J. clin. Path. **22,** suppl. (Coll. Path.), **3,** 82-86

Analysis of hospital bacteriological data

R. M. STIRLAND, VALERIE F. HILLIER, AND M. G. STEYGER

From the Departments of Bacteriology and Virology and Social and Preventive Medicine of the University of Manchester

Machine analysis and reporting has been widely applied to the numerical results of clinical biochemistry and haematology. Few have applied computers to routine bacteriology, although they have been of great value in taxonomy. This is unfortunate but stems, perhaps, from failure to appreciate that although reports may contain several different sorts of information, all of them may be given numerical values. This conversion, however, requires the bacteriologist to conform to a standard methodology to which some might object.

Another reason for this delay is probably that most of the available computer languages and compilers tend to be too inflexible for easy adaptation to the diversity of information which clinical bacteriologists may want to analyse. The present investigations were therefore undertaken to discover the suitability of a compiler, designed for survey analysis, for obtaining statistical and epidemiological information from the reports of a hospital bacteriology laboratory. Bacteriology is inseparable from epidemiology with its implied ability to make predictions derived from past experience. At its simplest a knowledge of the prevalence and drug sensitivity of specific pathogenic bacteria should increase the precision of clinical diagnosis and treatment. Within the laboratory, this information could also be used as a form of quality control by drawing attention to changes in reporting practice.

PROCEDURE

These experiments were conducted in the Manchester Royal Infirmary where a conventional reporting system is used. After issuing the report copy, the file cards were sorted into a small number of major categories to produce at least three visual alignment checks of punching error in each batch.

The method of coding has already been described (Report, 1968). This was deliberately compressed by using multiple punching, so that virtually all types of information could appear on a single punch card. This had certain advantages, including limitation of storage space, the need to teach operators only a single code, and allowing fixed-field files to be formed easily on magnetic tapes. Although this code may appear difficult, it was easily mastered by about 15 clerical assistants between July 1964 and December 1968, each taking four to eight weeks to achieve competence. A weekly output of about 600 reports took about 12 hours to sort, punch, and verify.

Cards were punched on a simple hand punch with manual verification. In certain experiments the rate of error was investigated by comparing the output of two or more operators and by using range checks on the card sorter and computer. These indicated that about 1% of cards contained any punching error. Most of these were unimportant, and less than 0·1% had errors which might affect the clinical interpretation of the report. This was less than that introduced by the technician writing the report itself. In addition, technicians' errors were likely to be 'logical' and hence less easily detected.

Computer processing was on the University of Manchester Atlas computer situated about $\frac{3}{8}$ mile (1 km) from the laboratory. For this, and other reasons, a batch input of about 2,000 cards was convenient. The following facilities were used: card input of raw data, program input on paper tape, magnetic tape storage, and line printout of tables.

Direct computation with a tabulated output from 2,000 cards was performed on a few occasions without the use of tape storage. The amount of computing time required for this, in competition with other users, however, made this procedure slow and unreliable. Much greater economy and speed was achieved by breaking down the processes into smaller units. Cards were read in off line to one magnetic tape and then transferred by a standard program to a second tape (Table I). This transfer involved the following computations. (1) Check ranges and print error report. (2) Convert primary image into an array of numbers. (3) Pack drug sensitivity results into smaller array. (4) Arrange reports in uniform files of

TABLE I

PROCEDURE FOR COMPUTERIZED ANALYSIS OF RECORDS

Procedure	*Programming*
1 Read in cards to magnetic tape	None (off-line)
2 Translate to second tape and file in 500s	Standard
3 Open files, categorize, tabulate, and print tables	Variable

500, occupying 71 blocks each of 512 computer words (a standard tape of 2,000 blocks could thus contain about 14,000 reports). (5) Print serial number of cards filed, and a 'password' for later access to the file.

Programming was performed within the compiler developed by Adelstein, Collard, Downham, McKay, Rutovitz, Stein, and Susser (1965). This compiler, an Atlas Autocode, was designed to give exceptional freedom of coding for input information on cards or paper tape. The freedom permitted on cards has been described (Report, 1968); this allowed great economy in space required to encode a given amount of information. For analysis, any of this information can be categorized as required with complete freedom to suit the needs of particular investigations. These may be for either specific or class retrieval, and the product of these searches can be printed out directly or stored on magnetic tape for future reference or modification. A program to consult the files for a simple tabulation requires only the addition of eight to 17 instructions to the general program. These would be of the following kind: variate specification instructions, table specification instructions, data source instructions, control, setting, and/or operating instructions, and storing and/or printing instructions. Requests like this to answer particular questions can usually be met quite easily.

The only significant addition required to the compiler was a set of special routines to assist in economy in storing the results of drug sensitivity tests.

Although this work has been primarily orientated to survey analyses, similar procedures and input can easily be used for the production of individual reports. These consist simply of the tabulated contents of single reports instead of cross-tabulation of many.

The main difficulties encountered were concerned with the bulk of information. After only four years we have enough cards to fill nearly 10 magnetic tapes, with the prospect of an almost endless accretion of new records. Disc storage may help in searching files of this size, but whatever their importance to the individual patient, about 90% of these records lose their individual importance in the laboratory within a few months. The problem is to define which 10% should be retained. We are,therefore, actively concerned with methods for creating cross-indexed subfiles for long-term retention.

FINDINGS

Initial investigations showed that three tables (Table II) gave useful information about the source and type of specimens sent to the laboratory, the proportion of them giving positive cultures, and the types of bacteria isolated. It was very quickly apparent that these values were remarkably constant (Tables III, IV, and V), and that deviations were usually of interest. We, therefore, prefer a system which files the full monthly statistics on magnetic tape and prints only those which differ from prescribed ranges.

Computer analysis can bring to light errors which

TABLE II

STANDARD MONTHLY ADMINISTRATIVE TABLES

Table No.				
1	Ward or unit		Number of specimens	
2	sending specimen	(24[1])	in major categories	(9)
3	Type of specimen	(23)	Culture result	(6)
4	Organism isolated	(39)	Type of specimen	(5)

[1]Number of categories used.

TABLE III

NUMBER OF REQUESTS FROM MAJOR USERS IN JANUARY IN THREE SUCCESSIVE YEARS

User	*Year*		
	1965	*1966*	*1967*
Inpatient Units			
M1	179	202	220
M2	94	92	114
M3	101	97	88
M4	136	93	112
M6	91	100	175[1]
S1	44	56	64
S2	132	100	116
S3	80	70	79
S4	130	170[2]	174
S5	108	115	176[3]
Respiratory	114[4]	82	51
Other	128	139	136
Private Patients' Home	109	131	128
Outpatients			
General	418	463	419
VD clinic	56	50	53
Casualty	172	138	41[5]
Other	94	102	126
Totals	2,186	2,200	2,272

[1]New consultant appointed
[2]New test introduced
[3]Survey of respiratory tract infections
[4]Chance variation in intensive investigation units
[5]Casualty officers stopped swabbing simple septic lesions.

TABLE IV

NUMBER OF MAIN CATEGORIES OF SPECIMENS RECEIVED IN JANUARY ON THREE SUCCESSIVE YEARS

Type or Source of Specimen	*Year*		
	1965	*1966*	*1967*
Blood	49	57	64
Cerebrospinal fluid	154	110	127
Throat and mouth	66	55	58
Nose and antrum	19	32	14
Sputum	380	331	425
Eye	2	7	1
Wounds, etc	308	362	199[1]
Urethra and vagina	88	75	89
Urine	959	1,088	1,141
Bile	5	3	1
Faeces	83	124	98
Cavity fluids	34	39	29
Other specimens	21	3	9
Totals	2,196	2,200	2,272

[1]Casualty officers stopped swabbing simple septic lesions.

TABLE V

NUMBER OF EACH BACTERIAL SPECIES IDENTIFIED IN JANUARY ON THREE SUCCESSIVE YEARS

Organism	*Year*		
	1965	*1966*	*1967*
Staphylococcus aureus	189	191	144[1]
Staphylococcus albus	86[2]	38	37
Strep. pyogenes (group A)	15	10	9
Strep. agalactiae (group B)	4	3	4
Streptococcus sp. (group C)	1	2	3
Streptococcus sp. (group G)	1	4	1
Streptococcus viridans	14	8	8
Streptococcus faecalis	25	28	25
Streptococcus, anaerobic	12	11	3
Pneumococcus	28	41	69[3]
Clostridium welchii	3	1	1
Diphtheroid	19[2]	6	3
Neisseria gonorrhoeae	12	1[4]	14
Pasteurella haemolytica	0	0	1
Pasteurella septica	2	0	0
Acinetobacter anitratus	9	4	1
Achromobacter sp.	4	2	5
Alcaligines sp.	3	1	1
Pseudomonas aeruginosa	88[5]	67	28
Haemophilus influenzae	54	60	155[3]
Bacteroides sp.	2	7	4
Escherichia coli	140	130	173
Klebsiella sp.	59	52	103[6]
Citrobacter	2	1	3
Providencia	5	0	5
Proteus mirabilis	100	91	85
Proteus morganii	6	41[6]	7
Proteus rettgeri	5	1	5
Proteus vulgaris	6	1	8
Salmonella sp.	0	6	1
Shigella sonnei	0	0	2
Candida albicans	19	22	12[7]
Candida sp. (not *albicans*)	11	10	37[7]
Totals	924	886	990

[1]Casualty officers stopped swabbing simple septic lesions.
[2]Too many irrelevant saprophytes reported
[3]High incidence of bronchopulmonary infections
[4]Batch of agar inhibitory to *N. gonorrhoeae*
[5]Outbreak of *Ps. aeruginosa* in Respiratory Care Unit
[6]Endemic *Pr. morganii* infections (1966) replaced by Klebsiella in Urological Unit
[7]Germ-tube test wrongly interpreted gave too low proportion of *Candida albicans*.

might otherwise pass unnoticed. An example of this is shown in Table VI which contrasts the reported resistance to sulphonamides in consecutive three-month periods. This difference was explained by investigating the interpretations put on the test results by the two technicians mostly involved. Of the two, it was found that technician B, with the lower incidence of resistance, was providing the correct interpretation.

Until 1964, all specimens of urine were cultured to detect significant bacteriuria which was present in 20% of specimens. A change to screening on the basis of clinical information and microscopy, to exclude from culture the 40% unlikely to give a positive culture, made no detectable difference to the previous isolation rate. Confidence in the reliability of this method was reinforced by analysis

TABLE VI

ISOLATIONS SHOWING SULPHONAMIDE RESISTANCE FROM URINARY TRACT INFECTIONS IN OUTPATIENTS (1967)

Organism	*Percentage Reported Resistant*		
	Jan-March (Technician A)	*Apr-June (Technician B)*	*Number Tested*
Escherichia coli	23	15	574
Klebsiella sp.	10	6	111
Proteus mirabilis	45	9	220

TABLE VII

APPROXIMATE INCIDENCE OF BACTERIA PER 10,000 ISOLATES AT MANCHESTER ROYAL INFIRMARY 1964-68

Species	*Incidence*
Achromobacter sp.	80
Acinetobacter anitratus	60
Actinomyces israelii	1
Aeromonas hydrophila	22
Alcaligines bronchiseptica	0·2
Alcaligines faecalis	14
Bacillus sp.	10
Bacteroides sp.	70
Bordetella pertussis	0·2
Brucella abortus and melitensis	1
Citrobacter	36
Clostridium welchii	20
Clostridium, other	1
Corynebacterium diphtheriae	0·2
Corynebacterium sp. (diphtheroid)	100
Escherichia coli	1,600
Escherichia coli (enteropathic)	5
Haemophilus influenzae	760
Klebsiella sp.	700
Lactobacillus sp.	1
Mima polymorpha	6
Moraxella lwoffii	6
Mycobacterium tuberculosis (human)	50
Mycobacterium tuberculosis (bovine)	2
Mycobacterium, other	1
Neisseria gonorrhoeae	120
Neisseria meningitidis	2
Neisseria sp.	10
Pasteurella haemolytica	8
Pasteurella septica	6
Propionibacterium	2
Proteus mirabilis	1,200
Proteus morganii	120
Proteus providenciae	15
Proteus rettgeri	40
Proteus vulgaris	40
Pseudomonas aeruginosa	450
Pseudomonas sp.	10
Salmonella sp.	14
Serratia marcescens	1
Shigella sonnei	22
Shigella, other	1
Staphylococcus albus	600
Staphylococcus aureus	2,360
Streptococcus pyogenes (group A)	120
Streptococcus agalactiae (group B)	60
Streptococcus sp. (group C)	35
Streptococcus sp. (group G)	40
Streptococcus, other haemolytic	10
Streptococcus sp. (*viridans*, etc)	150
Streptococcus faecalis	400
Streptococcus, anaerobic	95
Streptococcus pneumoniae	540
Veillonella sp.	0·4
Other bacteria	2

TABLE VIII

TETRACYCLINE-RESISTANT HAEMOLYTIC STREPTOCOCCI

Lancefield Group	*Percentage Tetracycline Resistant*		*Number Tested*
	1963-65	*1966-68*	
A	27	42	752
B	29	48	235
C	26	28	185
G	36	36	199

TABLE IX

BLOOD CULTURE ISOLATIONS FROM JUNE 1964 TO DECEMBER 1968[1]

Organism Isolated	*Significant*	*Contaminant*	*Total*
Staphylococcus albus	7	317	324
Staphylococcus aureus	113	16	129
Diphtheroids	0	114	114
Streptococcus viridans	48	10	58
Escherichia coli	45	0	45
Bacillus sp.	0	44	44
Klebsiella sp.	20	2	22
Proteus mirabilis	14	3	17
Pseudomonas aeruginosa	11	5	16
Achromobacter sp.	2	9	11
Pneumococcus	11	0	11
Streptococcus, anaerobic	6	0	6
Acinetobacter anitratus	4	1	5
Streptococcus faecalis	1	3	4
Streptococcus sp. (group G)	3	0	3
Aspergillus sp.	0	2	2
Bacteroides sp.	1	1	2
Brucella abortus	2	0	2
Candida albicans	2	0	2
Citrobacter	1	1	2
Moraxella wolffi	1	1	2
Salmonella dublin	2	0	2
Salmonella typhi	2	0	2
Salmonella typhimurium	2	0	2
Clostridium welchii	1	0	1
Brucella melitensis	1	0	1
Neisseria sp.	0	1	1
Penicillium sp.	0	1	1
Strepromyces sp.	0	1	1
Totals	300	532	832

[1]The 300 isolations believed to be of clinical importance came from 156 patients.

of the results obtained when general culture screening was reintroduced for trial periods. At the same time, screening for significant bacteriuria has continued to be of value in certain defined populations, but this more flexible routine permits an economy of about 10,000 culture plates a year.

While some information has local value only, some could serve as a useful basis for comparing the results from different laboratories. Thus Table VII shows the standardized frequency of identification of bacterial species in this laboratory. Extensions of this type of table, related to the type and source of specimens, could prove a useful basis for comparing and improving methods of identification, interpretation, and reporting.

Many results can be obtained by the sensible use of manual methods; Table VIII was obtained in this way. This was certainly quicker and cheaper than a computerized survey of six years' results stored randomly.

Long-term results requiring correlation with other information obtained outside the computer can still be assisted by it. Table IX gives the results from 4,200 blood cultures which required the retrospective interpretation of the significance of the positive cultures. In this series, 7·1% gave a significant isolate, while 12·7% were probably contaminated. Further analysis of this high proportion of contaminated cultures indicated that at least half was due to careless technique. Most of the staff taking these cultures produced a contamination rate of about 5%, but a minority had much higher rates. The introduction of *Bacillus* sp was invariably associated with the use of unsterile swabs for cleaning the skin, and Gram-negative contaminants appeared when skin disinfectants other than iodine were used.

Tables X and XI relate the incidence of particular types of drug-resistant *Staph. aureus* with information on the prescription of antibiotics from the Pharmaceutical Department. In these instances the final merger was done by hand although the compilation was assisted by machine. Another survey in which computer and manual methods have been merged is given in Table XII. The inclusion of

TABLE X

METHICILLIN-RESISTANT *Staph. aureus*

Year	*No. of Patients Infected*	*Penicillins*				
		Total (ka)	*Penicillin G*	*Ampicillin*	*Cloxacillin/Methicillin*	*Cephaloridine*
1962	2	11·5	4·2	3·6	3·7	—
1963	23	10·7	4·0	3·8	2·9	—
1964	49	12·5	3·0	7·6	1·8	0·1
1965	127	27·5	2·5	17·8	7·0	0·2
1966	178	39·8	2·6	29·4	7·4	0·4
1967	180	48·4	1·5	41·6	4·6	0·7
1968	241	43·8	1·1	33·5	7·8	1·4
Totals	800	194·2	18·9	137·3	35·2	2·8

TABLE XI

RELATION BETWEEN USE OF TRIPLE ANTIBIOTIC SPRAYS AND INCIDENCE OF PATIENTS INFECTED BY *Staph. aureus* RESISTANT TO NEOMYCIN AND BACITRACIN

Year	*No. of Sprays*	*No. of Patients Infected*
1956	1	0
1957	155	0
1958	219	0
1959	184	0
1960	169	0
1961	927	0
1962	1,500	45
1963	1,225	118
1964	149	88
1965	248	74
1966	436	66
1967	524	63
1968	318	49
Total	6,175	513

TABLE XII

FREQUENCY OF PRIMARY ISOLATION OF THREE COMMON MULTIPLE-RESISTANT STRAINS OF *Staph. aureus* FROM DIFFERENT SITES (1962 TO 1968)

Site of Primary Isolation	*Methicillin Resistant* (%)	*Phage Type 52/80* (%)	*Neomycin-Bacitracin-Resistant* (%)
Wound and skin	46·4	58·9	69·5
Respiratory tract	43·5	34·0	20·3
Urinary tract	8·3	6·4	8·8
Blood	1·4	0·6	0·6
Faeces	0·4	0·1	0·8
Total patients infected	800	506	513

'phage type' and 'primary isolation', which are not at present included in our computer survey, does not lessen the assistance to be gained from it in categorizing the other variables in the table. In general we have found it most profitable to proceed towards computerization of information by preliminary manual simulation and pilot trials in order to define our needs more precisely.

Perhaps the most important fact which we have re-discovered during these investigations is that when the normals have been defined, only the changes are interesting.

SUMMARY

Bacteriological reports from a hospital laboratory have been encoded on punch-cards and analysed on an Atlas computer using magnetic tape storage. Examples show some of the surveys made possible by the technique. Apart from the simpler epidemiological information which can be obtained, data processing may give valuable guidance on quality control in clinical bacteriology.

We should like to thank Mr D. Y. Downham for his assistance in the earlier stages of this work.

REFERENCES

Adelstein, A. M., Collard, P. J., Downham, D. Y., McKay, J., Rutovitz, D., Stein, Zena, and Susser, M. W. (1965). *Compiler for Survey Analysis: Guide to use of a Compiler consisting of Atlas Autocode and a number of inter-related A.A. routines for the Analysis of Surveys.* Department of Social and Preventive Medicine and Department of Bacteriology, University of Manchester.

Report (1968). Data Processing in Clinical Pathology. *J. clin. Path.*, 21, 231.