Reprinted from the "SA Medical Journal", Vol. 50, 23 October 1976, pages 1832 - 1836

LKW 78 (Byvoegsel—Suid-Afrikaanse Tydskrif vir Laboratorium- en Kliniekwerk)

An Integrated Cardiorespiratory Pathology Information System

C. L. SOSKOLNE, B. GOLDSTEIN, I. A. HAFFAJEE, GRACE COWLING

SUMMARY

An integrated computer system which is in its initial year of implementation has been developed to handle pathological data of cardiorespiratory organs. On the single entry of data captured at the source, the interactive computer system generates reports for administrative requirements by law, updates a data base of cardiorespiratory histopathological data for research purposes and facilitates inquiry for periodical management reports. The system is based on multiple-source documents which have numerically precoded choice option statements as well as free text provision for those statements where precoded options are not provided. The computer system has replaced manual system procedures most effectively in terms of cost.

S. Afr. med. J., 50, 1832 (1976).

The implementation of the Pneumoconiosis Act in 1956 required that a standardised method of examination of the cardiorespiratory organs of deceased miners be instituted. Since that time it has been possible to compare pathology reports and to assess changes in the degree of pneumoconiosis and the severity of chronic obstructive lung disease. By 1959, the information contained in the examination reports was recorded for subsequent statistical analysis. The system of the edge-punch card was introduced and, with various refinements, was still operational in 1974.

The use of the edge-punch card is described in related applications in the literature.[1-3] Demands on the pathologist's time in clipping the cards and checking the manually typewritten reports in the NRIOD application were, however, disproportionate to the benefits of such a system. The data could not be retrieved easily and the demands on human resources continued to grow.

Apart from the difficulties in statistical analysis by means of the data base established in edge-punch card form, considerable manpower was required to compile and check close to 3 000 reports which were prepared annually and which had numerous phrases in common.

With the establishment of a Department of Biostatistics at the NRIOD in mid-1973, current procedures at the Institute were appraised from the data processing viewpoint. The inter-relatedness of medical and administrative information[4] was seen to apply.

As the pathology techniques and reports had been standardised, it was apparent that the arrangement of the terminology on a single-source document could with advantage yield the standard tabular-selection-(menu-) type of application for each autopsy. An extensive pathology data base with rapid and reliable access could thus be established. The postmortem data, related to the industrial service record and the periodical medical examinations during life, have proved to be an invaluable source of research material. The greater utilisation of pathology data is proving of value in research generally[5] and is essential to computer-assisted diagnosis.

The management system described in this article was developed to utilise the various levels of expertise and disciplines at the NRIOD more efficiently. An integrated information system was designed which would, on a single capture of data at its origin and a single data-entry operation, generate typewritten reports, update the data base and facilitate inquiry.

FLOW OF INFORMATION AND ORGANISATIONAL REQUIREMENTS

The application of modern computer technology in medicine is not new, but at the NRIOD such procedures as described here are an innovation. Histopathology lagged behind other fields which have been automated for 'in order to utilise the histopathologic data mechanically, elaborate systems of data coding must be devised'.[6] Experience has led to standardised questionnaires which lend themselves to numerical coding. Much emphasis has also been given to the expeditious furnishing of reports in clinical pathology environments in general.[7] Histopathology at the NRIOD was seen to be well suited to automation. A management system facilitating *ad hoc* requests for information has great potential in the pathology of diseases. The one-time recording of data at its origin reduces possible transcription errors, giving a more valuable data base through only a moderate change (indeed, simplification) in the over-all organisation.

The 'Old' System

A minimum of 3 independent forms and 1 index card relating to the examination had to be identified. This clerical function was the responsibility of the Records Department. A summary of the deceased's file was also compiled, and the relevant information transcribed onto the face of the edge-punch card and on a summary sheet. Each of these forms was then dispatched to an examining pathologist. These forms required voluminous handwritten

Department of Biostatistics, National Research Institute for Occupational Diseases of the South African Medical Research Council, Johannesburg

C. L. SOSKOLNE, B.SC. HONS
B. GOLDSTEIN, M.B. B.CH., F.F.PATH.
I. A. HAFFAJEE
GRACE COWLING

Date received: 4 November 1975.

additions, since they were designed for manual assessment and not arranged in the tabular-selection-(menu-) type system. At the completion of the examination all forms were collated and the final report was typed manually, with up to 9 carbon copies for various bodies and certain registers. Once prepared, all documents relating to the final report were sent to the pathologist concerned to be checked and signed. While checking the report, the pathologist would clip the edge-punch card. A copy of the report together with the edge-punch card would then be sent to the Department of Biostatistics where an industrial history would be assessed from additional service records. A person with some knowledge of pathology was required to check the clipping of the edge-punch card against the typed report. Errors were brought to the attention of the examining pathologist for checking and correction. The summarised industrial history was recorded on the reverse side of the edge-punch card which was then filed away, according to year of death and type of mining. A disproportionately large amount of the examining pathologist's time and the time of numerous other personnel was required in the checking and recording of information. The integrated and automated system would enable respective personnel to perform the functions for which they were trained more fully.

The Integrated System

With the tabular-selection-(menu-) arrangement contained in a multiple-source document, the Records Section is required to identify the document only once. Unique source document serial numbers appear on each page of each source document.

On completion of all aspects of the examination, the forms are checked for clarity and completeness. The data, together with the industrial history, are entered via a telecommunications terminal to a large digital computer. With various edit routines built into the computer programme, most items which are rejected at the point of data entry are corrected within the Department of Biostatistics. Very few queries are returned to the examining pathologist. While several workers[8-14] have made extensive use of variable format, restricted format and natural language (free-text English) data-entry systems, the NRIOD system makes use of numerically coded data-entry primarily. This takes full advantage of the exhaustive lists of possibilities developed over many years of experience with a manual system at minimal computer cost as opposed to the costs involved in the development of thesauri and related software in the free-text approach. Besides, the significantly fewer data-entry operations required in the numerically coded approach itself makes the numerical coding method attractive, although substantial effort has to be placed in the development of the source documents. This point of view is not unique; Crocker,[15] in reviewing automation trends in anatomical pathology, points to the dilemma of choice between a natural language-based system and one of numerical coding. Coded diagnoses are the simplest to programme, requiring less sophisticated software as well as hardware.[16] Simon *et al.*[16] utilise coded language in their radiology reporting system.

System Procedures

The system is composed of 10 main procedures:

(i) *The Bureau library* checks for the uniqueness of each record.

(ii) *Bio and sum data* provide for standard fixed format personal history and service estimate data.

(iii) *Macro and micro data* provide for fixed format statement options relating to possible macroscopic and microscopic pathological examination findings.

(iv) *Additional comments list* produced from additional comments acceptable under each subsection allows for 260 character position statements not provided for in the standardised options. This ensures no restriction in the reporting of findings. A list is generated monthly for the perfection of the source document standard options provided for.

(v) *The occupation breakdown* or detailed service history generates a map on the hard-copy report of a man's exposure history.

(vi) *The update* procedure facilitates the skipping of complete sections where no data exist. The improved version of the system allows backward or forward skips while entering a new case; also, direct access of old records (disc-based) for correction of 'old' information or the addition of new information.

(vii) *The working list* generates a list of selected identifying cross-reference information relating to reports, computer-generated on special stationery.

(viii) *Monthly statement of account* containing summary facts relating to a month's reports are generated with a unit cost breakdown to satisfy accounting needs.

(ix) *The monthly stats file* procedure compacts disc-based records, writes onto tape and generates the additional comments list and monthly statement of account.

(x) *The master stats file* is a tape-based translation of the monthly stats file containing relevant statistical parameters for the creation of a cardiorespiratory pathology data base.

The Report Generation Programme and Method of Interactive Data Entry

The report generation programme system is written in PL/1 and consists of a main (calling) procedure, 4 data capture procedures, a report generating procedure and procedures generating hard-copy versions of information required for administrative purposes. The entire system is implemented on an IBM-370/158 computer. The programme requires a minimum of 230K of memory with the overlaying of 3 of the data capture procedures for execution. Access to the programme is via a typewriter terminal under the time sharing option facility.

The use of the pointer variable in the selection of character string options from the statement option library greatly reduces the amount of core storage required for the permanent storage of the numerous and lengthy standardised character string options.

Programmatically, allowance is made for the selection of singular or plural statement options and the entry of any 'additional comment' that the pathologist may wish to make. The following is an example:

PROGRAMME INTERACTION — PROMPT FOR PULMONARY VESSELS DATA

Request: *Enter macdata*: *pulmonary vessels.*
Reply: *2.2/3.2/6.*
Request: *Enter additional comment.*
Reply: *Are congested.*

A 'SCAN' procedure assigns each of the character string option codes to its relevant statement in the 'statement option library'. The 'additional comment' typed at the terminal is directly assigned to a variable name. The pathologist's macroscopic findings for pulmonary vessels in this example will then be interpreted by the computer as follows:

PULMONARY VESSELS
SHOW MODERATE ATHEROMA
THROMBO-EMBOLISM IS PRESENT ON THE LEFT SIDE
ARE CONGESTED

Data Base

In addition to the generation of postmortem reports, the programme creates fixed-length records of each case dealt with on a permanent stream-oriented file. Each record has a fixed linesize of 2 000 characters. The information contained in this 'record library' is used for the creation of a permanent data base called the 'master stats file' which is a tape-based translation producing statistical parameters through the application of diagnostic criteria to the statement options selected in the 2 000-character record reducing the record length to 748 characters for statistical research purposes.

The identification of each record is based on a unique number allocated by the Medical Bureau for Occupational Diseases. A 'bureau number library' is used by the programme to check the uniqueness of each record (case) before a report is entered at the terminal. Additional identifying numbers are also recorded with a view towards record linkage which it is hoped will be achieved in the future when the records of periodical clinical data are computerised and the industrial service records updated.

Why PL/1?

The character manipulation facilities of PL/1 are well utilised in the programme. Large-scale use is also made of the on record, on error and on endfile conditions. The former two conditions allow data entered incorrectly to be re-entered. To varying degrees of complexity the if-then-else, do-end, do-while, begin-end combinations and the substring, length and index character manipulation built-in functions have been implemented in the programme. Also, use is made of a print file to allow a neat format despite the variable length of reports (Fig. 1).

SOUTH AFRICAN MEDICAL RESEARCH COUNCIL

NATIONAL RESEARCH INSTITUTE FOR OCCUPATIONAL DISEASES

TELEPHONE 724 0331
P.O. BOX 4788

JOHANNESBURG
2000

PATIENT: X X X AGE 36 BUR NO: X X X
SERVICE: REF NO:RBL-3155002-1
PLATINUM: 8,9YRS. DATE OF DEATH: 02.08.75 P NO: 8413/75
GOLD : 6,3YRS. DATE RECEIVED: 13.08.75 LAB NO: 4858/75
DATE OF EXAM: 03.09.75
DATE OF REPORT:02.10.75

CAUSE OF DEATH: RUPTURED AORTA.
REPORT ON THE EXAMINATION OF CARDIO-RESPIRATORY ORGANS.

MACROSCOPIC REPORT **********
LUNGS
RIGHT LUNG: 744G, LEFT LUNG: 342G.
HEART
WEIGHT: 262G.
REGIONAL GLANDS
ARE MODERATE ENLARGED AND MODERATELY PIGMENTED.
ARE FIBROSED.
BRONCHI
SHOW NO GROSS CHANGE.
PULMONARY VESSELS
SHOW NO GROSS CHANGE.
PLEURAE
ARE MODERATELY PIGMENTED.
THERE ARE NO PALPABLE PLAQUES.
TAGS OF FIBROUS ADHESIONS AND A CHRONIC NON-SPECIFIC PLEURISY ARE PRESENT OVER THE RIGHT SIDE.
LUNGS ON SECTION
SHOW A MODERATE INCREASE IN PIGMENT.
THE PIGMENT IS AGGREGATED IN DISCRETE NODULES AND A FEW OF THESE ARE PALPABLE.
SECTIONS HAVE BEEN TAKEN.
THERE IS NO EVIDENCE OF EMPHYSEMA.
THERE IS NO OVERT TUBERCULOSIS IN THE LUNG SUBSTANCE.
HEART
IS NOT ENLARGED.
IT DOES NOT SHOW ANY SIGNIFICANT PATHOLOGICAL CHANGE.
MICROSCOPIC REPORT **********
REGIONAL GLANDS
THE SECTIONS SHOW:
DIFFUSE FIBROSIS IN WHICH THERE ARE FOCI OF CONCENTRIC FIBROSIS.
LUNGS
SECTIONS TAKEN FROM PALPABLE NODULES IN THE LUNG SUBSTANCE SHOW:
PARENCHYMAL LYMPH NODES IN WHICH FOCI OF SILICOTIC COLLAGENISATION ARE PRESENT.
THERE IS INTRA-ALVEOLAR HAEMORRHAGE AND OEDEMA.
BRONCHI AND BRONCHIOLES
THERE IS AN INSIGNIFICANT DEGREE OF CHRONIC BRONCHIOLITIS.
A SECTION OF THE MAIN BRONCHUS SHOWS AN INSIGNIFICANT DEGREE OF GLAND HYPERPLASIA.
I AM OF THE OPINION THAT **********
SILICOSIS IS PRESENT IN THE LUNG PARENCHYMA.

DR E. KAGAN
MEDICAL OFFICER

2086 HIERDIE VERSLAG IS OOK IN AFRIKAANS VERKRYGBAAR

Fig. 1. Computer-generated medicolegal report.

Programme: General

The programme checks for inconsistent data, selects the 'option' statements programmatically for standard portions of the report, selects/corrects the opinion for cases where this is necessary and prints specially formatted reports on special stationery.

The cost for the data capture and report generation of one case on special stationery is approximately R0,75 (for a 100-line report).

With the single entry of data into an interactive terminal facility (on-line), standard reports on special stationery are computer-generated and typed, and the data base is updated. Reports are generated on special pre-printed 3-part stationery (Fig. 1). The final report is perused within the Department of Biostatistics and approved. The report, together with the collated source document is then presented to the pathologist for his signature.

The system is depicted in flowchart form in Fig 2. The three integrated aspects of the system are those of report generation, data base updating on single data entry and inquiry.

To implement this automated system, the 'new' system was introduced over two stages. The first stage involved 3 months over which time programmes and the system were improved, programme irregularities eliminated and staff educated in and familiarised with the new procedures.

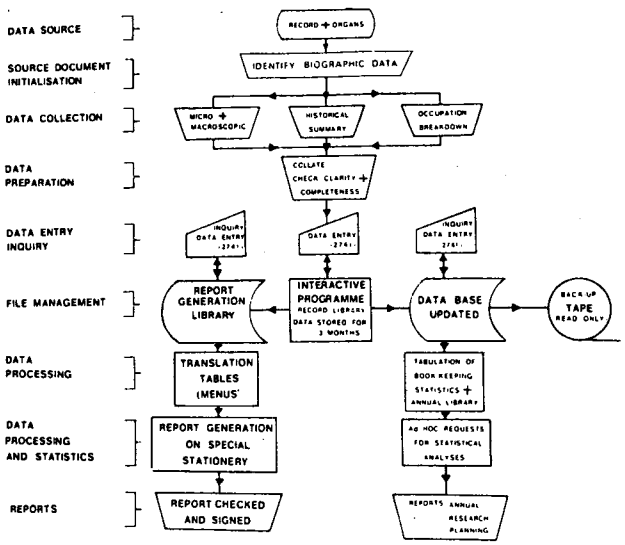

Fig. 2. Integrated cardiorespiratory pathology information system. Report generation, data base update and inquiry flowchart.

ADVANTAGES

Direct and immediate advantages include the following:

1. The new source documents are more readily completed, saving up to 50% of the pathologist's time and allowing him to devote more of his time to management and research.

2. The need for continually expanding typing services, filing facilities and register maintenance is eliminated; the system requires as little as 5 - 10% of what was previously required in typing under the manual system.

3. The tedious and time-consuming task of checking is minimised; with checking at the source prior to data entry the likelihood of error is minimised.

4. As a management system, the facility of direct access inquiry encourages the use of the data base. The massive data base facilitates *ad hoc* requests for information. In the past, the data base on edge-punch cards was used for vital statistical purposes alone. Research is facilitated through the advanced statistical procedures (for example, the BMD and/or Crosstabs packages) that can readily be applied to the data-base.

5. The introduction of a computerised system with due cognisance given to data privacy reduces the risk of infringements on confidentiality. Computer-orientated staff are required to adhere to a minimum ethical code and programme locks are built into the data base.

6. The need to maintain individual manual registers on certain factors falls away. Up-to-date lists are generated as called for.

7. *Ad hoc* analyses, annual reports, statements of account, data necessary for planning, sophisticated and standard research analyses and reports are readily produced.

8. A large, accessible data base, essential for research into computer-assisted diagnostic techniques in cardiorespiratory disease, is being established and is growing daily through the integration of routine, service functions with research needs.

FUTURE CONSIDERATIONS

The ready access to an already existing remote job-entry terminal facility dictated the *modus operandi* to be employed in the initial development and implementation of the system.

It is anticipated that soon after the system is totally integrated into the NRIOD procedures and the staff are more computer-orientated, the introduction of video terminals (which will display the tabular selection scheme) may eliminate the need for an actual hard-copy source document. However, as the turn-around time on reports need not be rapid, the costs involved in a video system may not be justified.

In view of the extent to which this system has been developed, and the limited workload in terms of daily production, it is unlikely that a more 'dedicated' system would be introduced. The facility as part of a larger shared system will probably suffice for some years to come. The optimisation of the present system programmes to utilise direct access concepts such as access vectors for the statement options is currently being introduced.

DISCUSSION AND CONCLUSIONS

Such extensive automation in cardiorespiratory pathology is unknown elsewhere. The usefulness and effectiveness of the system has been borne out by the general agreement, even among the older examining pathologists, that the adjustment required in the completion of a new source document is more than balanced by the advantages derived from the integrated system. The release of the complete range of personnel — from typists to examining pathologists — for more interesting, less repetitious and hence more satisfying work, has been welcomed. Research in the field of pneumoconiosis is already under way, making extensive use of the continually updated data base. The integrated cardiorespiratory pathology information system serves also as an innovative procedure at the NRIOD to generate familiarity with computerisation which, in turn, is generating sufficient confidence in management to encourage computerisation and automation in other application areas within this and other institutions.

We wish to thank Mr Andrzej Dabrowski, of the National Research Institute for Mathematical Sciences at the Council for Scientific and Industrial Research, without whose assistance in the programming the system could not have been achieved so expeditiously, and Mr David H. Deane, of the Department of Biostatistics, who has been responsible for the development since writing of a more efficient version of the system programmes.

We also thank Mrs Pat J. O'Reilly for her part in the data entry aspects; all staff of the NRIOD whose work ensured success in the implementation of the system, in particular Mrs Vivian P. Mumford who has been responsible for all data control functions, and Mr Victor V. Tshandu whose

efforts have ensured the continual recording of historical data.

REFERENCES

1. Athreya, B. H., Athreya, R. B. and Coriell, L. L. (1970): Amer. J. Dis. Child., **119**, 53.
2. Kartashova, O. Ya., Pitirimov, B. V. and Sokolova, L. M. (1968): Arkh. Pat., **30**, 67.
3. Khaime, C. B. and Volosova, E. M. (1968): *Ibid.*, **30**, 63.
4. Lameer, C. and Fokkens, O. (1973): Radiol. Clin. Biol., **42**, 529.
5. Stosiek, P. and Schmidt, K. (1973): Zbl. allg. Path. path. Anat., **117**, 191.
6. Miller, A. S. and Pullon, P. A. (1969): Oral Surg., **28**, 702.
7. Editorial (1968): Lancet, **1**, 963.
8. Röttger, P., Reul, H., Sunkel, H. and Klein, I. (1970): Meth. Inform. Med., **9**, 35.
9. Paplanus, S. H., Shepard, R. H. and Zvargulis, J. E. (1969): Lab. Invest., **20**, 139.
10. Lacronique, M. (1971): Meth. Inform. Med., **10**, 30.
11. Gaynon, P. and Wong, R. L. (1972): *Ibid.*, **11**, 152.
12. O'Connor, M. L. and Smith, J. C. (1973): Amer. J. clin. Path., **59**, 357.
13. Platt, R. C., Wong, R. L., Lantner, K. W. and Gaynon, P. S. (1974): Comput. Biomed. Res., **7**, 83.
14. Hercz, L., Laszlo, C. A. and Reesal, M. (1974): *Medinfo 74*. Amsterdam: North-Holland Publishing Co.
15. Crocker, D. W. (1972): Hum. Path., **3**, 1.
16. Simon, M., Leeming, B. W., Bleich, H. L., Reiffen, B., Byrd, J., Blair, D. and Shimm, D. (1974): Radiology, **113**, 343.