Thorax, 1978, **33**, 257–260

Comparison of reports on lung function tests made by chest physicians with those made by a simple computer program

D. M. GEDDES, M. GREEN, AND P. A. EMERSON

From Westminster Hospital, London SW1P 2AP

Geddes, D. M., Green, M., and Emerson, P. A. (1978). *Thorax*, **33**, 257–260. **Comparison of reports on lung function tests made by chest physicians with those made by a simple computer program.** The numerical results of 60 sets of pulmonary function tests were submitted to five consultant chest physicians for independent reporting. The chest physicians' reports were compared with the reports generated by the on-line computer reporting system in routine use in the pulmonary function laboratory at Westminster Hospital.

There was good agreement between the reports of the chest physicians among themselves and with the computer. The individual reports were compared with the consensus opinions of the physicians and the computer. The computer's decisions differed from the consensus opinion in 4·6% of instances. The comparable figures for the five physicians were respectively 2·9%, 2·9%, 2·9%, 4·0%, and 4·6%. Decisions differing from the consensus were due to mistakes or actual sustained disagreements of opinion. The physicians made more mistakes than the computer even though they were performing under test conditions. They made up for this, however, by producing fewer genuine disagreements from the consensus opinions.

It is concluded that in routine day-to-day practice the computer report will be as consistently useful as the chest physicians' reports and more immediately available.

Computer programs for the analysis of spirograms and for test reporting have been described for use with a distant computer (Schonfield *et al.*, 1964) and for on-line systems in the USA (Earle and Domize, 1969; Chiang, 1972). Pack *et al.* (1977) have described a computer system which is in use at the Glasgow Royal Infirmary for processing data from pulmonary function tests and giving an automatic interpretation. The data are processed in batches and are not on-line. An on-line computer system for the analysis of spirometry has been in use in the pulmonary function laboratory at Westminster Hospital for five years and for actual interpretation and reporting on the results for over a year. The details of the system will be described elsewhere.

Clinicians generally accept the validity of any data produced by such systems but are much more sceptical about the value of interpretative reports generated by machines. Published validations of computer reporting programs have been limited to a simple review of the computer-generated reports (Hoffer *et al.*, 1973; Ellis *et al.*, 1975). We therefore set out to compare the reports of our computer with those made by experienced chest physicians on the same sets of results and to see how the chest physicians agreed or disagreed among themselves.

Methods

Routine spirometry is carried out on an Ohio Electromed model 780 spirometer, and peak flows are recorded with a Wright peak flow meter. The subject performs a maximum inspiratory and expiratory manoeuvre from which the forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), maximum mid-inspiratory and mid-expiratory flow rates (MMIFR, MMEFR), FEV_1/FVC, and MMEFR/MMIFR are calculated by the computer and displayed on a small oscilloscope in the laboratory. The tests are repeated after inhalation of 100 μg of salbutamol aerosol in those patients with evidence of

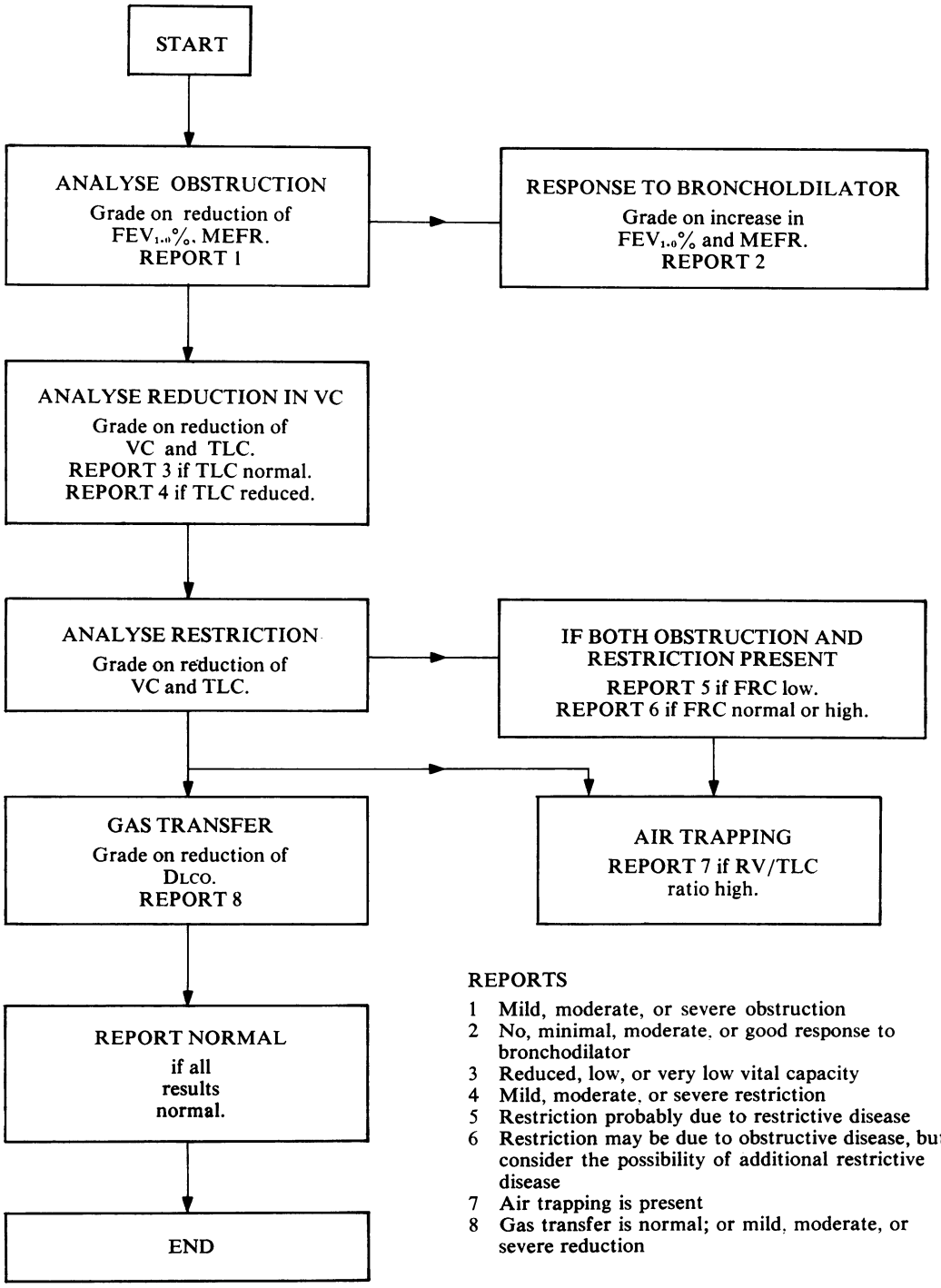

Figure *Simplified flow chart showing the main steps of the lung function test reporting program.*

airflow obstruction. At the clinician's request, lung volumes are also measured by helium dilution and carbon monoxide gas transfer by the single-breath method. The technician enters the results of these tests, and such patient data as age, sex, height, and weight, on a push-button keyboard in the laboratory. The computer then calculates predicted normal values, which are displayed alongside the spirometric data together with an interpretative report on the results.

The Figure summarises the main steps of the program and indicates the range of phrases from which the computer makes its selection for the report. The criteria by which the computer decides the degree of obstruction or restriction are described in the full program, which is available on request. In general, a reduction of between one and two standard deviations below predicted is described as minimal, between two and three as moderate, and more than three as severe.

Sixty consecutive sets of lung function test results generated in this way were reported independently by five consultant chest physicians, and the reports were compared with those produced by the computer. All 60 results contained spirometric measurements. Forty-seven included lung volumes and 35 gas transfer measurements. The physicians were informed of the terms in which the computer reported and were asked to keep their reports brief and to use, as far as possible, the same terminology as the computer. The reports were compared by giving a numerical value to the assessment of the severity of each abnormal pattern. Each report contained several decisions as to the pattern and severity of abnormality. For example, airways obstruction was scored from 0 to 4 : 0=no airways obstruction; 1=possible/minimal; 2=mild; 3=moderate; 4=severe. After all the reports had been completed, the 'correct' score for each pattern was taken as the average of the scores given by the computer and the five physicians. Any score differing from the average by two or more was regarded as differing significantly. Whenever a difference occurred the reporter was asked to report the relevant tests again. Differences were then classified as 'mistakes' when the second report no longer differed from the average, and 'disagreements' when the second report maintained an independent view.

Results

Tables 1 and 2 show the number of times the computer and each of the physicians gave a report that was different from the average. Differences are classified as 'mistakes' or 'disagreements', as explained above. A 'mistake' by the computer was due to an incorrect entry of data or incorrect copying of the report from the display on to the form. Overall agreement between physicians and the computer and among physicians was very close, and there was very little scatter in the opinions for any one assessment. The number of different decisions that had to be made by each of the physicians and by the computer was 175. On this basis it will be seen from Table 2 that the computer made three 'mistakes' in 175 decisions (1·7%) whereas the physicians made 24 'mistakes' in 875 decisions (2·7%), even though they were performing under test conditions. In day-to-day practice they might be expected to make more.

On the other hand, the computer produced five 'disagreements' in 175 decisions (2·8%) whereas the physicians had only six 'disagreements' in 875 decisions (0·7%).

Taking into account both the 'mistakes' and 'disagreements', the computer's decision differed from the consensus decision in 4·6% of instances. The physicians as a group performed marginally better; their individual decisions variously differed from the consensus in 2·9% to 4·6% of instances.

Table 1 *Differences between the individual reports and the average of all reports, for each functional abnormality*

	Obstruction	*Restriction*	*Reversibility*	DLCO	*Total*
Computer	1	4	3	0	8
Physician					
1	0	1	0	4	5
2	2	1	1	1	5
3	1	2	2	0	5
4	2	1	0	4	7
5	1	3	2	2	8
Total	7	12	8	11	38

Table 2 *Differences divided into 'mistakes' and disagreements (see text)*

	'Mistakes'	*Disagreements*	*Total*
Computer	3	5	8
Physician			
1	5	0	5
2	4	1	5
3	3	2	5
4	6	1	7
5	6	2	8
Total	27	11	38

Discussion

This study shows good agreement between the computer reports and those of the chest

physicians. When account is taken of the probability that the physicians might be expected to make more mistakes when working under the stress of day-to-day practice, it seems that the computer report is as consistently useful as the chest physicians' routine report. It has the added advantage of being immediately available.

The finding of such close agreement between all reports in this survey was perhaps surprising and emphasises how uniformly the descriptive terms such as obstruction, restrictive, and reversible are used. The 11 disagreements were assessed in detail. The most frequent problem was the assessment of a restrictive defect in the presence of airflow obstruction and air trapping. The incidence of 'mistakes' was similar for all reporters and is presumably due to misreading of figures or lack of concentration since the mistaken report could be explained in no other way. The computer 'mistakes' are important since there is a tendency to believe that numbers become infallible when there is a computer in operation. A number of steps in the processing of the computer data are still prone to human error, particularly when numbers are transcribed by the technician. The number of 'mistakes' produced by physicians was probably minimised since the knowledge that a study was in progress was likely to have stimulated more concentration than usual. We would expect more 'mistakes' to occur in routine reporting. The computer has the considerable advantage of consistency in the operation of a busy laboratory.

The idea of using a computer to report on lung function tests is anathema to some clinical physiologists. Apart from the understandable distress of seeing a hard learned human skill readily performed by a machine, two objections worthy of consideration have been voiced. The first is that lung function tests should be interpreted as an extension of the clinical examination and lose much of their value when reported in isolation. This may be so if clinical diagnosis is carried out as a subjective art but, if clinical diagnosis is to be made on a logical basis the result of each test should first be stated objectively and in isolation, and only then applied to the other information already available.

The second objection is that information may be lost if the figures are not actually interpreted by the clinician. The degree to which this may happen clearly depends on the quality of the program and the interpretative skill of the clinician. An expert clinical physiologist will clearly get more information from the actual figures than from a report derived from a poor program. A good computer report produced immediately is, however, very helpful to clinicians who may be less skilled in interpreting respiratory physiology.

The descriptive terms used in our reporting are the standard ones but they do describe arbitrary patterns, and there is a view that the ideas of restrictive and obstructive ventilatory defect should be abandoned altogether. However, no satisfactory alternative system has been forthcoming. The descriptive terms used are widely taught, and this study shows that in practice they are used surprisingly uniformly. For this reason we conclude that they are a useful means of summarising and communicating lung function data.

We are most grateful to Dr L. Capel, Professor T. Clark, Dr G. Laszlo, and Dr K. Saunders for taking part in the reporting game. Mr A. Bovington carried out the actual lung function tests. We also acknowledge the help of Mr P. E. Dann, Mr J. R. Hunt, and Dr T. D. Preston, of the Medical Computer Centre, Westminster Hospital, London.

References

Chiang, S. T. (1972). Application of desk-top computer in clinical spirometry, *American Review of Respiratory Disease*, **106**, 614–617.

Earle, R. H., and Domize, D. (1969). On-line computer analysis of pulmonary function tests: The forced expiratory spirogram. *Chest*, **56**, 247–248.

Ellis, J. H., Perera, S. P., and Levin, D. C. (1975). A computer programme for calculation and interpretation of pulmonary function studies. *Chest*, **68**, 209–213.

Hoffer, E. P., Kanarek, D., Kazemi, H., and Burnetton, G. O. (1973). Computer interpretation of ventilatory studies. *Computers and Biomedical Research*, **6**, 347–354.

Pack, A. I., McCusker, R., and Moran, F. (1977). A computer system for processing data from routine pulmonary function tests. *Thorax*, **32**, 333–341.

Schonfield, E. M., Kerekes, J., Rademacher, C. A., Weihrer, A. L., Abraham, S., Silver, H., and Caceres, C. A. (1964). Methodology for computer measurement of pulmonary function curves. *Diseases of the Chest*, **46**, 427–434.

Requests for reprints to: Dr P. A. Emerson, Westminster Hospital, Dean Ryle Street, Horseferry Road, London SW1P 2AP.