Cesar: cervical smear analyser and reader

A new approach to evaluating cells in cytological preparations

I. M. P. DAWSON, C. P. HEANLEY, A. C. HEBER-PERCY,
AND J. K. TYLKO

*From Tetronics Research and Development Corporation Ltd., Faringdon, Berks., and the
Department of Clinical Histochemistry and Cytology, Westminster Hospital and Medical School, London*

SYNOPSIS Cesar presents an automated multiparameter approach to pattern recognition in the rapid detection of malignant cells from cervical and other cancers. Cells are cybernetically classified. They are collected into a container into which a nuclear stain may be incorporated and laid as a track on 35 mm. film using a disposable mapping pen device. The film track (Cytotrack) may be scanned using conventional optics and provision may be made for electronic sensing devices measuring a number of parameters, some of which are not normally used in conventional microscopy.

The object of this communication is to give an outline of a new approach to the evaluation of cells in a variety of cytological preparations, with particular reference to smears from the female genital tract. Here we are concerned chiefly with the main features of this method and the reasoning which led us to adopt it.

There is no doubt that exfoliative cytology has come into its own as a diagnostic tool in the detection of cancer and precancerous conditions. With trained observers, adequate facilities, and above all sufficient time, a high degree of accuracy in diagnosis can be obtained. Its disadvantages, particularly for mass screening, are obvious: it is time-consuming and expensive in terms of actual cost per positive case diagnosed and in skilled laboratory personnel who might be more profitably engaged. A less obvious, but perhaps more grave disadvantage is that the diagnosis of a cell as malignant or non-malignant depends more on a subjective impression in the mind of an individual, linked to the memory of the appearance of previous similar cell images, than to any measurement capable of quantitative qualification. Thus the diagnosis, as generally practised, is an art rather than a science. This makes communication between cytologists somewhat limited since impressions must inevitably contain the subjective element. Let us consider more closely what is meant by the 'general impression' of a vaginal or cervical smear. This is conveniently obtained by scanning at low magnifications ($\times 80$ to $\times 200$). At this level two types of information are

derived, one relating to the quality of the smear itself, *e.g.*, uniformity of spreading and staining, presence of artefacts, etc., and one relating to the cell populations, *viz.*, the ratios between the numbers of the different type of cells, their state of preservation and folding, tissue fragmentation and clumping, presence of leucocytes and erythrocytes, bacterial infection, etc. No actual measurements or counts are normally taken at this stage in routine examinations. Even the gradation of hormonal effects (Schmitt, 1953) does not necessitate actual counting. It is only in the more detailed study of a cell or cell groups carried out at $\times 300$ to $\times 400$ magnifications that certain 'pseudo-quantitative' measurements are occasionally introduced. The most common are the cytoplasm/nucleus ratio; the karyopyknotic index and the eosinophilic index. The first may be of great help in histology where the orientation of the cells observed is known, but becomes of little value for separate cells whose orientation is not certain. The second, in spite of many attempts at strict defining as simple superficial cell index, is still in practice defined in at least three different ways, the differences being reflected in the estimation of oestrogenic activities. The third introduces an original element of doubt since the most commonly used stain, that of Papanicolaou, often gives variable staining, making the distinction between the eosinophilic and cyanophilic cells difficult or impossible (Smolka and Soost, 1965). The use of all these 'pseudoquantitative' methods necessitates lengthy counting and/or measuring procedures which are clearly not practicable in routine analysis.

In attempts to overcome these difficulties many workers have turned to automatic or semiautomatic ways of evaluating exfoliated material. Basically, two types of instrument have been developed.

The larger group of analysers comprises the cyto-analysers proper. These vary from the early type of scanning devices based on the use of the Nipkov disc through the object plane scanners (Tolles and Bostrom, 1956) to a sophisticated method of chromosome analysis (Ledley and Ruddle, 1966), relying on contour resolution (Mellors and Silver, 1951; Mellors, Glassman, and Papanicolaou, 1952; Sawyer and Bostrom, 1958; Bostrom, Sawyer, and Tolles, 1959; Tolles and Bostrom, 1956). These methods tend to prefer fluorimetric stains and all use not more than three parameters, more often one or two. The main attempts have been directed to the reduction of various sources of noise and increase of the rate of information throughput. These methods follow closely television developments but appear to pay relatively little attention to the pattern recognition itself. The other type of instruments dispense with the preparation of the orthodox smear but instead use a suitably dispersed suspension of the exfoliated material. For instance, the Coulter method, in which the relatively non-conducting cells are suspended in an electrolyte solution, attempts to measure the impedance to the electric current as a function of cell volume.

The above instruments share one common feature: they attempt to measure not more than three parameters. This is, in our opinion, a basic misunderstanding. Instead of aiming at recognizing the 'general pattern' of the smear by considering a wide range of parameters, as the basic requirement of diagnosis, a number of optoelectronic or purely electronic devices have been developed, all of which are capable of measuring a very limited number of parameters—most frequently just one. These parameters, *e.g.*, the average diameter of a nucleus, or the approximate amount of a certain component of a nucleus, although in themselves very useful adjuncts in the diagnosis of malignancy, cannot be by themselves relied upon (Iversen, Gordon, Cowell, and Watson, 1966).

Employing a few parameters only, is like trying to describe an individual merely in terms of his height and weight or sex; it may enable one to tell a fat man from a small thin woman, or a frankly malignant cell from a benign one, but it is of less help in distinguishing two individuals or cells which closely resemble one another.

THEORY

A fundamentally new approach in this field is needed. The situation may be summarized as follows: the processes responsible for the changes occurring in the cells to be exfoliated or already exfoliated cells are

B	A	C
Cell differentiator	Stratum	State differentiator
Operator →	Operand	← Operator

FIG. 1. The cybernetic principle of the three letter code adopted in the C.F.C.

cybernetic by nature, that is, they cannot be explained by simple cause-and-effect relations, but involve complex feedback processes. Hence attempts at studying such systems by measuring one or two parameters must inevitably give at best a very distorted and frequently quite misleading picture.

The immediate answer would be a truly multiparameter approach. This is in fact the method used by the eye-brain system of a pathologist. To date, truly multiparameter mechanized systems have not been seriously contemplated, not only because of their alleged complexity, cost of the apparatus and limitations of existing computers, but chiefly because it was thought that extracting a large number of parameters would slow down the throughput of the machine thus invalidating one of its main objects, speedy analysis.

In Cesar project we have purposefully rejected the temptation of a single parameter approach and chosen the quantitative multiparameter one. The main difficulties mentioned above are being solved in five ways:—

(1) by presenting the exfoliated material to the machine in the form of a very narrow continuous track, containing dispersed cells, whose total width can be observed under a low power objective;

(2) by arranging a large number of sensing stations along the track;

(3) by providing as far as possible all the U.V. and visible light at various required frequencies from one large polychromatic plasma source;

(4) by introducing in addition to the existing parameters a new range of cell-defining parameters, each contributing its own fraction of information;

(5) by adopting an unambiguous cell classification capable of very sensitive quantitative morphological or chemical differentiation.

The last mentioned cell classification is used by us in the Cesar fundamental classification (C.F.C.) which is based on a cybernetical concept of operand-operator relationship. Each form is expressed as a three-letter code of which the central letter acts as an operand, while the first and last act as operators (Fig. 1). The central letter represents a stratum which, for all the cytoforms, is based on histo-embryological considerations.

We have defined five strata: endoderm (N); mesoderm (M); ectoderm (C); extraneous (T); and artefacts (A). The first letter in the C.F.C. is a cell differentiator, which may be defined as a sum total of operations (the operations themselves need not be known) which when acting upon the stratum will produce a particularly differentiated cell. Further subdifferentiation at all levels, *e.g.*, tissue, cytoplasm, or nucleus, is also possible by the use of brackets and suffix numbers. It is immediately obvious that only certain combinations of strata with cell differentiators are possible or 'permitted' while others are 'forbidden'. The third and last letter of the C.F.C. is also an operator which acts as a state differentiator. Typical state differentiators include, amongst others, phenomena like pregnancy pattern, karyomegaly, malignancy, and alkylating agent patterns.

The use of C.F.C. permits further application of operators (subdifferentiation) so that ultimately an extremely sensitive description is possible. Classification of this type is known as 'open-ended', *i.e.*, it is constructed in such a way that no cross ramifications will occur when new operators are used, permitting the system to be expanded without any loss of argument or introduction of ambiguity. The chief advantage of using this type of classification lies in the ability to describe unambiguously any form at any given state.

Each form, be it a single cell, or an artefact, can be described by a large number of parameters. Each parameter contributes a certain limited amount of information only. The percentage of total information available from a single parameter will take an asymptotic form, the information content never exceeding a certain limit (Fig. 2). Our aim in cell classification is to use a large number of parameters at reasonably high information contents: that is, we need a large number of graphs like Figure 2.

Two other factors in which our measurements differ from the conventional are also important. When we measure a particular parameter, instead of writing an average for a large number of measurements we write the upper and lower values of a bracket within

FIG. 2. Asymptotic character of one parameter observation related to total information content.

FIG. 3. Venn diagrams applied to differentiation between the values of parameters: *a* = values quite different (dis-joint); *b* = certain values common to both codings (con-joint); *c* = one set of values wholly contained within the other—differentiation not possible.

which the set of measurements has fallen. When the same parameter is measured for two different cells three situations may occur (Fig. 3). In normal pattern recognition one looks for similarities, *e.g.*, Fig. 3 C; in Cesar we are looking for differences, *e.g.*, we seek parameters which will produce A or B.

The second important factor, which is linked with the one above, is that of information fields. The information in any cytoform is most unlikely to be isotropic, *i.e.*, uniformly distributed, but instead will show various degrees of anisotropy (Fig. 4). We have observed that in cytological preparations the information content frequently increases, whether at the level of cytoplasm, nucleus, or subnuclear forms, as we approach the periphery, rather than at the centre of the particular structure. These observations are supported by the work of McMaster (1965). It is clear that in order to measure this type of information field different parameters must be adopted from those which measure isotropic information fields.

For these reasons we have set out to measure a large number of cell parameters. Most of them are relatively new in cytology and aim at giving a much more sensitive measure of symmetry, shape, and contour development of a cell; others measure the position and eccentricity of the nucleus with respect to the cytoplasm; yet others measure the optical density of various components at different wavelengths.

FIG. 4. Information fields: *a* = anisotropic; *b* = isotropic.

APPARATUS AND METHODS

If a large number of different measuring devices are to be applied to a single preparation of exfoliated cells, the conventional stained slide becomes relatively useless. From an information point of view it is a very inefficient way of presenting material for scanning since the distribution of cells is highly uneven, giving rise to waste space where there are no cells, and regions where the cells are too clumped to be discernible. For these reasons this concept was rejected and instead a cytotrack was introduced. Cytotrack is the backbone of Cesar (Figs. 5 a, b, c). In its present stage it comprises a 6-foot long rail with a longitudinal groove carrying a 35 mm. perforated uncoated cellulose acetate or polyester film. Material taken from the cervix with an Ayre's spatula is suspended in a suitable fluid into which a stain may be introduced if necessary. The film is driven at any required speed along the rail while a narrow track, which may vary from 0.1 to 0.7 mm., is laid along it by a specially developed apparatus, which consists of a

FIG. 5a, b, c. Diagrammatic representation of (a) cytotrack film; (b) cytotrack rail with transport mechanism and track laying apparatus; (c) detail of track laying apparatus.

high precision two-leafed mapping pen attached to micro-manipulators. The suspension carrying the exfoliated cells is fed gradually between the leaves of the pen, while the whole liquid index is stirred by a simple pneumatic device. Each sample is followed by a few clear washes which ensure that all the cells are exuded. Deposition carried out in this way produces unclumped and well dispersed cells. Under certain operating conditions a limited amount of segregation according to size takes place. The larger cells tend to occupy the central portion of the track, while the smaller ones move toward its upper and lower boundaries. It is possible to lay a number of parallel cytotracks with this apparatus and still retain enough space to carry at the edges magnetic tracks, which may be used for storing patient data, or be used as markers for critical recall to any given position along the longitudinal axis.

Cytotracks can be laid irrespective of the means of collecting the cells. The choice of fixative or preservative is governed by best retention of cell structure and shape and staining consideration rather than by track laying limitations.

A wide range of pH permits acid, neutral, or basic staining and a further advantage of this method is the ability of staining in suspension. This is particularly applicable to Fuelgen staining where uniform staining is frequently very difficult.

Alternatively staining, differentiation, or leaching and restaining may be easily accomplished after track deposition by passing the film through a series of troughs. The cells dry in air, and although no mountant has been used, the track with the cells in it remains well anchored to the film base. Although spray varnishes can be used, we have found that they are not necessary: so far our cytotracks remained without any visible deterioration for a period of six months. The use of thin film with no cover is very advantageous to subsequent microscopy. Not only does it reduce considerably internal scattering, heat absorption, and optical errors, but it also permits direct oil immersion work. Surplus oil left on the tracks appears to have no effect on the track itself, but usual precautions against accumulation of dust must be observed. The general feature of cells in a cytotrack is their enhanced cytoplasm-background contrast. This is most likely due to staining in suspension and the absence of glass surfaces.

The reading of cytotrack is achieved by a number of scanning stations, most of which can be energized by light or ultra-violet from one central source. This source can be conveniently provided by a plasma column issuing from a constricted nozzle plasma jet and/or radio frequency plasmas (Tylko, 1966). These devices emit copiously in the visible and ultra-violet regions. The required intervals of the spectrum can be delivered to the cytotrack by relatively simple means.

A number of scanning stations are placed above the cytotrack along the rail. These employ different types of sensors and operate at different magnification levels, each contributing a different type of information. Briefly we are aiming at developing three groups of sensors.

The first operates at very low magnification and its object is to detect the presence and quality of cells, the degree of clumping or overlapping and folding of the superficial population, and also distinct colour changes.

Its role is similar to that of scanning a slide at low magnification. A simple and very cheap type of such a scanner, based on a photodiode matrix and the Schmidt trigger circuit, has been constructed and used on cytotracks stained with Papanicolaou. It is possible with this arrangement to count the individual cells and differentiate between them on the basis of their colour by simple colour filtering.

The second group of sensors operates at a higher magnification level, comparable to that of $\times 300$ to $\times 600$. We are considering here, in addition to the well proven flying spot type of scanning with its numerous variants (McMaster, 1965), a number of new approaches. For instance, the use of an additional cylindrical lens purposefully distorts the image by elongating it approximately $5 \times$ in the direction of film travel. This device allows a photoelectric matrix sensor more time to lock itself onto the objects.

By using very powerful sources of light such as those generated by RF plasmas it is possible to produce an image of the cell on a phosphor screen with special decaying characteristics. Once the image is produced the sensor reads it from the phosphor screen without any reference to the actual cyttrack. Controlled decay phosphors have a further advantage in this context, because it is possible to take their light values at fixed intervals of time when various less bright portions of the image have disappeared and in this way to obtain measurements of areas of various light intensities.

It is also possible by a slightly displaced imposition of a positive and negative image to obtain a sharply defined contour, similar to that of the phase contrast microscope, which in turn may be scanned to give the length of any perimeter.

Another important type of contour sensor with which we are experimenting is based on the use of a family of curves known as quasi-ellipses (Fig. 6). These can be plotted (or generated electronically) according to the simple equation:

$$(x/a)^n + (y/b)^n - 1 = 0$$

When $n = 2$ the quasi-ellipse becomes an ordinary ellipse, and when $n = \infty$ the quasi-ellipse becomes a rectangle. It is possible to inscribe a cell of any contour into a given quasi-ellipse so that there are three points of contact of which two are the extremities of the longest chord in a cell. One may inscribe in this way a given cell into many types of quasi-ellipse but there will be only one value for n , for which the differences between the area of the quasi-ellipse and cell will be at a minimum. Such a measurement gives a sensitive measure of contour development.

Sensors at this level may also be made to measure selectively different colours. This is particularly important to us since we are aiming at combining fluorochrome stains such as Acridine Orange with an ordinary differential stain like that of Shorr, or a simplified Papanicolaou method. Using this technique one increases not only the amount of information, but also its reliability. Thus the fluorochrome image can be differentially scanned for the green, yellow, and red areas respectively.

The third and last group of sensors aims at extracting some information about the volume of the cytoforms rather than their flat images. A certain amount of information may be obtained by relatively simple and well known

FIG. 6. Family of quasi-ellipses.

techniques such as 'underfocusing' and 'overfocusing', *e.g.* by allowing the objective to oscillate at a predetermined amplitude. Readings from such sensors can be called upon specially in cases where the differentiation between the end-on image of an elongated cell or a spherically shaped cell are difficult.

This group includes the moiré pattern type of sensors in which trains of light pulses, *e.g.*, generated by linear gratings, are transmitted jointly with the cell image. The sensors are placed at different angles of elevation and inclination to the plane of the cyttrack with consequent production of different rates of pulses. When a cell enters such a field a disturbance occurs in the train of pulses which is in turn read by the different sensors. From the known positions and inclinations of the sensors and the change in the pulsating pattern it is possible to extract some topological data. We are also using moiré pattern principles in discerning fine structures of nuclei and cytoplasm, *e.g.*, vacuolization or chromatin clumping.

In order to evaluate our sensors and optimize their design a considerable number of manual observations using ordinary microscopes has to be carried out. These measurements are carried out on particularly carefully prepared smears, with well defined cell populations. The measurements are carried out simultaneously by standard microscopy using calibrated gratules and micrometers, from photomicrographs and planimetry, and from projection microscopy. In addition to the areas of the cytoplasm and nuclei, we are measuring a number of new parameters of which we shall mention here the most important.

In every cytoform or artefact of a given contour S , it is always possible to find at least one chord of maximum length. In the rare cases where there is more than one such chord, the ambiguity will not affect the statistically valid argument. Denoting this chord by L.C. (longest chord) we can derive a number of important relations which are summarized in Figure 7.

Having defined the L.C. we bisect it at a point p and note the two intercepts made at the points of contact with the 'upper' and 'lower' perimeter of the cell.

Ignoring for the time being the nucleus, we can now

FIG. 7. Typical measurements in a cell leading to the evaluation of moments of inertia and other characteristic parameters.

define the moment of inertia with respect to the L.C. in the usual manner, by summing the products of the elements of masses and the squares of their respective distances from the L.C. axis, thus: —

$$M.I.L.C. = A_1x_1^2 + A_2x_2^2 + A_3x_3^2 + \dots = Ax^2 \dots (1)$$

It is important to remember that instead of area we can substitute mass or some measure of light absorption or any other continuous property within the boundary S.

We now extend a similar argument to the nucleus, N, whose M.I.L.C. is referred also to L.C. The 'specific mass' of nucleus is weighted differently from the mass of cytoplasm. We have now a two-body moment of inertia which will represent a value very sensitive with respect to the periphery.

In addition we have also the values of the two intercepts PQ and PR, which will give us, in the case of exfoliated cells, an important measure of symmetry.

Furthermore, we can now also calculate moments of inertia about the QR axis.

Each of these parameters, *i.e.*, the longest chord, the perpendicular QR at the midpoint P, the intercepts QP and PR, and similar measurements in the nucleus, all convey a certain amount of information which helps to differentiate between cells. These extremely simple measurements also allow us to compute first moments of area (centroids) and mean heights of the 'upper' and 'lower' portions of the cell as well as the 'left' and 'right' portions *i.e.*, mean heights of the LQC; LCR; LQR and QCR areas respectively. The eccentricity of the nucleus is obtained in terms of the distance of the nucleus centroid from the longest chord and from the perpendicular QR.

The above operations giving the topological and (optical) densitometric data are all capable of being automatized, while the overall duration of such operations which hitherto was the main problem of single parameter cytoanalysers, can be considerably improved by relying as far as possible on low magnification scanners, calling in

the high magnification ones only when necessary. Needless to say, this mode of operation could only be accomplished by relying on a narrow cytotrack.

DISCUSSION

The next 10 years are likely to see a vast increase in the demand for routine cytological investigations. It is not difficult, for example, to forecast requests for urinary scanning on workers in selected chemical industries or routine sputum examinations on cigarette smokers. Any cytological service which is offered has certain prerequisites. It must be accurate without appreciable false positive or false negative results. It must be cheap, sparing of manpower, and speedy. Storage of material and back-references must be readily available, compact and easy. We do not think that all of these desiderata can be met by the present methods of manual scanning and individual interpretation. Two factors seemed of paramount importance. First, there must be some system of language for classifying all possible cell patterns and coding them so that each can be referred to without ambiguity and so that the description will be as precise for a physicist or mathematician as for a cytologist. This we feel has been solved by the adoption of a cybernetical system and the use of a cytomorphic code, which is capable of indefinite expansion to accommodate any cell.

The second factor concerns the placing into its appropriate category of each individual cell which is shed. Of the methods at present used in cytology the most accurate is the combination of human eye and brain which, without conscious analysis, probably make use of four or five different cell parameters. The mechanical methods known to us utilize one, two, or three parameters. To obtain a really high information rate about a cell it is necessary to measure with reasonable accuracy a very large number of individual parameters, and to look for differences in pattern rather than similarities. A limited number of parameters only are available for measurement by conventional techniques and besides making full use of these it is necessary to make use of deliberate distortions and unusual methods of measurement to obtain maximal information. Such techniques involve rethinking the conventional methods of presentation of material and have led to the concept of cytotrack. There is no doubt that further experiments will lead to alternative methods of staining or to other methods using unstained cells. What we wish to emphasize is that when large numbers of cells have to be scanned reliably and quickly a multi-parameter scanning device feeding its observations into a computer system is a more rational approach than a large number of readily fatigued human eyes and brains which can find better employment elsewhere. With

the approach suggested above, the large number of parameters does not appreciably slow the processing time and therefore appears to offer the best of both worlds. At the same time the Cesar machine does not isolate the human operator. The cells on the cytotrack can be viewed by any orthodox microscope arrangement or projected onto a screen.

At the present stage (1966) the Cytotrack itself and the simple sensors are working satisfactorily. Other sensors are only in the early stages of development.

Finally, we feel that if the cytotrack principle could be generally adopted, a very good start would be made, leading not only to the eventual complete automation of routine scanning, but to what in the long run is perhaps even more important, to the gathering of more reliable quantitative cytological data in general.

ADDENDUM

Since this article was submitted several improvements have been made in the collection of specimens and the laying of Cytotrack.

After cervical scraping, the end of the Ayre spatula is now deposited directly into a disposable plastic container containing preservative with

simple nuclear strains incorporated as required. The conical end of this container plugs directly into a disposable plastic graphion or mapping pen which is automatically adjusted to lay the cells onto 35 mm. film. This method eliminates any possibility of cross contamination.

REFERENCES

- Bostrom, R. C., Sawyer, H. S., and Tolles, W. E. (1959). In *Medical Electronics. Proceedings of the Second International Conference on Medical Electronics*, Paris 24-27 June 1959, edited by C. N. Smyth, p. 479. Iliffe, London.
- Iversen, S., Gordon, W. J., Cowell, M. A. C., and Watson, E. R. (1966). *Brit. med. J.*, **1**, 1209.
- Ledley, R. S., and Ruddle, F. H. (1966). *Scientific American*, **214**, 40.
- McMaster, G. W. (1965). *A technique for the digital scanning of normal and malignant cervical cells*. Personal communication.
- Mellors, R. C., and Silver, R. (1951). *Science*, **114**, 356.
- , Glassman, A., and Papanicolaou, G. N. (1952). *Cancer (Philad.)*, **5**, 458.
- Sawyer, H. S., and Bostrom, R. C. (1958). *I.R.E. National Convention Record*, Part 9, p. 37.
- Schmitt, A. (1953). *Geburtsh. u. Frauenheilk.*, **13**, 593.
- Smolka, H., and Soost, H. J. (1965). *An Outline and Atlas of Gynaecological and Cytodiagnosis*, 2nd ed. E. Arnold, London.
- Tolles, W. E., and Bostrom, R. C. (1956). In *Airborne Instruments Laboratory Inc. Some Theoretical and Practical Aspects of Microscanning*. July 1956, Mineola, New York.
- , — (1956). *Ann. N.Y. Acad. Sci.*, **63**, 1033.
- Tylko, J. K. (1966). *Improvements in or Relating to High Temperature Apparatus*. Prov. Spec. No. 39236/66.