FORMULATION OF A SCORING SYSTEM BASED ON DISCRIMINANT ANALYSIS OF RISK FACTORS IN DEVELOPMENT OF CERVICAL CANCER—A COMPARATIVE STUDY OF CASES OF CANCER, DYSPLASIA AND CONTROL GROUP*

by

USHA SARAIYA,** M.D., D.G.O., F.I.A.C., F.I.C.S.
P. C. GUPTA,*** M.Sc., D.Sc.
MAYA LULLA,**** M.D., F.C.P.S.

and

MOHINI GARUD,** M.D., D.G.O.. M.I.A.C., F.R.C.O.G.

In this study from Cytology Clinic (AMWI) of Cama and Albless Hospitals Bombay known epidemiological factors responsible for the development of Cancer Cervix have been analysed and compared in three groups of women, those with cancer, dysplasia and a control group.

In our 12 year experience of managing cases of dysplasia, it was evident that close clinical and cytological supervision of such a large number of cases was proving difficult. It was necessary to classify those who were at a higher risk for developing cancer.

The Socio-economic profile of these patients was studied and a simple scoring system was evolved in 1976. By this we were able to select the high-risk patients and concentrate on their follow-up.

Subsequently, a comparision of scores in three groups of women those with cancer, dysplasia and a control group was undertaken which validated the earlier findings.

However, to give quantitative significance to each risk factor, computer analysis was undertaken and the original scoring system was then modified.

Data from 290 cases for whom complete and reliable information was available was computerised. These cases were as follows: 76 cancers, 112 dysplasias and 102 normal controls. On the basis of univariate analysis, several variables were excluded and only the most important independent variables were retained for multivariable computor analysis.

These variables were age at marriage, years of married life, gravidity, presence of infection and trichomonal infection. Age of patient was excluded as it is not independent of age at marriage and years of married life.

The technique of discriminant analysis is utilised to statistically distinguish be-

* *Paper from Cytology Clinic (AMWI), Cama & Albless Hospital supported partly by a Research grant from I.C.M.R. (New Delhi), Read out 26th Conference of Federation of Obstetrics & Gynaecology Societies of India.*

** *Hon. Cytologist, Cytology Clinic (AMWI) Consultant Obstetrician Gynaecologist, Cama & Albless Hospital & Associate Prof. of Obst. & Gynaecology Grant Medical College, Bombay-400 008.*

*** *Statistician, Tata Institute of Fundamental Research.*

**** *Research Officer, Cytology Clinic (AMWI) Cama & Albless Hospital.*

Accepted for publication on 16-2-83.

tween two or more groups of cases which are defined by the particular research situation. In this case, that being the risk of developing cancer cervix.

To distinguish between the groups a number of discriminating variables are selected that measure the characteristics on which the two groups are expected to differ.

One or more linear combinations of these variables are then formed, the analysis of which enables one to discriminate between the two groups, in other words to tell them apart.

Once the discriminant functions have been derived, it is possible to pursue the two research objectives of analysis and classification.

For classification, this technique is used after computation. Once a set of variables is found which provides satisfactory discrimination for cases with known group memberships, a set of classification functions can be derived which will permit the classification of new cases with unknown memberships. Then, the combination of characteristics that classify a patient in dysplasia or cancer group can be used to prognosticate on women who are currently normal.

As a check of the adequacy of classification functions and the scores derived from it, we can classify the original set of cases, to see how many are correctly classified by the scores and the variables being used.

The technique of discriminant analysis described by Klecka (1975) has been successfully utilised by Bibbo *et al* (1976). Bartels (1980) and others to study the difference in the chromatin appearance in intermediate cells from patients with uterine cancer and those with normal cytology.

Table I gives the scoring for each factor. This scoring was arrived at by the computer analysis which gave the precise quantitative significance of each factor taking into account the effect of all other risk factors at the same time. To the total score thus obtained a constant of—17 was added which then subdivided the scores into positive and negative. Negative score indicated a normal individual and a positive score indicated a possible disease status either dysplasia or cancer.

TABLE I

Scores for Each Group for Classifying Control vs. Dysplasia or Cancer

Variables	Classes	Scores
1. Age at Marriage	Less than 13	1
	14-16	3
	17 and above	0
2. Years of married life (Age of respondant-Age at marriage)	Up to 19	0
	20-29	4
	30-39	7
	40 and above	17
3. Gravity	Upto 4	0
	5 or more	4
4. Infection	Absent	0
	Present	25
5. Trichomonus	Absent	0
		5
Constant	—	—17

Table II gives the analysis in 2 groups. Of those with the disease category of cancer or dysplasia 69.15+ showed a positive score and 30.85% a low score. Whereas in the control group the figures were reversed 18.63% had high score and 81.37% had low score.

Table III gives the scoring for classifying cancer and dysplasia cases. To the total a constant of—7 had to be added to arrive at either a positive or negative score. Positive score indicated cancer, whereas a negative score indicated dysplasia.

Table IV gives the scores in the 2 groups, dysplasia and cancer. Here again the findings are clear. Out of 76 cases of cancer, 69.7% had a positive score and 30.3% had a negative score. Whereas in

TABLE II
Classifications on the Scores

Groups	No. of Cases	Positive Scores	Negative Scores
Cancer-Dysplasia	188	130 (69.15%)	58 (30.85%)
Control	102	19 (18.63%)	83 (81.37%)

TABLE III
Scores of Each Group for Classifying Cancer vs. Dysplasia

Variables	Classes	Scores
(1) Age at Marriage	Less than 13	—3
	14-16	10
	17 and above	0
(2) Years of married Life (Age of respondent Minus age at marriage)	Up to 19	0
	20-29	8
	30-39	10
	40 & above	6
(3) Gravidity	Up to 4	0
	5 or more	7
(4) Infection	Absent	0
	Present	—5
(5) Trichomonus	Absent	0
	Present	—11
Constant	—	—7

TABLE IV
Classifications of Dysplasia vs. Cancer

Groups	No. of Cases	Positive Scores	Negative Scores
Cancer	76	53 (69.7%)	23 (30.3%)
Dysplasia	112	34 (30.4%)	78 (69.6%)

dysplasia 30.4% had positive score and 69.6% had negative score.

From the above it appears that about 70% cases are correctly identified in each group and that in about 30% in each group the identification fails.

The cancer and dysplasia cases, which were correctly identified in the primary classification have been considered separately for the second classification, in Table V. In this group once again it was found that 70% of cancers and 20% dysplasia had a positive score and that 30% cancers and 78.31% dysplasias had a low score.

TABLE V

Classification for Selected Cancer and Dysplasia Cases Identified Correctly in Primary Classification

Group	Positive Scores	Negotive Score in each group
Cancer	47	15
62 Cases	(68.09%)	(31.91%)
Dysplasia	83	65
128 cases	(21.69%)	(78.31%)

Therefore, the authors feel that the modified scoring system is more precise and is of greater significance in analysing the risk factors for cancer cervix.

When resources and facilities are limited, this simple scoring system can be used to advantage in selecting "the High-Risk" group for close clinical and cytological monitoring.

References

1. Bartels, P. H.: Analyt Quart Cytol 2: 19, 1980.
2. Bibbo, M., Bartels, P. H., Chen, M., Harris, M. J., Fruttmann, B. and Weid, G. L.: Acta Cytol. 20: 249, 1976.
3. Klecka, W. R.: "Discriminant Analysis in Statistical Package for Social Sciences" Nie NH, Hull CH, Jenkins JG, Steinbrenn I, Bent DH, Seconded New York Mcsraw Hill, 1975.
4. Lulla, M., Khan, S., Garud, M. and Saraiya, U.: J. Obstet. Gynec. India. 30: 359, 1980.
5. Saraiya, U., Lulla, M. and Gupta, P. C.: Accepted for publication by J. Obstet. Gynec. India.