Improved Papanicolaou smear reporting through the use of automated data entry

GPS Yeoh, KW Chan, WY Ng

The implementation of an automated data entry and report generation system using an optical scanner and commercially available image processing program is described. This method could be easily adapted for use in other fields of medical research where the compilation of a large amount of repetitive data is involved, such as the filling in of questionnaires. Using an optical scanner for data entry improves the efficiency of report generation, thereby improving the turnaround time of reports. Reports are standardised and more easily understood by referring doctors. Data is also standardised and validated and is more amenable for quality assurance analysis, in the reminder service for patients, and gives a performance analysis of smear takers.

HKMJ 1997;3:256-60

Key words: Papanicolaou smear; Cervical intraepithelial neoplasia; Primary prevention; Automatic data processing

Introduction

The use of the Papanicolaou (Pap) smear for cervical cancer screening has resulted in a dramatic reduction in the mortality due to cervical cancer in many western countries. The most impressive achievements can be seen in Scandinavian countries and in British Columbia, Canada, where organised programmes are used. Moderate improvement has been seen in Australia and Hong Kong, where less well organised systems are in place. The mortality rate for cervical cancer has decreased from 9.6 to 4 per 100 000 patients in Queensland, Australia, following promotion of the Pap smear.[1] The age-standardised incidence of cervical cancer in Hong Kong was 14.1 per 100 000 and the mortality rate was 3.9 per 100 000 in 1992.[2]

An organised cervical cancer screening programme is a complex multi-disciplinary process involving medical and paramedical health professionals, including epidemiologists, educationalists, and publicists. The establishment of a central registry and database is required to coordinate the programme. However, the less complex ad hoc model of opportunistic screening that exists in Hong Kong, involving patients, physicians, and laboratories, can be effective if properly co-ordinated. The laboratory can play a central role in co-ordination and thereby improve cervical cancer screening through the effective use of information technology and data management to ensure quality assurance and good patient follow up. We report our experience in the design and implementation of an automated cytology data entry system in our laboratory in an attempt to help improve cervical smear reporting.

Histopathology Unit, Canossa Hospital, 1 Old Peak Road, Hong Kong
GPS Yeoh, MB, BS, FHKAM (Pathology)
KW Chan, MB, BS, FHKAM (Pathology)
WY Ng, B Appl Sc

Correspondence to: Dr GPS Yeoh

Materials and methods

System design

The current practice in many cytology laboratories in Hong Kong is to issue non-standardised written reports to clinicians. These reports, however, are difficult to index and retrieve. The difficulty in retrieval can be overcome by including coding such as that of the SNOMED system or by standardising the reports. Our objectives in designing the automated data entry system were to:

1. Use commercially available optical scanning devices and programs to minimise costs and development time.
2. Develop a one-step automated data entry procedure to minimise transcription errors.
3. Validate data.
4. Compartmentalise data to enable easy manipulation, retrieval, and analysis to be done.
5. Create a written report from the scanned-in data.

6. Interface with the existing laboratory information system.

Hardware

The current computer systems in our laboratory consist of a server and a network of six personal computers (PCs). All PCs are IBM clones with 486 processors and hard disk drives of 250 to 1100 megabyte capacity. The operating system used is DOS 6.22. The scanner (Fujitsu ScanPartner 10, Fujitsu Computer Products of America Inc., San Jose, Ca, US) has a sheet feeder and can scan 10 pages per minute. This is attached to a PC with 8 megabytes of RAM through an Adaptec SCSI card interface. Each PC is connected to an HP laser printer for the printing of reports.

Software

The form is created and printed on A4 paper by a commercial printer. A commercially available Windows-based program called FORMation (Sea Dragon Design, Heathmont, Victoria, Australia) is used for controlling the scanner and translating the marks on the form into meaningful data. The program achieves this by a page mapping process that relies on a form containing alignment markers (Fig 1). The program allows the user to easily define and compare the optical density of various regions of the form. Undefined areas of the form are ignored, which allows for more efficient use of computer resources. The threshold of each region is also user-definable. The program outputs user-defined data separated by carriage returns depending on whether a particular region has been marked.

Once the data have been captured, it is passed on to an in-house validation program (written using Powerbuilder) for verification. Considerable care has to be taken to ensure that the data acquired is exactly as marked on the form. Validation rules have to be constructed to detect not only extraneous data (i.e. errors in data capture) but also human error in filling in the forms to ensure that no nonsense data or reports are generated. All possible combinations of marks on the form have to be checked. As an example, the combination of no malignant cells seen (NMCS) in the Cancer Reading section and C for colposcopy under the Recommendations section is not allowed and the program will flag this as an unacceptable combination. The construction and testing of the validation rules was tedious and time-consuming, but the end result is a program with intelligence. The validated data are then placed in the appropriate data cells in a temporary database before export into the existing laboratory information system, which is an in-house database program written in FoxBase (DOS version) for report printing.

Performance assessment

For a two-week period, the time taken for manual data entry (start time and end time) by each typist and the number of reports generated were recorded. Reports that were rejected by the pathologists at 'signing out' and that needed amendments during this time were collected and analysed at the end of the trial period. A similar exercise was repeated over another two-week period using the automated system (Fig 2).

Results

System performance

The total time required to generate 1554 reports by the manual method during the two-week trial period was 51 hours 40 minutes or about 120 seconds per report, compared with 54 seconds using the automated process (1622 reports in 24 hours 21 minutes), an improvement of 222%. The manual data entry, which was performed by two typists could be done by one typist using the automated system.

System accuracy

The scanner accurately reads the marks. During a pilot study prior to routine use of the automated system, four marker errors were found out of 234 sheets, each with 175 marker regions i.e. four of 234 x 175 marks read, giving a scanner error rate of 0.01%. Human error in filling the dots on the sheets was more common, with 15 errors on 234 sheets (6.4%). All the errors (scanner and form filling) were detected by the validation program. After one month of operation, the error rate in form filling was reduced to about 3.0% (48 sheets of 1622). This compares well with the approximately 6.0% data entry error rate found (93 reports of 1554 requiring amendments) using the manual system.

Discussion

A dramatic reduction in mortality due to cervical cancer can be achieved through the use of an organised Pap smear screening programme. In the absence of a mass cervical cancer screening programme, laboratories with properly organised databases can play a role in improving cervical cancer screening. The database can be used to improve quality assurance. The main impediment for laboratories in the past has been the relatively high capital and recurrent cost involved in setting up and maintaining the database and the labour cost of data entry. Optical Mark Reading (OMR)

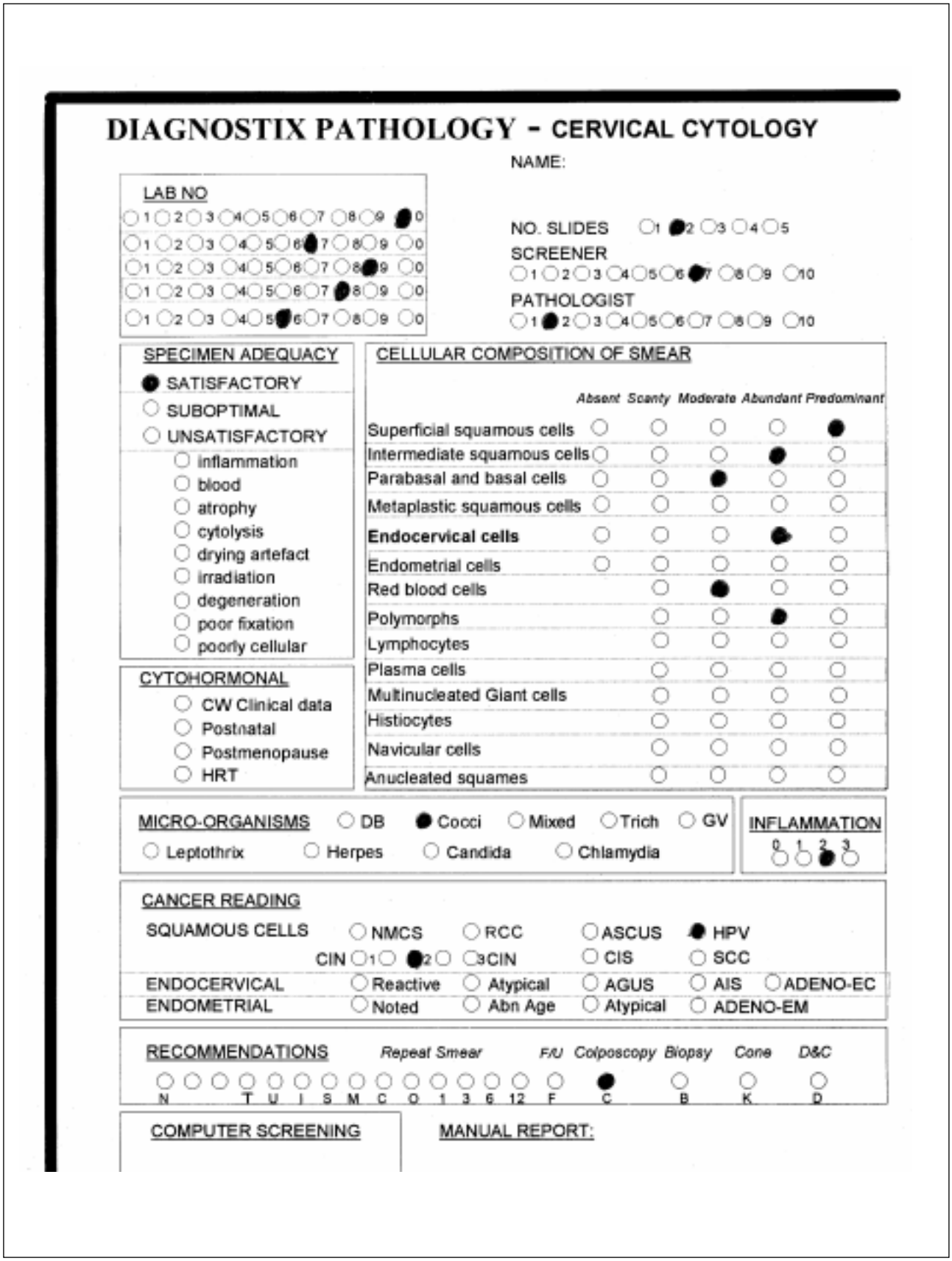
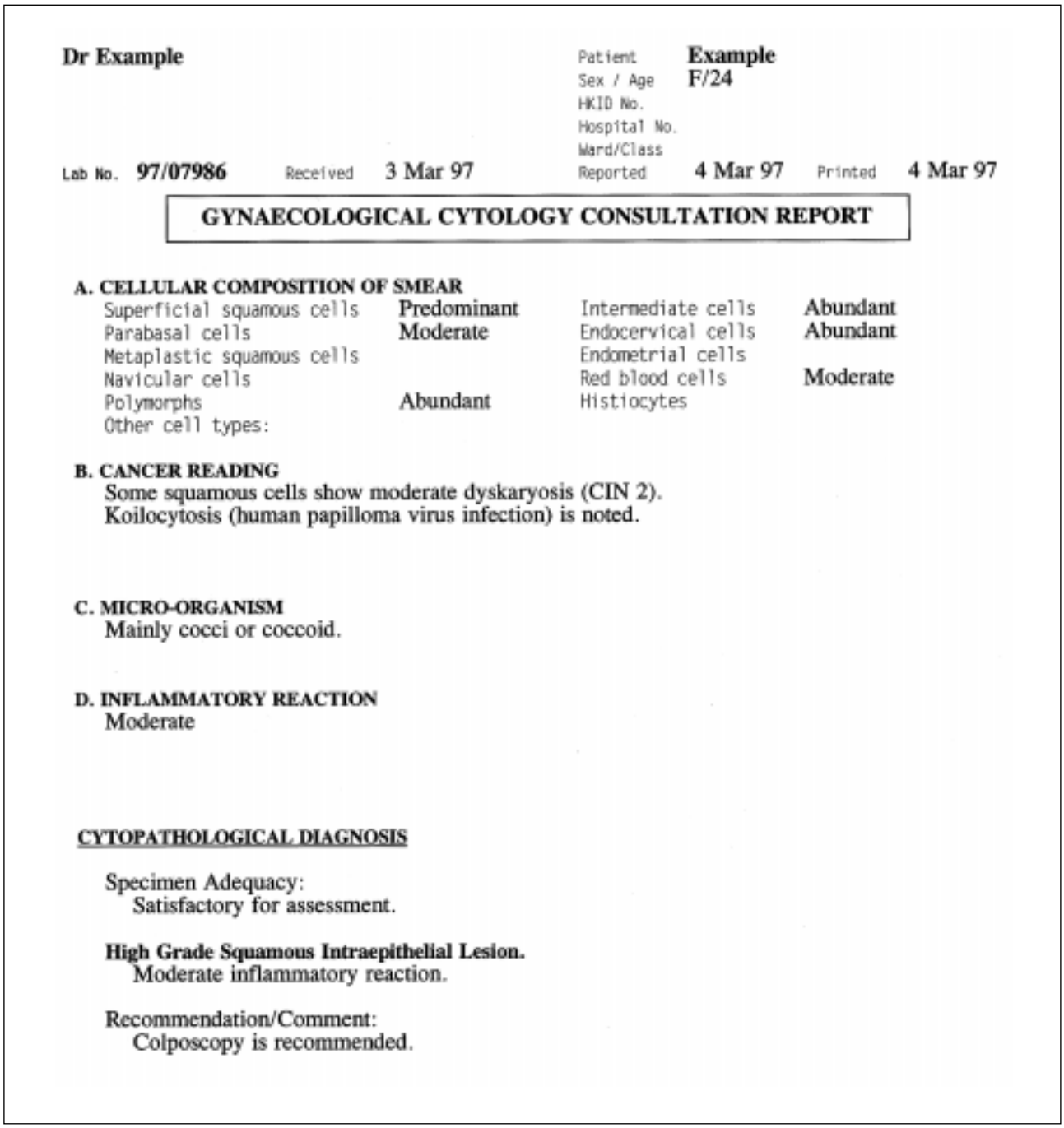


Fig 1. Pap smear worksheets. The form is filled by blackening the appropriate circle with a 2B pencil. Completed forms are scanned, the data extracted and validated. All data in the form are reported except for the number of slides and screener information, which is stored in the database for laboratory reference only

has been widely used for automated data entry. This technology is well established and reliable. However, the capital cost of an OMR card reader is relatively high and the specially printed cards are expensive. Image scanners for PCs are now used in place of the traditional OMR reader. The advantage of using a scanner is the significantly lower capital cost (10%-20% of the cost of an OMR reader) and the cheaper running cost. Forms printed on A4 paper may be used in place of the specially printed OMR cards. With the dramatic fall in the cost of hardware and the availability of commercial optical scanning programs, an automated data entry system can be set up for less than HK$30 000.

One benefit of the set form and pre-coded comments is the standardisation of reporting in our laboratory. The coded comments are agreed upon after much discussion between the pathologists and cytotechnologists involved. The use of standardised terminology means that referring doctors are familiar with our reports, regardless of who the reporting pathologist is. Although the majority of our reports are



Dr Example

Patient **Example**
Sex / Age F/24
HKID No.
Hospital No.
Ward/Class

Lab No. **97/07986** Received 3 Mar 97 Reported 4 Mar 97 Printed 4 Mar 97

GYNAECOLOGICAL CYTOLOGY CONSULTATION REPORT

A. CELLULAR COMPOSITION OF SMEAR

Superficial squamous cells	Predominant	Intermediate cells	Abundant
Parabasal cells	Moderate	Endocervical cells	Abundant
Metaplastic squamous cells		Endometrial cells	
Navicular cells		Red blood cells	Moderate
Polymorphs	Abundant	Histiocytes	
Other cell types:			

B. CANCER READING
Some squamous cells show moderate dyskaryosis (CIN 2).
Koilocytosis (human papilloma virus infection) is noted.

C. MICRO-ORGANISM
Mainly cocci or coccoid.

D. INFLAMMATORY REACTION
Moderate

CYTOPATHOLOGICAL DIAGNOSIS

Specimen Adequacy:
Satisfactory for assessment.

High Grade Squamous Intraepithelial Lesion.
Moderate inflammatory reaction.

Recommendation/Comment:
Colposcopy is recommended.

Fig 2. Example of a report generated by the system corresponding to the blackened circles in the form illustrated in Figure 1

generated by the automated process, we have on occasion issued non-standardised manual reports for smears with equivocal changes as it is not cost-effective to design a comprehensive form to cover all possible situations.

One of the performance indicators of a pathology laboratory is the turnaround time. With the implementation of the automated data entry for cervical cytology, our typists are able to transcribe the histology dictation sooner. This has resulted in improved histology turnaround times.

Laboratories have long recognised the need for quality assurance (QA) and QA programmes are commonly present in clinical pathology laboratories. However, QA in cytology laboratories is more difficult to implement due to the subjective nature of the interpretation and the predominantly text-based reports, which are difficult to analyse statistically. These problems can be overcome by breaking down the reporting of cervical smears into small components that are more amenable to meaningful comparison and study. Breaking down the reports into smaller components means that more information has to be entered, hence the need for an automated process. The resultant database can then be used to improve the screening process through QA programmes for the laboratory and clinicians. Statistics on the workload and performance of individual cytotechnologists and pathologists can be produced at regular intervals. All high-grade lesion reports can be retrieved at regular intervals and followed up. Statistics can be provided to clinicians on the presence or absence of endocervical cells in smears. A clinician's rate of endocervical cell content in smears can be improved by being given feedback statistics.[3] Laboratories can assist clinicians by making sure that women requiring further attention are followed up. The database can be used to flag all patients who are overdue for follow up smears and reminder letters can be generated and sent to the relevant clinicians. Reminder systems have been shown to increase the number of patients returning for overdue Pap smears.[4]

The system of data entry described here can be easily adapted for use in other fields of medical research where the projects involve the collection of vast amounts of repetitive data such as surveys, questionnaires, and standardised treatment protocols.

References

1. Free K, Roberts S, Bourne R, et al. Cancer of the cervix—old and young, now and then. Gynecol Oncol 1991;43(2):129-36.
2. Hong Kong Government. Hong Kong Cancer Registry Annual Report, 1992. Hong Kong: Hospital Authority, 1996.
3. Yeoh GP. Factors affecting the presence of endocervical cells in cervical smears. Proceedings of the 20th Annual Scientific Meeting of the Australian Society of Cytology; 1990 Oct 12-14; Brisbane. Brisbane: Australian Society of Cytology, 1990.
4. McDowell I, Newell C, Rosser W. Computerized reminders to encourage cervical screening in family practice. J Fam Pract 1989;28:420-4.