because quantitative methods provided either no answers or the wrong answers to important questions in both clinical care and service delivery.[1] If you still feel that qualitative research is necessarily second rate by virtue of being a "soft" science, you should be aware that you are out of step with the evidence.

In 1993, Pope and Britten presented a paper to the BSA Medical Sociology Group conference entitled "Barriers to qualitative methods in the medical mindset," in which they showed their collection of rejection letters from biomedical journals. The letters revealed a striking ignorance of qualitative methodology on the part of reviewers. In other words, the people who had rejected the papers often seemed to be incapable of distinguishing good qualitative research from bad. Somewhat ironically, qualitative papers of poor quality now appear regularly in some medical journals, whose editors have climbed on the qualitative bandwagon without gaining an ability to appraise such papers. Note, however, that the critical appraisal of qualitative research is a relatively underdeveloped science, and the questions posed in this chapter are still being refined.

Thanks to Professor Nick Black for advice on this article.

1 Black N. Why we need qualitative research. *J Epidemiol Community Health* 1994;48:425-6.
2 Denkin NK, Lincoln YS, eds. *Handbook of qualitative research.* London: Sage, 1994.
3 Mays N, Pope C, eds. *Qualitative research in health care.* London: BMJ Publishing Group, 1996.
4 Abell P. Methodological achievements in sociology over the past few decades with specific reference to the interplay of qualitative and quantitative methods. In: Bryant C, Becker H, eds. *What has sociology achieved?* London: Macmillan, 1990.

The articles in this series are excerpts from *How to read a paper: the basics of evidence based medicine.* The book includes chapters on searching the literature and implementing evidence based findings. It can be ordered from the BMJ Publishing Group: tel 0171 383 6185/6245; fax 0171 383 6662. Price £13.95 UK members, £14.95 non-members.

5 Britten N, Jones R, Murphy E, Stacy R. Qualitative research methods in general practice and primary care. *Fam Pract* 1995;12:104-14.
6 Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. *JAMA* 1993;270:2093-5.
7 Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993;270:2598-601.
8 Guyatt GH, Sackett DL, Cook DJ. Users' guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;271:59-63.
9 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? *JAMA* 1994;271:389-91.
10 Jaeschke R. Guyatt G. Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;271:703-7.
11 Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. *JAMA* 1994;271:1615-9.
12 Laupacis A. Wells G. Richardson WS. Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994;271:234-7.
13 Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. *JAMA* 1994;272:1367-71.
14 Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. A. Are the results of the study valid? *JAMA* 1995;273:1292-5.
15 Richardson WS, Detsky AS. Users' guides to the medical literature. VII. How to use a clinical decision analysis. B. What are the results and will they help me in caring for my patients? *JAMA* 1995;273:1610-3.
16 Hayward RSA, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? *JAMA* 1995;274:570-4.
17 Wilson MC, Hayward RS, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. B. Will the recommendations help me in caring for my patients? *JAMA* 1995;274:1630-2.
18 Drummond MF, Richardson WS, O'Brien BJ, Levine M, Heyland D. Users' guides to the medical literature XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? *JAMA* 1997;277:1552-7.
19 O'Brien BJ, Heyland D, Richardson WS, Levine M, Drummond MF. Users' guides to the medical literature XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? *JAMA* 1997;277:1802-6.
20 Greenhalgh T. Getting your bearings (deciding what the paper is about). *BMJ* 1997;315:243-6.
21 Kinmonth A-L. Understanding and meaning in research and practice. *Fam Pract* 1995;12:1-2.

Any questions

Use of a statin for reducing cholesterol levels

If a patient with coronary disease already drinks a glass of wine and eats a piece of fruit daily, eats fish several times a week, dresses salads with olive oil, exercises regularly, and takes a β blocker and vitamin E is it still worth while prescribing a statin to lower "normal" cholesterol concentrations for five or more years? Even if there is a reduction of 30% in relative mortality what is the absolute advantage?

All the habits of this patient may reduce the risk of coronary heart disease, but when serum cholesterol is above 5.0 mmol/l or the total ratio of cholesterol to high density lipoprotein is above 5 more specific treatment should be considered. A statin will reduce the absolute risk in such patients by about 7% over a five year period[1] and more if the cholesterol concentrations are higher but less if they are lower. In other words, for 100 patients treated with a statin for a "normal" cholesterol there will be one coronary event less a year.

Michael Oliver, *emeritus professor of cardiology, London*

1 Scandinavian Simvastatin Survival Study Group. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian simvastatin survival study (4S). *Lancet* 1994;344:1383-9.