A MULTIFACTORIAL APPROACH TO ANTENATAL FETAL MONITORING

G.V.P. CHAMBERLAIN (*),
G.B. NEWMAN (**), A. GILLESPIE (***),
I. SOMMERVILLE (*), S. BISWAS (*),
S.C. CAMPBELL (****),
E. YOUSSEFNEJADIAN (*), J. DEWHURST (*)

(*) Institute of Obstetrics Queen Charlotte's Hospital for Women, London W6
(**) Department of Mathematics, the Middlesex Hospital, London W1
(***) Now at University of Adelaide, South Australia
(****) Now at King's College Hospital, London SE5

SUMMARY

The prediction of the fetus at higher risk is commonly attempted using laboratory tests. It is difficult to measure the precision of these tests and so a prospective survey has been performed comparing six such tests and the clinical assessment of fetal progress in the last weeks of pregnancy. The ultrasound assessment of the fetal biparietal diameter, urinary oestrogens, plasma oestriol, human placental lactogen, the activities of heat stable alkaline phosphatase and cystine amniopeptidase were investigated in 189 patients. No single test was significantly effective in the prediction of perinatal death, low Apgar score or weight retardation, even the most precise misclassifing approximately 25 % of the cases. Combining the tests improved prediction significantly and weighting factors were constructed to predict the small-for-dates babies and those with low Apgar scores. It was concluded that, while these indices were too cumbersome for regular use, such a grouping of tests may indicate what the clinician does when he weighs up many factors before making a decision about taking any action about induction of labour and further combinations should be sought.

INTRODUCTION

One of the aims of antenatal care is to predict the high risk fetus so that the best time for delivery can be determined and hypoxia in labour anticipated. There are now numerous antenatal laboratory tests which try to help the clinician. Some are hallowed by years of use, others are recent introductions backed only by the enthusiasm of those who pioneered them. It was decided to examine the predictive tests used in our clinic to assess their usefulness, comparing them with the clinical assessment of fetal progress. The first step was to see if any single test would help to predict fetal outcome better than any other. If no significant differences were found, it was decided to try grouping the tests to see if this and the addition of weighting factors helped to improve prediction rates.

METHODS

One hundred and eighty-nine women were assessed from the 30th to the 36th week of pregnancy. After full discussion the women agreed to allow urine, blood and ultrasound tests to be performed every two weeks for the rest of the pregnancy. Thus the group included a higher proportion of the better motivated patients and those who had problems in previous pregnancies.

Fetal testing was related to some neonatal outcome. Many significant factors do not make themselves apparent until later in neonatal or infant life, but it was decided in the short term that the outcome could best be measured by:

1) the perinatal mortality rate;
2) the Apgar score at birth;
3) the birthweight in proportion to the gestational age.

The first includes the stillbirths and the early neonatal deaths; this is an absolute measure, but there were few deaths in this series. The Apgar score as divided by results of 0-4 and 5-10, for a good prognosis cannot be given for babies with scores below 5. The last factor is concerned with the removal of growth support in the latter

weeks of pregnancy for poor placental exchange may be related to retarded fetal growth.

Standard charts relating birth weight to gestation were used ([6]) those beneath the fifth percentile of weight for gestational age were considered small-for-dates. Thus, the criteria for chronic placental insufficiency were lightness for dates and for acute placental insufficiency, an Apgar score below 5.

The population was divided according to the fetal outcome:

1) there were 5 deaths (4 stillbirths and 1 neonatal death) and 184 women whose liveborn babies were still alive at the end of a month;

2) there were 16 women who produced babies with Apgar scores of 4 or less at one minute;

3) there were 15 women who produced live babies who were small-for-dates (S.F.D.) and 168 whose babies were not.

This group of 189 infants had a perinatal mortality of 26 per 1000 total births, higher than the Queen Charlotte's Hospital background mortality at that time (21 per thousand); 8.7 % of the live had an Apgar score of below 5 whilst about 6 % would have been expected from the Hospital population. Similarly, 8.2 % of women had live babies with a birth weight below the fifth centile, whilst from the total hospital population 5.5 % would be expected. Thus, the patients in the survey produced babies who were a little worse off than the total population, but since it is the tests of fetal outcome themselves that are being assessed, it was felt that this did not invalidate the method of investigation.

The following seven tests were performed on patients at alternate weeks at antenatal visits from 30 weeks onwards.

1) Clinical assessment of fetal size.

2) Ultrasound assessment of the biparietal diameter of the fetal head (BPD).

3) 24 hour excretion of urinary total oestrogens.

4) Plasma unconjugated oestriol levels.

5) Plasma human placental lactogen (HPL) levels.

6) Plasma heat stable alkaline phosphatase (HSAP) activity.

7) Plasma cystineaminopeptidase (CAP) activity.

The clinical assessment was performed by palpatation. The biparietal diameters of the head were measured by an ultrasonic technique using the Diasonograph 4101 (Nuclear Enterprises Ltd.) as described by Campbell (1968) ([4]) Urinary total oestrogens were estimated by the method described by Brown et al (1968) ([3]) and the plasma unconjugated oestriol by radio-immunological methods described by Youssefnejadian et al (1973) ([7]); human placental lactogen levels and the activities of heat stable alkaline phosphatase and cystineaminopeptidase were determined by previously described methods (Biswas et al, 1972) ([2]).

STATISTICAL ANALYSIS

A multifactorial analysis of all seven factors collected over ten weeks on 189 patients would have presented problems, so the data was consolidated into groups so that like could be compared with like. Since many of the observations were widely spaced and showed no clear evidence of a linear trend, we used a method of interpolating the data reported by Campbell and Newman (1971) ([5]). An attempt was made to mimic the intervals of routine antenatal care by using readings taken during each of the four week periods 30-33 week and 34-37 week. Although there were readings in the 37-41 week period, analysis was confined to the weeks before this, for it is from 30 to 37 weeks that the clinician needs the most help from fetal monitoring tests.

Each mother provided at least one reading in each of the four week periods and, since these were usually multiple measurements, a single measurement had to be chosen. Target dates were set at 32 and 36 weeks; if the mother had a measurement taken during the target date week, that reading was taken. If the mother had a test on each week on either side of the target week, the average was taken. Otherwise, the test within the stated four week period, which was closest in time to the target date, was taken.

Despite the number of pregnancies studied, there were only 15 small-for-dates babies, 16 with Apgar scores of 4 or less and 5 perinatal deaths. Within the seven testing methods observed, it was hoped that some measure could be found which distinguished the high risk groups.

RESULTS

1. *Single tests to predict outcome.*

The number of patients, the mean and standard deviation for each measure are shown in tables 1 and 2. Table 1 gives the results of mothers with normal births, while table 2 shows the corresponding figures for those with the poorer outcomes. Several measures appeared to give statistically significant differences from the normal babies by a « t » or « F » test, and these have been marked with an asterisk in table 2. These statistical tests must be treated with some caution, as we did not have sufficient information to judge the symmetry of the distribution of results reported.

Consideration was given to a special cut-off value for each measure, trying to determine a value beyond which a mother can be described as being a high risk for having a baby affected by the selected criterion. Cut-off values were chosen for the ability of a test to demonstrate a statistically significant difference between the normal babies and those who were small-for-dates or had low Apgar scores. The choise was empirical and influenced by a sharp change in the risk, or a significant number of patients at, apparently, very high or very low risk, e.g. no patient with an HPL at 32 weeks of 5 μg or more had a small-for-dates baby.

This discreet analysis presented our missing variables with a distinct cut-off effect which would be concealed by any large spread of the normal values; to some extent, it also mimicked the clinical process of taking action at certain levels of the predictor variables (tables 3 and 4). An attempt was made to summarize the effectiveness of each test from the number of patients who would be misclassified by a dicision taken at approximately halfway between the normal and abnormal. This showed that even the most effective test applied singly would misclassify approximately 25 % of cases.

Table 1. *Mean, standard deviations and numbers of readings for normal babies.*

Measure	Target Dates (week)	Normal Babies Mean	Normal Babies Number of Observations	Normal Babies SD
Clinical estimation of weeks of gestation	32	32	(145)	1.5
	36	36	(149)	1.2
Biparietal Diameter (cm)	32	8.50	(90)	.38
	36	9.24	(113)	.36
Urinary oestrogens (μg/24)	32	12,536	(119)	6092
	36	18,076	(138)	7330
Plasma unconjugated oestriol (ng/ml)	32	11.3	(53)	10.1
	36	17.1	(60)	12.9
HPL (μg/ml)	32	5.15	(123)	2.01
	36	5.49	(127)	2.13
HSAP activity (units/ml)	32	.69	(124)	.39
	36	1.14	(124)	.63
CAP activity (units/ml)	32	11.87	(123)	5.33
	36	16.17	(128)	6.54

Table 2. *Mean, standard deviation and numbers of readings for abnormal babies (Units as in Table 1).*

Measure		SFD less than 5 %ile) Mean	Number	SD	Less than 5 Apgar Mean	Number	SD	Perinatal Deaths Mean	Number	SD
Clinical	32	31 *	(12)	2 *	31	(10)	2	32	(4)	3
estimation	36	34 *	(10)	2 *	35	(14)	2	35	(3)	1
Biparietal	32	8.1 *	(11)	.6	8.5	(9)	5	8.4	(4)	1.3
diameter	36	8.9 *	(13)	.4	9.2	(12)	5	8.5 *	(2)	.5
Urinary	32	9,581	(11)	2333 *	9,390	(11)	3,016 *	14,100	(4)	8531
oestrogens	36	14,627	(13)	6508	16,353	(15)	6,514	17,530	(3)	13077
Plasma	32	8.3	(7)	5.5	14.9	(2)	13.9	20.5	(1)	—
oestriol	36	26.1	(6)	3.54 *	11.4	(10)	6.3	16.0	(2)	3.8
HPL	32	3.1 *	(11)	1.2 *	5.4	(11)	2.3	4.2	(3)	2.7
	36	3.5 *	(12)	2.1	5.2	(16)	2.1	4.1	(4)	2.5
HSAP	32	.71	(11)	.50	.47 *	(11)	.18 *	.60	(4)	.31
activity	36	1.04	(12)	.54	.84 *	(16)	.37 *	.88	(3)	.20
CAP	32	8.9 *	(11)	4.0	9.0 *	(11)	3.2 *	10.0	(4)	5.0
activity	36	13.9	(12)	7.0	12.6 *	(16)	4.6	16.7	(3)	3.3

* Indicates different from normal at 5 % significance level.

Table 3. *Proportion of patients producing a baby who was small for dates (less than 5th percentile) using a stated cut off value for each measure.*

		Cut off Value	Number of Patients	Proportion SFD (< 5 %ile)
HPL	(32 w)	< 5.0	71	15 % (9-25 %)
HPL	(36 w)	< 3.7	37	19 % (9-25 %)
CAP activity	(36 w)	< 8.6	15	27 % (10-51 %)
Biparietal diameter . . .	(32 w)	< 7.8	5	60 % (20-91 %)
Biparietal diameter . . .	(36 w)	< 8.8	12	33 % (13-60 %)
Clinical estimation . . .	(32 w)	≤ 30 weeks	31	19 % (9-35 %)
Clinical estimation . . .	(36 w)	> 35 weeks	118	3 % (1- 7 %)
Plasma oestriol	(32 w)	> 5	74	0 % (0- 5 %)

The 90 % confidence interval of these proportions given in brackets. (Units as in Table 1).

Table 4. *Proportion of patients producing a baby with an apgar score of less than 5 using a stated cut off value for each measure.*

		Cut off Value	Number of Patients	Proportion with Apgar less than 5
HSAP activity	(36 w)	< 1.29	109	14 % (9-21 %)
CAP activity	(36 w)	< 8.6	15	27 % (10-51 %)
Plasma oestriol	(36 w)	< 19.3	54	19 % (11-30 %)

The 90 % confidence interval of these proportions given in brackets. (Units as in Table 1).

Table 5. *Index for predicting small for dates babies with the 90 % confidence interval in brackets*

	Number of Patients	Proportion Small for Dates (less than 5 %ile)
Index less than 94	23	30 % (16-50 %)
Index greater than 94	68	1.5 % (0-6 %)

Index = (10 × BPD at 32W) + (3 × HPL at 36W).

Table 6. *Index for predicting low apgar score with the 90 % confidence interval in brackets.*

	Number of Patients	Proportion with Apgar less than 5
Index less than 37	39	18 % (9-31 %)
Index greater than 37	24	0 % (0-13 %)

Index = (Cap at 38W) = $\frac{\text{(U. Oestrogen at 32W)}}{1000}$ + (4 × Plasma Oestriol at 36W).

2. *Grouping and weighting of tests to predict outcome.*

The examination of the effect of more than one factor to predict low birth weight or low Apgar score was done using a stepwise linear regression technique (Armitage, 1971). The weights found by the linear regression analysis were rounded to provide any easily computed index of risk and tables 5 and 6 show in summary those indices found to best predict small-for-dates babies and those with a low Apgar score. Again, too few deaths prevented the inclusion of perinatal deaths in any significant analysis.

(a) *Predicting small-for-dates babies.*

The biparietal diameter at 32 weeks was the best predictor of small-for-dates babies, though this was only marginally better than its values at 36 weeks and both the HPL measurements. Any of these could be used with equal efficiency, but it would require a much larger series to establish any difference between them. Sixty per cent of babies with a BPD (32w) of less than 7.8 cm were small-for-dates, a risk that is unlikely to be less than 20 % even allowing for possible sampling variation. The BPD at 36 weeks, the clinical size and HPL levels at 32 and 36 weeks were statistically associated with low birth weight ($p < 0.05$, tables 2 and 3). CAP activity at 36 weeks was associated with low birth weight in the discreet analysis (table 3), but only HPL contributed additional information over that already contained in the BPD at 32 weeks. This improvement in prediction was statistically significant ($p < 0.05$). Clinical size at 36 weeks was the next best variable to contribute further information, but this might have been a change effect. The index chosen, therefore, was:

(10 × BPD at 32w) × (3 × HPL at 36w)

The index ranges from 79.8 to 121 for the 91 cases in which it could be computed with risks of small-for-dates as given in table 5.

(b) *Predicting an Apgar score of less than five.*

In attempting to find the best index, the CAP activity at 36 weeks was the best single variable and the predictive value of the other variables could be explained in terms of their relationship with CAP activity at 36 weeks. The best additional contributions were from urinary oestrogens at 32 weeks and plasma oestriol at 36 weeks and, although these were not statistically significant, they were included in the index. The index chosen, therefore, was:

CAP activity (36w) + Urinary oestrogen (32w) + plasma oestriol (36w) : 1000.

This index could be simplified for, if the urinary oestrogen levels were expressed in mg/24 hr., the denominator of a thousand could be removed. The index ranged from 19 to 69 in the cases in which it could be computed. The results are shown in table 6.

CONCLUSIONS

Data such as we have analysed have shown certain trends, but even the best combined indices might not be confirmed in a larger series of cases or in a different population. The patients studied in this survey had a slightly poorer reproductive record than our whole hospital population but, as might be anticipated in a prospective study, the indices had a striking ability to predict healthy babies rather than to pick out unhealthy ones.

Recently there has been much speculation about the use of routine biochemical screening of total antenatal populations. Our results would indicate that any of the tests assessed, if used singly, would not be very precise but, in selected high risk groups of pregnant women, the same tests run in combination might help to determine those fetuses at even more risk. Here might lie the most economic and effective use of fetal monitoring in pregnancy.

In our series, if the definition of risk was moved halfway between normal and abnormal, even the most effective test misclassified a quarter of the patients. Important actions like the induction of labour might have been taken on the results of such individual tests and this analysis emphasizes how few of the tests we examined could be the sole deciding factor in decision making.

Because physical signs alone were not predictive enough in the field of fetal medicine, obstetricians turned to placental and fetal function tests in the first place, but our results show that, although certain tests are fractionally better indicators than clinical estimates, they do not improve prediction greatly alone. They will be added to clinical jujdgement, for this includes the weighing up of many features to make a final decision, and, if a combination of tests is used, a significantly more precise measure of risk can be obtained.

BIBLIOGRAPHY

1) Armitage P.: *Statistical Methods in Medical Research,* Blackwell, Oxford, 1971.
2) Biswas S., Hindocha P. & Dewhurst C. J.: J. Endocr. *54,* 251, 1972.
3) Brown J. B., MacLeod S. C., MacNaughton C., Smith M. A. & Smyth B.: J. Endocr., *42,* 5, 1968.
4) Campbelll S. C.: *J. Obst. Gynaec. Brit. Cwlth., 75,* 568, 1968.
5) Campbell S. C. & Newman G. B.: *J. Obst. Gynaec. Brit. Cwlth., 78,* 513, 1971.
6) Thompson A. M., Billewicz W. Z. & Hytten F. E.: *J. Obst. Gynaec. Brit. Cwlth., 75,* 903, 1968.
7) Youssefriejadian E. & Somerville I. F.: *Steroid Bioch., 4,* 659, 1973.