one when she can legally terminate the entire pregnancy.[1] There are, however, grounds for caution if there is a lasting effect on the surviving twin.

Through the Lone Twin Network, the Multiple Births Foundation meets a large number of adults whose twin died at or before birth and who feel the loss profoundly. How far this self selected group is representative is not known. Research on the long term psychological sequelae, including their prevalence and relative seriousness, is therefore badly needed.

ELIZABETH BRYAN
Medical director

Multiple Births Foundation,
Queen Charlotte's and Chelsea Hospital,
London W6 0XG

1 Berkowitz RL. From twin to singleton. *BMJ* 1996;313:373-4. (17 August.)

Fifth of babies born after assisted conception have retinopathy of prematurity

EDITOR,—In vitro fertilisation and similar reproductive techniques are now capable of producing dangerously large numbers of fertilised human eggs, as shown by the woman who was expecting octuplets recently.[1] But are such women, and the doctors who look after them, fully aware of the dangers should these fetuses survive?

Recently, McKibbin and Dabbs published an important paper entitled "Assisted conception and retinopathy of prematurity."[2] Their conclusion was that "of the babies born after assisted conception treatment 20% fulfil the ROP [retinopathy of prematurity] screening criteria. ROP of any stage was present in 23% and this group also accounted for a large proportion of those reaching stage 3 disease and of those requiring treatment." What this means in simple language is that there is a high probability that such babies will be blind. I would ask such potential mothers: do they really want to bring eight blind babies into the world?

D P CHOYCE
Consulting ophthalmologist

Southend Hospital,
Westcliff on Sea SS0 0RY

1 Dyer C. Selective abortions hit the headlines. *BMJ* 1996;313:380. (17 August.)
2 McKibbin M, Dabbs TR. Assisted conception and retinopathy of prematurity. *Eye* 1996;10:476-8.

Use of fertility treatments needs more careful oversight

EDITOR,—The occurrence of drug induced multiple ovulation followed by multiple pregnancy represents a failure of procedures designed to protect both individuals and society from the adverse effects of medicines.[1] The manufacture and quality control of the medicines themselves, in this case gonadotrophin products, have been statutorily regulated in the United Kingdom since the Medicines Act of 1968 and, more recently, under European Union legislation. The strength of each medicine is standardised against international standards prepared and distributed under the auspices of the World Health Organisation, which keeps up to date with changing technology.[2]

The pharmacodynamic effects of gonadotrophin treatment in individual cases need to be carefully monitored to ensure an appropriate response. One method of doing this, by use of plasma oestradiol assays, is also regulated in the United Kingdom through laboratory accreditation and the external quality assessment schemes for hormone assays.

The effects of ovarian stimulation have long been known.[3] In 1927 "superovulation" was first described (in mice), and by the 1960s the requirement for careful monitoring of human gonadotrophin treatment had been established. Has not the time come when the use of these technologies needs more careful oversight, as is already the case for the pharmaceutical manufacturers and assay laboratories on which they depend?

Fertility treatment has profound personal, ethical, and social implications, as has been recognised in the work of the Human Fertilisation and Embryology Authority in the complex area of in vitro fertilisation and related techniques. It should not be difficult to extend this philosophy, principle, and practice to the use of fertility drugs in less complex, but still hazardous, treatments.

STEPHEN JEFFCOATE
Consultant

PO Box 188,
Dorking,
Surrey RH4 2FE

1 Dyer C. Selective abortions hit the headlines. *BMJ* 1996;313:380. (17 August.)
2 Jeffcoate SL. Standardisation of human follicle-stimulating hormone for therapy and diagnosis: report of the international workshop at the National Institute for Biological Standards and Control, March 1996. *Clin Endocrinol* (in press).
3 Jeffcoate SL. From TNA to DNA: 60 years of follicle stimulating hormone in gynaecology. In : Studd JW, ed. *The yearbook of the Royal College of Obstetricians and Gynaecologists, 1994.* London: Royal College of Obstetricians and Gynaecologists Press, 1995:55-67.

Terminations of pregnancy, not unplanned deliveries, increased as result of pill scare

EDITOR,—In October 1995 the Committee on Safety of Medicines gave a warning about the risk of thromboembolism in women taking oral contraceptives containing gestodene or desogestrel.[1] James Owen Drife questions whether there was an increase in the number of terminations of pregnancy resulting from women abandoning the pill as a result of this warning.[2] He suggests, instead, that labour wards were stretched by an increased number of deliveries of subsequent unplanned pregnancies.

Last May we reported a 9.9% increase in terminations of pregnancy carried out in Oxford from November 1995 to March 1996, inclusive, compared with the same period in each of the previous two years. In response to direct questioning 8% of women requesting a termination of pregnancy reported that their failure of contraception was a consequence of their having stopped oral contraceptives containing gestodene and desogestrel.[3]

Table 1—*Number of deliveries at John Radcliffe Hospital each month from April to August, 1994-6*

	1994	1995	1996
April	571	570	502
May	558	583	542
June	581	612	559
July	536	562	576
August	576	537	505
Total	2822	2864	2684

Analysis of the activity of our delivery unit indicates that there were 5.6% fewer deliveries in Oxford between April 1996 and August 1996, inclusive, than in the same period in the previous two years (table 1). Most women who became pregnant as a result of stopping oral contraception because of the pill scare would have delivered during this period.

These figures do not support Drife's views or expectations; they suggest that the adverse effect of the pill scare was an increase in terminations of pregnancy but not in unplanned deliveries.

T J CHILD
Clinical research fellow
I Z MACKENZIE
Clinical reader
M REES
Honorary senior clinical lecturer

Nuffield Department of Obstetrics and Gynaecology,
Women's Centre,
John Radcliffe Hospital,
Oxford OX3 9DU

1 Committee on Safety of Medicines. *Combined oral contraceptives and thromboembolism.* London: CSM, 1995.
2 Drife JO. Old men and young girls. *BMJ* 1996;313:368. (10 August.)
3 Child TJ, Rees M, MacKenzie IZ. Pregnancy terminations after oral contraception scare. *Lancet* 1996;347:1260-1.

Correcting outcome data for case mix in stroke medicine

Structure and process should be audited, rather than outcomes

EDITOR,—The method used by Richard J Davenport and colleagues to correct for differences in case mix before and after a stroke unit was set up could obscure genuine differences in outcomes due to changes in medical care.[1] A logistic regression model containing 19 prognostic variables, whose coefficients are derived from the study itself, is almost certain to overfit the data, so that some differences that are due to treatment may be spuriously "explained" by adjustment for case mix. The method may therefore be unfair to the "before and after" comparisons used in clinical audit studies.

Nevertheless, the paper contains a salutary warning about the dangers of non-randomised comparisons, particularly those that might be used to generate league tables of outcomes in different hospitals or units. The corrections for case mix used by the authors apply only to cases of stroke identified prospectively, which may differ considerably from those identified by the routine hospital coding system. We compared 340 cases of acute stroke (*International Classification of Diseases*, ninth revision, codes 431, 433-4, 436) identified from the hospital information system with those included on a prospective stroke register over 15 months in a teaching hospital in Liverpool. Of the 420 confirmed cases, 278 (66%) were on the hospital information system, but many patients with minor or non-paretic strokes (often misdiagnosed as transient ischaemic attacks) and some of those with rapidly fatal strokes were missed. Sixty two patients identified from the hospital information system had a false positive diagnosis of stroke; many of these patients had been admitted for other reasons, having had strokes previously. Thus the overall case mix was quite different from that reported by the authors.

Even when cases are identified prospectively, comparisons between different units can be hazardous. Not only would few hospitals be able to collect the detailed information on case mix used in Davenport and colleagues' study, but much of it is subject to wide interobserver variability.[2,3] For instance, in multicentre comparative studies for the European stroke database project we have found large differences in the proportion of total anterior circulation strokes, which have mostly been due to differences in doctors' willingness to assess key signs (such as visual fields) in drowsy or dysphasic patients.[4]

Because of these difficulties it is unlikely that valid purchasing decisions can be based on such comparisons among units in the foreseeable future. Randomised controlled trials are needed

to establish the relation between the structure or process of care and the outcome, and thereafter it is structure and process rather than outcomes that should be audited.[5]

DAVID BARER
Professor of geriatric medicine
JOHN ELLUL
Research fellow
Institute for the Health of the Elderly,
Newcastle General Hospital,
Newcastle upon Tyne NE4 OBE
CAROLINE WATKINS
Research coordinator
Aintree Stroke Unit,
Fazakerley Hospital,
Liverpool L9 7AL

1 Davenport RJ, Dennis MS, Warlow CP. Effect of correcting outcome data for case mix: an example from stroke medicine. *BMJ* 1996;312:1503-5. (15 June.)
2 Benbow SJ, Watkins C, Sangster G, Ellul J, Barer D. The availability and reliability of information on the premorbid functional status of stroke patients in hospital. *Clin Rehabil* 1994;8:281-5.
3 Shinar D, Gross CR, Mohr JP, Caplan LR, Price TR, Wolf PA, *et al.* Inter-observer variability in the assessment of neurologic history and examination in the stroke data bank. *Arch Neurol* 1985;42:557-65.
4 Barer D, Ellul J, and ESDB Collaborators. The European stroke database project: developing a common clinical language for stroke. *Eur J Neurol* 1995;2(suppl 2):11-2.
5 Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;311:793-6.

Further studies are needed to assess current outcome indicators

EDITOR,—Richard J Davenport and colleagues question the validity of the clinical outcome indicators published by the Scottish Office.[1] In fact, the results that they present suggest that the outcome indicators published in Scotland do their intended job—raising meaningful questions, nothing more—reasonably well.

The indicators in the authors' study—that is, the study data standardised for age and sex—suggest that mortality associated with standard care is higher than that associated with the care provided in a stroke unit. In a real context, such a result might well lead to a call for further analysis or audit to establish whether this difference remains after adjustment for case mix. The authors follow this logic and conclude that, "because of the imprecision of the corrected data, the results are still consistent with a moderate but non-significant beneficial effect." Thus the indicator is shown to raise a valid issue even in artificial circumstances far removed from those for which the indicators were designed, involving a comparison over time (with an associated major shift in case mix) rather than between hospitals at one point in time, and also involving a number of cases well below the threshold for publication.

The authors' study is not an appropriate way to test the validity of the Scottish outcome indicators. Even according to the rule of thumb that the authors use to justify their statistical procedure, the analysis may well have "overcorrected" by forcing the inclusion of 19 explanatory variables. Eighty nine deaths within 30 days does not represent five outcome events for each explanatory variable.

Davenport and colleagues, in collaboration with the producers of the Scottish outcome indicators, are engaged in a major study to assess the feasibility of routinely monitoring outcomes of stroke by looking at enhanced administrative data. This study will soon make possible a linked comparison, on a case by case basis, between outcomes derived from Scottish hospital discharge data as they currently stand and outcomes derived from data enhanced to enable adjustment for case mix. Such a comparison, based on adequate numbers and across five hospitals, will help us to come to a much more precise and balanced understanding of the strengths and limitations of the current outcome indicators and how to improve them. We look forward to collaborating in this analysis.

STEVE KENDRICK
Principal statistician
MARION BAIN
Consultant in public health medicine
Information and Statistics Division,
NHS in Scotland,
Edinburgh EH5 3SQ

1 Davenport RJ, Dennis MS, Warlow CP. Effect of correcting outcome data for case mix: an example from stroke medicine. *BMJ* 1996;312:1503-5. (15 June.)

Study should have had more patients or longer time scale

EDITOR,—We welcome Richard J Davenport and colleagues' reminder of the importance of correcting outcome data for case mix.[1] Given the size of their study, however, they should not be surprised or disappointed that they were unable to show a beneficial impact of their stroke unit on mortality.

The 12 month mortality before the introduction of their stroke unit in Edinburgh was 39%.[1] A meta-analysis of the effect of stroke units reports a 23% (95% confidence interval 4% to 37%) reduction in the odds of death at final review (median follow up one year) in the intervention group.[2] Such an effect would lead to a lowering of the death rate in Edinburgh to 33%. To be able to detect this change with 80% power and at a (two sided) significance level of 5% would require a study of 2066 patients (1033 before the introduction of the stroke unit and 1033 after its introduction). Davenport and colleagues' study was of 468 patients identified over 27 months. To be confident that the introduction of their stroke unit had a smaller impact on mortality than that predicted by the meta-analysis, Davenport and colleagues' study would have needed to be much longer (119 months as opposed to 27).

The study is a clear illustration of why it can be impractical to try to compare performance among or within providers by monitoring outcomes, even when they are adequately adjusted for case mix: large numbers of patients or long time scales may be necessary to overcome the play of chance.[3] In areas of health care in which the results of randomised controlled trials are available it is possible to ensure that clinical practice reflects this evidence by the use of appropriate measures of the process of care. In this case, there is good evidence that introducing a stroke unit saves lives. Purchasers would thus do better to ensure that their providers offer "well organised services for acute stroke patients which provide comprehensive care centred on an integrated multi-disciplinary team who have a specialist interest in stroke rehabilitation"[2] rather than seek to interpret their case fatality rates.

JONATHAN MANT
Clinical lecturer in public health medicine
Department of Public Health and Primary Care,
University of Oxford,
Radcliffe Infirmary,
Oxford OX2 6HE
NICHOLAS R HICKS
Consultant public health physician
Oxfordshire Health Authority,
Oxford OX3 9DZ
JOHN FLETCHER
Honorary research fellow
Unit of Health Care Epidemiology,
Oxford Institute of Health Sciences,
Oxford OX2 6HE

1 Davenport RJ, Dennis MS, Warlow CP. Effect of correcting outcome data for case mix: an example from stroke medicine. *BMJ* 1996;312:1503-5. (15 June.)
2 Stroke Unit Trialists' Collaboration. A systematic review of specialist multidisciplinary team (stroke unit) care for stroke inpatients. In: Warlow C, Van Gijn J, Sandercock P, eds. *Stroke module of the Cochrane database of systematic reviews.* Oxford: Cochrane Centre, 1996 [updated 23 February 1996].
3 Mant J, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995;311:793-6.

Authors' reply

EDITOR,—All of the authors of the letters about our paper seem to agree with its central tenet. David Barer and colleagues and Steve Kendrick and Marion Bain suggest that our regression models may have overcorrected the data, and we discussed this possibility in our paper. While multiple logistic regression modelling is far from perfect, we thought that attempting to correct for case mix in this way was a more sensible approach than simply ignoring the problem. As discussed, we were reassured that we obtained almost identical results when we applied other prognostic models, which had been derived from and validated on different stroke datasets.[1] We had developed these considerably simpler models (which rely on five or six variables, such as age and systolic blood pressure on admission) because we realised that collecting the amount of data included in our study is beyond the resources of most hospitals, at least in routine clinical practice. We are currently testing the applicability of these models in several hospitals (as referred to by Kendrick and Bain). We agree with the suggestion that the structure and process of care may be more appropriate measures of quality of care, although there are still considerable problems associated with this approach.[2,3]

We have conducted a similar study to that of Barer and colleagues, comparing the accuracy of routine hospital coding statistics with our stroke register.[4] Of 566 patients registered as having a stroke, 84 (15%) were not given a code (*International Classification of Diseases*, ninth revision) for stroke; although better than Barer and colleagues' experience, this is far from perfect.

Finally, while we agree with Kendrick and Bain that the purpose of the Scottish outcome indicators is to raise meaningful questions, we are aware that they have been interpreted by many people as comparative league tables of hospitals' performance, and the purpose of our paper was to show the dangers of using the data as a direct measure of quality of care. As an example of this problem, the *BMJ* recently published a news item which suggested that the clinical outcomes initiative could identify hospitals which "seemed to perform badly."[5] We were interested to note that this report also raised doubts about whether the outcomes initiative would continue.

RICHARD J DAVENPORT
Senior registrar
MARTIN S DENNIS
Senior lecturer
CHARLES P WARLOW
Professor of neurology
University of Edinburgh,
Department of Clinical Neurosciences,
Western General Hospital,
Edinburgh EH4 2XU

1 Counsell C, McDowall M, Slattery J, Dennis M. Prediction of functional outcome following stroke using a validated model [abstract]. *Cerebrovasc Dis* 1996;6(suppl 2):50.
2 Davenport RJ, Dennis MS, Warlow CP. Can non-randomised studies reliably measure the quality of care provided for stroke patients? [Abstract.] *Cerebrovasc Dis* 1996;6(suppl 2):50.
3 Davenport RJ, Dennis MS. Assessing the quality of care. *BMJ* 1996;312:185. (20 January.)
4 Davenport RJ, Dennis MS, Warlow CP. The accuracy of Scottish morbidity record (SMR1) data for identifying hospitalised stroke patients. *Health Bull* 1996;54:402-5.
5 Christie B. Scotland has wide health care variations. *BMJ* 1996;313:578. (7 September.)

Measuring health outcomes

Negative values should be allowed in ratings of quality of life

EDITOR,—Ability to assess treatment and the outcomes of diseases continues to improve, and the schedule for evaluating individual quality of