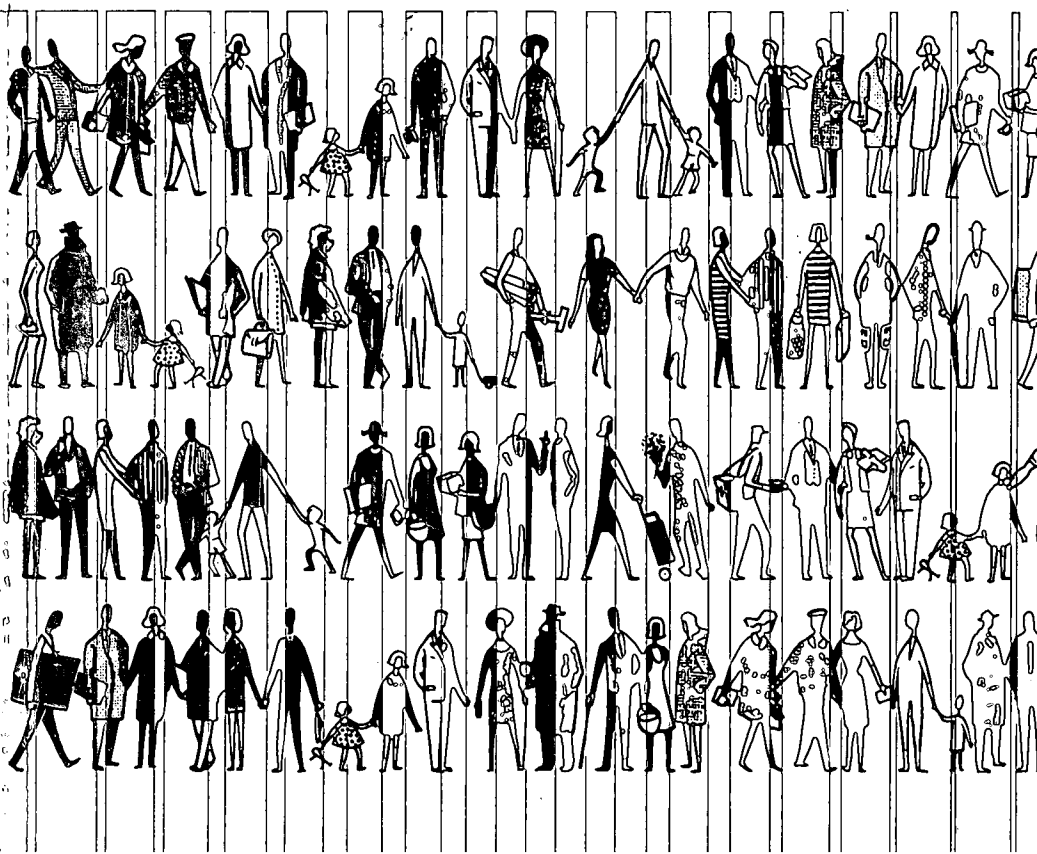
THE FUTURE—AND PRESENT INDICATIVES

PROBLEMS AND PROGRESS IN MEDICAL CARE
ESSAYS ON CURRENT RESEARCH NINTH SERIES

PUBLISHED FOR THE NUFFIELD PROVINCIAL HOSPITALS TRUST
BY THE OXFORD UNIVERSITY PRESS



THE FUTURE—AND PRESENT INDICATIVES
PROBLEMS AND PROGRESS IN MEDICAL CARE
ESSAYS ON CURRENT RESEARCH NINTH SERIES
OXFORD

The future—and present indicatives

Problems and Progress in Medical Care
NINTH SERIES
ESSAYS IN CURRENT RESEARCH
EDITED BY GORDON McLACHLAN

The future—and present indicatives

Published for the
Nuffield Provincial Hospitals Trust
by the Oxford University Press
London New York Toronto
1973

Oxford University Press, Ely House, London W1

GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
CAPE TOWN SALISBURY IBADAN NAIROBI DAR ES SALAAM LUSAKA ADDIS ABABA
BOMBAY CALCUTTA MADRAS KARACHI LAHORE DACCA
KUALA LUMPUR SINGAPORE HONG KONG TOKYO

ISBN 0 19 721380 4

Designed by Bernard Crossland

Printed and bound in Great Britain by
Hazell, Watson & Viney Ltd
Aylesbury, Bucks

Contents

Introduction

It is hazardous to forecast the course of history, but one certainty is that the radical re-formation of the administrative arrangements of the National Health Service taking place in 1973/4 will be of a somewhat different character and therefore have different stresses from that of 1947/8. Then, regionalization and the grouping of services for more effective operation were not exactly novel ideas but they were certainly untried in practice. In the hectic months which straddled the appointed day for the introduction of the new National Health Service, the immediate emphasis was on the formation of new-style administrative units at regional and local level, which on 5 July 1948 assumed responsibility for the hospital and specialist, and the general medical services; and this initiated the operation of health services in the UK on a national scale.

Now after years of experience of a slow but steady policy of internal rationalization in the several parts, designed principally to redistribute resources more equitably, the period has ended with the realization that in the next stage there must now be administrative integration of primary, secondary, and the medical part of community services if the concept of medical care as a single entity is to have meaning in practice. The moving forward of time has, however, brought about a considerable maturity in the planning process. The grouping of hospitals into distinct administrative entities inevitably carried in its train a sharpening appreciation on the part of the authorities of the constituents of need, and the strong contrast between need and demand. Operational experience has also drawn

attention to the imbalance in resources deployed between hospital-based and community care, and to the desirability of a more positive role for medical practice in health education and prevention.

It is indeed possible to see that the first twenty-five years of the NHS has contributed towards a better understanding on the part of most concerned, the public as well as the professions, of the realities and complexities of all that is involved in the setting up of the large-scale enterprise which is the NHS, and a special feature of the end of this era is the emphasis on establishing priorities to make better use of resources for services designed to meet what is being perceived, albeit dimly, as the greatest needs. The days are therefore gone when it could be assumed that resources could be summoned up to meet demand. It is being accepted that authorities have to choose between competing claims, and this calls for a revolution in thinking about the areas for exploration.

It seems clear that within obvious limits, the next stage of rationalization, with the merging of most if not all the services concerned with illness, must allow for rather more fluidity in the direction of resources. But straight away there is a major problem not so simple to grasp as those of the 1940s, 1950s, and 1960s which generated demands for capital investment in buildings, equipment, and manpower. This is the assessment of the quality of medical care in practice, and its bearing on the deployment of resources clearly inadequate to meet all the demands made with varying degrees of sophistication, from every direction of the health compass. To obtain sensible answers to this question focuses attention on the need for better information.

This is the reason for the publication of this collection of essays on the eve of the new NHS as an indication of some vital if not so obvious issues relevant to the necessary continuous review of medical care. If at first glance they appear to go somewhat beyond the immediate problems likely to be facing the managers appointed for the new venture in 1974 they are being no means esoteric. It is also perhaps appropriate to observe that the style of such managers and the way they address themselves to the problems of medical care in the first years of the reformed NHS will do much to mould the nature and direction of medical care in the future.

The importance of Sir George Godber's introductory essay needs no stressing, and the brilliance and comprehensiveness of the line of attack needs no bush. It is both an inspired piece of writing and a pointer to a way of thinking about dimensions that go beyond the

ordinary day-to-day anxieties of administration. It is a reminder too not only of the astonishing eclecticism of Sir George's mind but of the greatness which has characterized his tenure of the office of Chief Medical Officer.

At the same time, the kind of thinking exemplified by the essays from Professors Knox and Ashford will have increasingly to be made, tentatively at first no doubt, by management using the skills of the multidisciplinary teams which are to be a feature of the reformed health service. It is of course only one part of the problem to provide the mechanism and means. Mr Yates demonstrates how the problem of not only translating data to information, opens into the other key question of how it is to be used. It is evident that to answer this satisfactorily will need a special educational effort in itself. But there are also immediate problems concerned with patient management and Dr Rhys Hearn's sharp delineation of the problems associated with the management of the front-line troops of acute medical care, the nursing force, provides a thoughtful comment on the nurse dependency factor of care.

Yet few are confident at the moment, that the base for the production of the information that will be required for a positive medical care service is adequate. Professor Alderson addresses himself to this question and deals not only with the function of the community physician in the new NHS, but also with the kind of tools that require to be forged. It is based on the experience of a unit that uniquely in England has twin foundations in the academic and service worlds. Indeed it is not hard to conclude that a well-defined strategy for the setting up of adequate information systems is urgently needed, if the new NHS is to fulfil its objectives.

GORDON MCLACHLAN

June 1973

3 Prince Albert Road
London NW1

1

The cosmological medical science?

Future developments in epidemiology

Sir George Godber
GCB, DM, FRCP
Chief Medical Officer,
DHSS

Future developments in epidemiology

Epidemiology first developed as the study of epidemics and crowd diseases in the terminology of Major Greenwood's book. It still has an important function in the field in dealing with the control of communicable disease and some other acute episodes of illness as, for instance, the Minimata Bay incident of mercury poisoning or the two outbreaks of poisoning with an insecticide accidentally contaminating flour and bread in Britain in the 1960s. Substantial outbreaks of communicable disease like the Aberdeen typhoid outbreak or the North Yorkshire paratyphoid outbreak in which traditional field epidemiology is necessary, still occur. Because the incidence of many of the communicable diseases has now been greatly reduced few medical officers of health have personal experience in the control of major outbreaks. The Public Health Laboratory Service is able to help to an extent that some of the best-known field epidemiologists of the first half of the century, such as Vernon Shaw, did not enjoy, but even they, with their national and regional system of laboratories, have too few people with field experience. Even the Department of Health and Social Security now deploys a much reduced number of doctors with special experience and is much more concerned with monitoring longer-term trends and developing immunization campaigns. Medical officers of health are now beginning to recognize that they need more support on a regional basis than they have had in the past and this will probably have to be based upon the Public Health Laboratory Service. It is one of the improvements that could follow the establishment of an all-purpose regional health authority.

4 *Future developments in epidemiology*

It is relatively simple to define what the requirements may be for the study of the epidemiology of communicable disease and for field support, and beyond recognizing the need for improvement of this kind this symposium will probably not want to pursue that side of the question further. It is worth noting, however, that it was only just before the Aberdeen typhoid outbreak, the most spectacular example of an epidemic derived from an infected tin of corned beef, that it was realized in this country that such an incident was not just a bizarre occurrence in a small canning plant unlikely to be repeated, but a real if small risk of any canning plant which did not use pure water for cooling. The North Yorkshire epidemic of paratyphoid fever last year was the third in this country in which cattle had actually been found to be infected with *Salmonella paratyphi* and to excrete the organisms in the milk. Both these sources of enteric infections are either not accepted by many people in other countries or not recognized as being such serious possibilities as we now know them to be.

Field epidemiology of a somewhat similar kind is needed occasionally for the investigation of outbreaks of chemical poisoning. In addition to the examples quoted from Britain and Japan, there have been the much larger episodes of Jake palsy in South Africa and the paralytic conditions as a result of poisoning through contaminated cooking fat in North Africa a few years ago. The small episode of cardiomyopathy believed to be due to the use of a cobalt-containing foaming agent in beer in Canada is of the same kind, as is the occurrence of Pink disease in babies given frequent treatment with a mercury-containing 'teething powder' in this country until some fifteen years ago. There have been many other occasions when an acute episode of chemical poisoning has been due to the use of unsatisfactory vessels for food, including poisoning with antimony and zinc. There have been two outbreaks within the last twenty years of acute lead poisoning in young children as a result of exposure either to lead-containing fumes or dust, following the burning of old battery cases as domestic fuel. These episodes need the same kind of field epidemiology but backed by additional expertise in chemical poisoning. The toxicologist is likely to be increasingly important to the exponent of preventive medicine because of the other episodes which may arise from exposure at work or at home through toxic chemicals such as DNOC or the organo-phosphorus insecticides in the past. Even the occupational exposure of nurses to sensitization to pencillin or the question of adverse effects of enzyme detergents need some-

thing of the same field epidemiological approach. We clearly need to have toxicological advice available on a far wider scale than we now possess for this kind of epidemiological investigation.

The work of the Adverse Reactions Committee of the old Committee on Safety of Drugs and now of the Medicines Commission is another kind of epidemiological study, but it does not involve the traditional practitioner of public health in the same way. Assuming that the ordinary safety investigations and clinical trial would have made a sufficient assessment of the short-term hazards of a new drug, there must remain some uncertainty about its future until it has been used in such numbers of patients as to indicate the extent of risk arising from personal idiosyncrasy of an unpredictable kind. One of the earlier known episodes of major damage from a drug formulation arose from the use of ethylene glycol in a preparation of sulphanilamide sold in the late 1930s when the excipient, not the drug, was the poison. The work of the American haematologists in collecting information about adverse effects of drugs on haemopoietic mechanisms, including for instance chloramphenicol, was the forerunner of the present monitoring methods. The Adverse Reactions Committee's best-known investigations concerned the oral contraceptives and they gave us accurate information about the finite but very small risk long before anything of a similar kind was available in other countries and later, in collaboration with Denmark and Sweden, demonstrated how even this tiny risk could be halved by using a different formulation.

The very different incident of sharply increased mortality from asthma over four years following the use of a pressure aerosol containing isoprenaline involved far more deaths before it was detected. That episode serves to show that the bizarre clinical incident of massive thrombosis in a particular group may be noted much more readily than even a substantial change in a relatively common condition. If phocomelia had not been such a bizarre and uncommon malformation, one wonders whether the thalidomide tragedy would have been detected at all. If there had been, say, a comparable numerical increase in Down's syndrome, would it have been noticed?

It is axiomatic that drugs with powerful effects for good are likely to have narrow margins of safety and to need closer control in use, but they are also more likely than drugs with less profound physiological and pharmacological effects to be capable of serious harm. We are certain to get a steady flow of new substances for both therapeutic

and diagnostic use and to need general reporting and careful analysis of known or suspected adverse effects. Not only do we need this nationally but we can get much quicker results if information available from reliable international sources can be pooled. Not only can we gain in this country, but other countries can benefit from the findings of those with sophisticated monitoring systems. The WHO unit for drug monitoring has only recently become operational and cannot be said to have contributed a great deal yet, but it is certainly necessary for the future. In our case it is supplemented by a great deal of two-way exchange with other national health departments and especially with the Food and Drug Administration of the United States and the Food and Drug Department in Canada. Sir Derrick Dunlop was recently awarded the Bisset Hawkins Medal of the Royal College of Physicians for his major contribution to preventive medicine by the establishment of the Safety of Drugs Committee's work. That probably was one of the biggest contributions to preventive medicine in this country in the 1960s.

It is not possible to have an exactly similar monitoring of food and the effects of additives or processing, but we do need to collect information from laboratory sources as well as medical and other reports and to have more facilities than we now possess for toxicological evaluation. The general principle of rejecting the addition of anything merely designed to improve the presentation of food, unless we can have reasonable certainty by laboratory investigation that it can have no adverse effect, is essential. The Committee on Medical Aspects of Food Policy had a pharmacological subcommittee, which plays a considerable part in this, and also a panel on carcinogenesis, and these are now replaced by corresponding subcommittees under the new Committee on Medical Aspects of Chemicals in Food and the Environment. Essentially this form of epidemiology has to be practised centrally and it must be admitted that the epidemiological aspects have not been so well studied hitherto as those which are essentially laboratory based. Laboratory findings in animals cannot be readily extrapolated to man and even the mega-mouse facility to be established at Pine Bluffs in the USA will only provide indirect evidence on a lot of these points. The difficult of this work was well demonstrated by the long argument about the cyclamates used as sweetening agents, and there are still those, in the industry at least, who express doubt about the wisdom of the withdrawal of cyclamates. If we had not had an alternative in saccharin, the decision not

to use cyclamates after they had been shown to produce bladder tumours in rats and to be metabolized to cyclohexylamine in a proportion of humans, would have been much more difficult. After all a very small risk might have to be accepted in order to obtain a greater advantage in cutting down the excessive use of sugar in the diet of diabetics or the obese. That is a decision that may have to be taken afresh now that some evidence of minute risks with saccharin has been put forward; but it is not a decision that should be forced upon us by the Delaney Clause which controls the US action.

Some of the nitrosamines are actively carcinogenic; nitrites, and nitrates used in curing of food may be capable of increasing the formation of nitrosamines during cooking or even digestion. They may interact with such drugs as terramycin with the same result. We can now identify and measure nitrosamines in a few parts per billion concentrations, but we do not know whether this means anything to health.

I suggest that problems of the relatively acute effects of drugs or food additives or similar substances present in food or elsewhere in the environment are much more readily manageable than those of the very long-term effects of small amounts of many substances which may be present in air, water, food, or soil. We have very little information about the part that really long-term exposure to some potentially harmful substance may play in the aetiology of chronic and degenerative or malignant disease. We are completely at the mercy of well-motivated, but perhaps not always broadly informed, scientists who may suddenly announce that concentrations of this or that may be dangerous, for instance, lead in the atmosphere, or even that deficiencies of some substances, for instance calcium in the diet, may produce long-term effects on the incidence of bone disease or fractures in old people. What they say may be right or just plausible and its disproof may be almost impossible. The fact is that the one universal, determant incident for all of us is being born because thereafter our death becomes inevitable. Just how are we to estimate whether any factors in our subsequent condition of life hasten or delay that entirely predictable outcome?

The development of chronic and degenerative disease with advancing age is something that we all accept as inevitable. We also accept that the incidence of malignant disease inevitably increases with advancing age. What we do not know is the extent to which external factors may influence the rate at which such developments arise. In

countries with more primitive living conditions and in this country when adverse factors were so much more prevalent, the chance of dying before the degenerative processes were far advanced was so great that their postponement was hardly a matter of serious concern. Now that so many of the acute episodes in earlier and middle life are controllable, the factors which determine the onset of degenerative and malignant conditions become much more important. One hundred and thirty years ago the expectation of life for men in England and Wales at age 40 was 26·6 years and for females 27·7; in Sweden the comparable figures were 24·3 and 27·2 years. In England and Wales by the 1960s the figures were 31·6 years for men and 37 for women; but in Sweden they were 34·7 for men and 37·8 for women and these differences are too great to be due simply to smoking, though smoking may play some part at least in men.

The smoking story is a good example of the sort of problem which faces us. The cigarette is the noxious agent which began to be used on a substantial scale at the beginning of this century and only achieved a high level of consumption during and after the First World War. In a way we were lucky that the marker of its effect was an otherwise uncommon cancer, yet mortality from lung cancer had risen markedly over a twenty-year period before the relationship was proved. Looking back it is possible to see that the epidemic began forty years ago at least and yet there were still leading medical figures who rejected the epidemiological evidence well into the 1950s. To them statistical evidence was not real; it had not the obvious quality of, say, a head injury or a bacterial infection for which a blood culture, an infected vehicle, and a source were demonstrable.

Once the cigarette smoke/lung cancer relationship was shown, the parallel epidemiology of chronic bronchitis was relatively easy, although there are other causal factors which were diminishing just as the full impact of smoking began to be seen. The link with cardiovascular disease was less easy but more important in the context of this symposium because it is an example of accelerated onset of a degenerative condition which increases with age. Without the marker of lung cancer, would it have been demonstrated at all? Yet atherosclerosis is almost the fundamental lesion in the process of ageing and it is produced at a distance from the broncho-pulmonary epithelium upon which the primary insult of inhaled cigarette smoke falls. So too are the other smaller effects (cancers of bladder, pancreas, and liver, peripheral vascular disease, or even amblyopia) observed at

a distance. So there must be both surface and absorbed agents. There can be endless argument about the mortality attributable to cigarette smoking, but it is reasonable to conclude that at least a tenth of deaths in Britain are hastened by smoking and the fraction may be much larger. Without the cigarette component in our mortality, life expectation in Britain would be almost as long as in Sweden.

Epidemiological identification of long-term effects is only likely to be made if some distinctive occurrence leads us to them; acting as a marker in the same way as lung cancer. But the marker does not have to be so large as that, nor so strikingly unusual as phocomelia. A lot of cancers have been identified as the result of unusual exposures: like the luminizer's cancer which was local at the site of exposure, unusual in its nature and fairly early in onset. Skin cancers in tar-workers, mule-spinners, or chimney-sweeps' cancers were unusual, localized to the area of insult and found in a particular occupational group. Mesothelioma in asbestos workers is atypical, but still took a long time to surface. Cancer of the urinary tract after exposure to naphthylamine was identified late but in its extreme form points very clearly to an excreted carcinogen. Acheson and his colleagues were able to identify nasopharyngeal cancers very long after exposure and in small numbers because they were unusual and concentrated in small exposed groups. The time lag before the effect illustrates another problem in the epidemiology. We may have been already exposed to carcinogens which no one suspects but which may yet have widespread effects to come.

Our studies almost always assume that there is a baseline incidence of naturally occurring cancers. Dr Higginson's suggestion that 80 per cent or more of all cancers are due to extraneous factors challenges that. If we look for general evidence there is plenty of it. The present secular trend of incidence and mortality in cervical cancer suggests strongly that some adverse effect operating thirty or forty years ago was at least less effective twenty years ago so that mortality is falling at ages 35–45 independently of anything screening or therapy has done. On the other hand, liver and breast cancer death-rates have varied slightly without known explanation. There is a marked social class variation in mortality from cervical cancer for no explained reason and it is there at all parities. There are many examples of wide variations in incidence of and mortality from cancers of the same site in different countries and different parts of the same country. The well-known example of Burkitt's lymphoma may be explained by the

distribution of an infective agent, but there are many small differences in baseline incidence that may have an explanation in some other factor which might be accessible to contrived change.

Because the factors involved operate slowly over a very long time and affect whole populations, their identification will certainly require better epidemiological methods than we have yet devised. It will not always be a matter of identifying an external factor; there may be protective factors operating variably too. It seems to me necessary first to accept the proposition that there are reasons for the present incidence and for modest variations which may occur in it. We can then look for the clue from an unusual clustering of cases which may tell us more.

But there are some obvious problems which just do not get examined. The epidemiology of the fracture of the femoral neck has not been examined in Britain at all, yet Nordin has asserted that the incidence is to be related to decalcification, the result of insufficient vitamin D in the diets of old people. He has no clear evidence, but neither has anyone else, although virtually all liquid milk in the USA is *fortified with vitamin D*, and that might have been expected to affect their experience if the thesis is right. There is no difference between the countries so far as one can tell. The other even greater problem of atherosclerosis both produces more clues and a prospect of earlier results. Behaviour can play a large part in this if exercise and diet are factors. Morris's work on activity and the vast amount of evidence on the effect of diet give some prospect of modifying incidence if we can modify behaviour. True, the evidence about unsaturated fats in the diet is extremely confused, but at least overweight and cigarette smoking are clearly contributory factors. But the evidence of effects from various metallic substances such as lead, vanadium, cadmium, cobalt, or chemicals like carbon bisulphide and the phenomenon of raised mortality in areas with soft drinking water all suggest that the environment may be variable in a way which could be beneficial.

We have currently a problem concerning the monoenoic acids in fish and rape-seed oil. They may be present in varying amounts in margarine. A diet with a proportion of its calorie content derived from such fatty acids well above the level in our own diet does lead to myocardial changes in rats. Do smaller amounts over longer periods affect men?

Except for the soft water theory, most of our evidence is of short-

term effects, perhaps because exposure to very small amounts has to be very long-term and may then produce only some modest variation of the rate of development of coronary artery disease or atherosclerosis with age which is reflected in a raised or lowered mortality at an earlier age than usual. We may at this moment be reassured about the effects of mercury in greater than normal amounts in fish from certain areas. We do not know whether the levels we now encounter may in the very long term accelerate some degenerative process. Part of our problem with some of the assertions about, for instance, harm from atmospheric lead, is that although we may fairly reject the evidence that they are right we cannot prove they are wrong.

The biological experiment in man takes such an unconscionable time and must not be contrived. Except where the natural experiment has been done for us (as with soft water or, in another context, naturally fluoridated water) we have to extrapolate from short-term experience in man or high dosage in animals. Our new starting point has to be that natural variations between areas or whole population trends have to have an explanation for which we must now seek.

We are only just becoming conscious of the presence of some potentially toxic substances in trace amounts in air, water, food, and soil, although there has been enough concern about artificially added more complex chemicals used as pesticides or for some presentational effect. But our concern hitherto has been with the relatively short-term effects. We must now consider decades or even lifetimes of exposure. The cyclamates were an example of that requirement, and brominated vegetable oils another, but these are known additives. The far more difficult problem of increases or decreases due to natural or artificial factors in concentrations of substances which often occur naturally may have a greater long-term effect on the health or longevity of populations. To this must be added the balance of advantage and disadvantage if an occasional adverse effect from idiosyncrasy has to be measured against a small but general gain. This last may be a problem of drugs or food additives mainly but it can occur even with a detergent.

There is another area, which might give earlier results because events move faster, in antenatal paediatrics. The adverse effects on the foetus of rubella in pregnancy are well known and we can now prevent them. Rhesus sensitization of the mother and therefore haemolytic disease of the newborn can be prevented. Congenital syphilis is entirely preventable. But we still have a high incidence of toxaemia of

pregnancy, of prematurity, of various congenital anomalies and of such conditions as respiratory distress in the newborn. A rather clumsy reference in my *Annual Report* to a curious coincidence of raised neonatal mortality four months following severe influenza outbreaks in the only two years of the last twenty-five when infant mortality rose, led to press comment that could frighten mothers instead of to medical interest in a possible connection. There may be nothing in it or a chance association of marked severity with early onset of the epidemic may have produced an effect which shows in the statistics because it all comes in one-quarter. Sweden has markedly lower perinatal mortality than Britain and there must be a reason. Of course our old enemy, the cigarette, features here too: Scott Russell said that one abortion in five in women who smoke in pregnancy was due to smoking, but that is only part of the story. We have been able to demonstrate the selective action of rubella virus on certain tissues at one stage of foetal development, but a lot of other less specific effects could be present unidentified because there is not a marker like congenital cataract or deafness or phocomelia.

If new epidemiological methods or the extension of old ones are necessary for a study of the very long-term effects of environmental or other factors in the genesis of chronic and degenerative conditions and malignant disease, they are equally necessary for investigation of the outcome of known incidents such as communicable infections in earlier life. We are accustomed to regard the acute infective episode as complete in itself, leaving sometimes definable sequelae-like paralysis following poliomyelitis or a fibrotic pulmonary scar following tuberculosis, but very little attempt has been made to study the late and sometimes cumulative effects of acute episodes in earlier life. Degenerative joint changes following repeated trauma are well enough known, as is the effect of pregnancy on the incidence of varicose veins, but little attempt has been made to check on the part which may be played by such conditions as whooping cough and measles in determining later chronic respiratory disabilities.

A possible correlation between over-crowding, earlier occurrence of communicable diseases, more (and more serious) episodes of lower respiratory infections and the known social gradient in chronic bronchitis has been suggested, but not proved. There has been some work on the late results of exposure to atmospheric pollution in promoting the development of chronic bronchitis. Chronic bronchitis does not affect everyone living in an area subject to atmospheric pol-

lution, though of course it is much more likely to affect cigarette smokers than others, but even the non-smokers have a higher incidence in such areas than non-smokers in rural areas. There must be either some other predisposing incident or genetic factor involved.

The point becomes important when one considers the justification for immunizing against measles with a very safe vaccine or whooping cough with a somewhat more hazardous one. Many doctors even are prepared to suggest that natural measles is a relatively mild disease, the prevention of which does not warrant using a vaccine with a risk of serious complication that may be of the order of one in half a million. The fact that mortality from natural measles is about eighty times this and that before vaccination some 8,000 children might be ill enough to need hospitalization as a result of measles is forgotten. In the course of a measles epidemic in winter, any given GP may see perhaps about forty cases. As not more than one in a hundred will require hospital admission he sees such a case perhaps twice in five years. His chance of seeing a fatal case in general practice is about one in twenty-five years. If he relies on his own personal experience for his judgement on such a matter as this, he may well reach the wrong conclusion.

This particular epidemiological problem is not simply a matter of injuries and acute infections, but we do need to know more clearly what the cumulative effect of the insults of a lifetime may be. We will only get this from record linkage on a much more sophisticated scale than the present GP record provides. I am not advocating a tremendous new exercise in phenomenology, but a serious study on lines made possible by record linkage and computer storage so that our successors may truly study an individual's health rather than intermittent episodes of disease or injury.

Physical disease has so far largely monopolized most doctors' concern and its study is easier just because its episodes and signs can more easily be objectively defined and recorded than can those of mental illness. We are at the beginnings of such studies in mental illness through registers such as those of Camberwell and Aberdeen, and there are currently investigations on misuse of drugs which are nearer to those attempted with communicable disease. Schizophrenia has perhaps provided the most obvious target but the epidemiology of depressive episodes may be much more important.

The epidemiology of the use or misuse of psychotropic drugs in medicine is really part of a much-needed study of our misuse of any

other potent drugs, some of which are certainly grossly wasted and some may be used in a way which makes them either potentially harmful or less than effective, because they are used insufficiently.

I suppose everyone here fully accepts the need for better planned investigation of the effects of treatment. The controlled clinical trial, difficult as it may be to organize, is essential to the proper assessment of any new therapeutic method, medical or surgical. There are some differences in results that are so obvious that a controlled trial is hardly necessary. In the early days of the maternal deaths inquiry it was observed that practically all the deaths from retained placenta occurred in women who had been sent from home to hospital with the placenta *in situ* and a simple switch to the use of the flying squad, evacuation of the placenta in the home and transfusion instituted there, virtually eliminated that cause of maternal death.

Similarly, on a much larger scale, it was impossible to hold back the use of L-dopa in the treatment of Parkinsonism when large supplies were available before the results of treatment had been fully assessed. The symptomatic relief was such that neither the profession nor the public would have brooked delay. The recent controlled studies of the value of tolbutamide and phenformin in the treatment of diabetes in the USA not only cast doubt on the therapy but also suggest that atherosclerosis may be promoted. Yet both drugs have been widely used for many years. It has been extremely difficult also to make a proper assessment of the value of cervical smears and, as a result, many young women whose positive smears indicated a condition which may have been more likely to stabilize than to go on to invasive cancer of the cervix, have been subjected at the least to anxiety, at the worst to treatment which might sometimes affect their prospects of child-bearing or even occasionally endanger life. We simply do not know the balance between gain and loss yet, but there is a price to be paid for wholesale treatment and we cannot be sure of the benefit. Perhaps the most obvious example of the need for a controlled clinical trial is also one of the most difficult for action. About 200,000 children a year were undergoing tonsillectomy ten years ago. In 1969 the number was 130,000 although in that year the children in the relevant age-group reached the highest figure for the 1960s. We have no idea whether the whole or only half or even less of the 130,000 operations were justified, nor whether the other 70,000 that might have taken place on the criteria adopted half a dozen years earlier would have been helpful or harmful to the children affected. We have

had some controlled comparisons of efforts of treating varicose veins and we can say that the injection-compression method offers at least no worse results and costs perhaps a tenth of the amount that hospital admission with operation would cost. I suggest that there are really two levels on which we need epidemiological investigation here. A carefully planned controlled trial such as that of different methods of home or hospital treatment of acute myocardial infarction published in 1971 would provide the basis for a fresh study of local organization and methods. At the district level, meaning the district served by a single district general hospital, we certainly need a community physician familiar with epidemiological methods and findings, able to review what is happening in the population of his district in morbidity and mortality and in the provision of and results of treatment and prophylaxis. It may be journeyman stuff but on it depends the general improvement of the quality of medical care.

This paper has attempted only to outline some possibilities for future development. Some areas have received a good deal more attention than others, mainly because the problems set by environmental factors are both among the most difficult and those most likely to excite public attention. The public did not notice 3,500 extra deaths from asthma in four years, although they were reported, but it becomes desperately concerned about the possibility of a death from an insecticide which in malarious areas has saved the lives of many millions. But in countries with sophisticated health services and high standards of living, factors which may postpone chronic, degenerative, or malignant disease and death may become identifiable within the current decade. Whether, once identified, they can be modified or removed, we do not know, but this kind of development could become far more important than the advancing edge of surgery. We must put more effort into the training of specialist epidemiologists, the indoctrination of the profession with the importance of the subject and the establishment of opportunities for them to do their job once they have been trained for it.

2

Developing theory for practice: evaluating options for screening

A simulation system for screening procedures

Computer simulations of cervical cytology screening programmes

E. G. Knox
MD, FRCP, FFCM
Professor of Social Medicine,
University of Birmingham

A simulation system for screening procedures

Summary. *The place of simulation procedures in the planning of providing medical care is described. The theme is developed in relation to the evaluation of options related to screening procedures and the basic requirements of a simulation system are identified. The development of a practical general system is described.*

INTRODUCTION

Any proposal to introduce a large-scale screening procedure as a definitive service, faces a number of taxing questions. How sensitive and specific is the test in the context of its proposed application? How acceptable is it to the population to which it is offered? To what extent will its sensitivity, specificity, and acceptability influence its efficiency? What is the natural history and significance of the untreated disease? To what extent does early diagnosis improve the effectiveness of available treatment? How many people will be harmed by false positive and false negative results: by the procedure itself or through unnecessary treatment? What will the test and consequential treatments cost in money and manpower? What resources will be saved by the scheme? How many lives will be saved?

These questions range far beyond the particulars of the diagnostic test; they are largely operational in character, and are concerned more with the problems of converting a procedure to a service than with the procedure itself. They are important questions for three much-quoted reasons. First, a screening procedure departs from

traditional medical practice in deliberately soliciting custom, and its advocates bear a greater responsibility for knowing that it offers a net benefit, than when the patient himself initiates the contract. Second, the yields of many procedures are so low, and the normals so outnumber those who can be helped, that a small measure of harm to normal people can in aggregate easily outweigh a substantial benefit to the sick. Third, a substantial redeployment of available resources, of money, manpower, equipment, time, and premises, must certainly harm competing proposals for innovation, or the maintenance and development of existing services. Indeed, the necessary resources may not always be available.

Practical experience suggests, however, that it is only in exceptional cases that all questions can be answered satisfactorily before a service is offered; on the contrary it is often necessary to offer a service in order to answer questions. It is necessary that these needs be met by a process of planned and controlled escalation in the scope and scale of a service, which must therefore embody the means for formulating and answering successive biological and operational problems, partly through its in-built information systems and partly through superimposed *ad-hoc* experiments. The formal image of this process consists of a continuing cybernetic cycle with alternating inferential and operational activities. Unfortunately, and to the detriment of developing health services, this necessary mechanism has seldom been successfully established, partly because it is administratively difficult to do so, but also because its nature is widely misunderstood.

The inferential half of the cycle, based upon methods and philosophies transplanted into medicine from other disciplines, is the better understood and most widely accepted part of the process. In the field of health services research it is expressed in a wide range of essentially statistical techniques, both observational and experimental. By contrast, the operational half of the cycle has so far largely excluded itself from any comparable disciplines of technique and from the tradition of open publication. This sequestration is expressed in the administrative separation of research and service functions. The responsible staffs are separated, commonly dwell in different organizations, and appear to be afflicted with different habits of thought. Indeed, the attitudes of some administrators, clinicians, and scientists, would suggest that they were in some kind of mutual opposition. If we are to overcome traditional antipathies and replace them by a cyclical process capable of promoting and controlling the develop-

ment of services such as mass screening, we shall have to bring all the steps within the scope of a common discipline of technique and a common tradition of open communication.

It is in this spirit that the present study was conceived and its objective is to explore the possibility of adapting to the field of medical policymaking, a technique which has already been developed to a considerable extent in areas outside medicine, namely the technique of computer simulation.

Policies and predictions

The justification of the simulation approach to policymaking depends on two simple propositions.

1. All rational decisions are based upon predictions.

2. All rational predictions are based upon models.

The first of these propositions is unexceptionable, at least to intuitive judgement, except for being so obvious as to be virtually a tautology; we are simply asserting that if a decision to act is not based upon a prediction of the outcome of the action, and of alternative actions (including inaction), then it is not rational.

The truth of the second proposition depends on what we mean by a model. The term can be used to cover any expressible system of systematic connections between available premises and their consequences, either those consequences which occur later in time, or those occurring in other circumstances; that is, the extrapolation is not necessarily time-based. Available species of model range from vague and essentially intuitive appreciations of relationships, through pictorial or diagrammatic (ie iconic) models, to explicit verbal and mathematical models.

It cannot be held to be self-evident that explicit models are always more effective tools of prediction than are intuitive modelling processes. But it is a matter of common observation that intuitive predictions are often wrong, that necessary premises and facts are often omitted from consideration in error, or because their necessity was not appreciated. The main grounds for preferring explicit modelling methods relate to their capacity for being communicated from one person to another, their usefulness as tools for settling differences between different intuitive judgements, and their capacity for resolving

the basis of a disagreement into three separate parts. Is it a question of arithmetic? Is it a question of the premises used? Is it a question of the postulated mechanism?

The most explicit (ie least ambiguous) models available to us are those expressed in terms of mathematical relationships but even here a range of types and a range of purposes can be distinguished. At one extreme we recognize the classical mathematical model in which a complex set of observed relationships (say, between biological estimations) is summarized in a relatively concise algebraic equation. The objective here is indeed to be concise, to summarize a complex situation in a form which is convenient, and which as a secondary function may express a dependency relationship in causal or mechanistic terms. At the other extreme we recognize models whose primary purpose is to express the process rather than summarize the result, and where comparisons between the outcome of the process and available observations are used mainly to validate assumptions about the process. We call these models 'simulations'. Conciseness is not their objective. They can in fact be quite complex, and they are used essentially to extrapolate beyond the limits of available observations either in terms of circumstance or of time. The results of these extrapolations are not necessarily obvious to intuitive examination; indeed, the chief value of simulation appears when, in a complex situation, it predicts a result which was unexpected.

The construction and use of simulation models enforce the clear separation of premises from the mechanisms, and when predictions fail to match observations they permit the separate adjustment of each. Their use separates irrelevant considerations from the argument and forbids the omission of necessary declarations about premises. Guesses are recognized as guesses. When written in a computer language they bring a disciplined and unambiguous syntax into play, and at the same time the statement of the model is converted into a powerful working device. As such it enables the user to explore a range of uncertainty for each of the required declarations and to express each uncertainty in terms of a range of consequences.

At a deeper level the design and use of simulation systems assists in identifying the basic nature of the constructs and concepts of rational argument. Thus, means, proportions, or percentages, for example, may be computed directly from data, and belong to the analytic or reductive aspect of the cyclic process referred to earlier. A 'natural history', a 'pathological process', or a 'behavioural pattern' of any

kind cannot be so computed; it must be declared, and it is the consequences of the declaration which may be computed. Concepts such as these latter ones belong to another part of the cycle, and we may refer to them as models. Models, if we include intuitive processes within this term, are widely used in medicine.

It is helpful, finally, to remember that all predictions based on models are conditional upon the truth of the premises (such statements always begin with 'if') and that all model systems are susceptible to *reductio ad absurdum*.

SIMULATION OPTIONS FOR SCREENING PROGRAMMES

The chief variables about which assumptions must be made in planning or predicting the result of a screening programme are:

1. The structure of the population to be screened, especially its age and sex distribution.
2. The natural history of the disease processes to be interrupted.
3. The sensitivities, specificities, and acceptabilities of the screening procedures available.
4. The effect of diagnosis and subsequent treatment upon the natural history of the conditions detected.
5. The proposed policy for the deployment of resources, notably the numbers and types of tests to be offered, the ages at which they are to be offered, and if more than one is offered, their sequence.

The problems of specifying these variables will be discussed and the methods chosen in attempting a practical simulation will now be described.

The population. Natural populations are dynamic entities subject to gains and losses through births and deaths, exhibiting immigration and emigration across geographical boundaries. They consist simultaneously of many cohorts, each with a different experience of exposure to the risk of disease and with a different history of medical care. The kind of problem which stems from this in the context of screening is illustrated by studies of cervical pathology, where the measured incidence of onset of carcinoma *in-situ* of the cervix in successive periods between ages 20 and 35 accumulate to a value which is far in excess of the measured prevalence of carcinoma *in-situ* and more advanced lesions, at ages 35–50 (1).

We cannot, without appeal to other evidence, distinguish between the two biological hypotheses (*a*) that carcinoma *in-situ* frequently

reverts to normal and (*b*) that the experience of the younger cohorts on whom the incidence estimates were based, is entirely different from that of the older cohorts on whom the prevalence estimates are based.

Problems of this kind add a dimension to the complexity of constructing a model and it was decided in the first instance to opt for a simpler system in which a single cohort of 10,000 persons was traced from birth to death according to the mortalities expressed in a standard life-table. Life-tables are themselves cross-sectional artifacts compiled from age-specific mortalities of different cohorts, but recent rates of change of total mortality in most western countries, where screening policies can seriously be considered, have been sufficiently low for this not to be a serious perturbation.

For purposes of the simulation a declaration of l_x values by five-year intervals from birth (10,000) to age 95 (arbitrarily zero) is sufficient, and linear interpolations of annual values are sufficiently accurate and can easily be computed.

Natural history. The life-table provides data specifying rates of transfer, by age, from alive to dead. It can be regarded as an overriding element of natural history. Completion of the picture requires further resolution, namely a declaration of 'states' other than 'alive' and 'dead of any cause'; also a specification of transfer rates between these states. The degree of complexity of these declarations is arbitrary; it needs only to be sufficiently complex to produce an adequate simulation of what is believed to occur, and sufficiently simple to be comprehensible.

For example, it might be sufficient for a particular application to divide the living population into 'normal', 'pre-glaucoma', and 'glaucoma'; that is, three states. In another it might be necessary to specify six levels of blood pressure. Another simulation might require 'normal', 'dysplasia', 'carcinoma *in-situ*', 'occult invasive cancer', 'early clinical cancer', 'late clinical cancer', and 'dead of this disease'. We could, if we wished, distinguish between 'normal' and 'reverted normal', or between 'progressive dysplasia' and 'regressive dysplasia', and so on to any necessary degree of resolution. In addition it may be necessary to declare the existence of states which occur only as the result of intervention, either in connection with the screening programme or independently of it. In the first group we might wish to declare the existence of 'coned normal', 'coned dysplasia', 'coned

carcinoma *in-situ*', 'coned occult', 'treated early disease', and 'treated late disease'. In the second group we may wish to declare 'hysterectomy for reasons other than cancer'. The latter would be especially useful in adjusting a computed rate for cancer of cervix from one based upon a count of heads, to one based upon a count of cervices at risk.

For the purposes of the simulation a facility was designed which permits the declaration of up to twenty-three pathological states in addition to three obligatory states namely 'normal', 'dead of this disease', and 'dead of other cause'. Cards are punched with the names of the declared states; the three obligatory states are assigned the serial numbers 1, 25, and 26, while the optional declarations are assigned serial numbers 2, 3, 4, . . . 24 in the order of their presentation.

The number of possible transitions between n live states, in either direction, together with one-way transitions from any live state to either dead state (25 or 26), is $n(n+1)$, and a complex pattern of recognizable states could readily generate several hundreds of transitions, each warranting the specification of a transfer rate. Furthermore, each rate might require modification according to the age of the patient and the duration of the current state. Therefore a comprehensive declaration requirement would often surpass the practicalities of patience, let alone the precision of current knowledge, and even on a large computer might require prohibitive processing time.

Limitation through restriction to orderly progressions would have barred many desirable options and the solution adopted was to allow the user of the system to specify as few or as many of the possible transfers as he wished, up to some arbitrary maximum (50, initially). A transfer function is specified as a pair of 'state-serial-numbers' representing the state from which and to which the transfer is to be made and this is followed by a statement of the transfer rate expressed per cent per annum. In fact a series of transfer rates is required, each appropriate to a different age or a different duration of the pathological state. Thus a series of pairs of numbers must be specified; for example, the sequence 22, 0·05, 32, 0·04, represents a specification that from age 22 the transfer rate is 0·05 per cent per annum, and continues at this level up to age 32 when it changes to 0·04 per cent per annum, at which level it will continue either indefinitely or until another positive integer specifies the next age at which a change will occur. Each transfer-type can accommodate up to twenty changes of

this kind. The complexities of specifying transfer rates both by age and by duration of state were avoided arbitrarily by limiting age-specifications to 'normals' and duration-specifications to 'other live states'. Thus, transfers which exit from state '1' treat the initial number of each pair as an age, while transfers which exit from other states treat this number as the duration of the state (strictly, the ordinal number of the current year).

The screening procedures. Each procedure has a name and the simulation provides a facility for stating this name on a card. This must be followed by a statement of the groups (ie state–serial–numbers) to which the procedure can be applied. In practice a procedure is applied to clinical types, undifferentiated according to their pathology, whereas the types specified in this simulation are defined in pathological terms. However, this difficulty is overcome by specifying the pathological types in a manner which permits their grouping into clinical classes. Thus, cervical smears would be applied to 'normal', 'dysplasia', 'carcinoma *in-situ*', and 'occult carcinoma', but not to 'treated cancer' or 'hysterectomy for causes other than cancer'.

Next it is necessary to make a statement of intent or of assumption concerning the state to which the person will be transferred through treatment or other decision, according to the initial state and the result of the test. In other words, each application of the screening procedure is associated with transfer specifications analogous with elements of the natural history. One decision is needed for a positive result and another for a negative result. It is possible of course that there will be no transfer (specified as a self transfer) but for most positive results a transfer will occur. For example, 'normal' may be transferred to 'biopsied normal', or 'severe hypertension' to 'treated severe hypertension'. Exits from states attained only through intervention can be specified in the previous natural history section.

In addition to the transfer specifications associated with each screening procedure it is necessary to declare an anticipated 'percentage positive' for each pathological type screened. Thus, cervical smear attached to type 'normal' might be specified as giving 0·01 per cent positive, whereas its application to 'carcinoma *in-situ*' might be specified as 80 per cent positive.

The simulation system was designed to accommodate more than one screening procedure and up to six (arbitrarily) can be listed. They can be applied independently at different ages or simultaneously at

the same ages, when they are assumed to be applied in the order of their presentation. If a later procedure is intended to be applied only to those selected as positive by an earlier procedure, this can be accommodated by transference to an intermediate pathological state declared specially for the purpose, and acted upon only by the later procedure.

The simulation approach throws some light upon the concepts of sensitivity and specificity. Sensitivity is usually defined as the proportion of those with the disease, who are detected by the test. Specificity is usually defined as the proportion of those without the disease, who are passed by the test.

Both proportions are declared in the simulation system in identical terms, as the 'percentage-positive' estimate attached to the different initial states. The formal identity of the two concepts is confirmed if we restate them in terms of states x and y and in terms of results p and q, instead of using polarized terms such as diseased and well, or positive and negative. Sensitivity and specificity are also seen to be incomplete specifications of accuracy in the sense that there may be more than one pathological state to be detected and more than one to be excluded. These indices are incomplete also in another sense; even in simple situations they do not together specify the balance of hazards and benefits which a test offers, and for this we need also to know the prevalence of the disease in the population. Finally, the denominators of the two indices are usually dependent upon the result of yet another test, so that estimates of sensitivities and of specificities based on observational studies are always relative rather than absolute values.

The fundamental lesson of the simulation analysis is that in strict terms, sensitivity and specificity are model concepts and, like a natural history (for example) are declarable rather than computable values. *Post hoc* evaluation of a screening programme is more appropriately expressed as a simple count of the known true positives, false positives, false negatives, and true negatives.

Resource-deployment policies. Some aspects of the policy options have already been covered in specifying what tests are to be made available at all, and to what clinical and pathological groups they are to be applied. The possibility of sequencing tests in particular orders on single occasions has also been noted. The main remaining option concerns the ages at which offers of tests are to be made, within which

option we can declare their total numbers. The simulation was designed to provide for the specification of up to thirty offers for each test, declared by stating the ages. In fact, this embodies a 'trick' facility for going beyond thirty offers through declaring a test more than once. In making each offer it is necessary to state also the anticipated percentage acceptance of the offer. Each application is therefore expressed as a pair of numerical values.

Acceptability is known, in practical situations, to vary from class to class of person. For example, different attendance figures are observed for cervical cytology screening according to social class, parity, and so on. This concerns the simulator when, as is sometimes the case, the criteria determining acceptability are also associated with different attack rates or different natural histories. The problem was met in the simulation system by introducing a facility for stating differential weightings of test acceptability, directly related to the existing pathological state. In practice these weightings are most conveniently entered at the time that the pathological states are initially named and allocated to their sequence numbers.

The approach to resource deployment used here differs from that which might be taken if the resource limits were decided in advance as a policy and the problem was only one of optimum development. This is seldom practical in the administration of individual services and, in addition, the primary medical problem concerns the balance of hazard and benefit offered to patients. The balance of net benefit to financial and other resource cost is usually a secondary consideration. Estimates of available resources are not therefore entered as data. The user of the system must himself take resource constraints into account when deciding the range of policy options to be explored.

Mechanism and results. The model is written in Fortran and, as indicated already, the inputs are prepared on punched cards. The basic sequence of the system is already evident and the first stages consist in reading and validity-checking of the various premises described. A certain amount of rearrangement is necessary together with subsidiary computational work (interpolations, totals, percentages, etc.), and they are printed out for visual checking of their correctness, and to label for future reference the consequences which will follow.

The basic computational sequence follows a cohort of 10,000 persons in year-by-year steps, each step leading to application in turn of

(*a*) the mortalities of the life-table, (*b*) transfers specified by the natural history, and (*c*) transfers resulting from applicable procedures. The transfers are effected on a deterministic and proportional basis rather than upon a stochastic basis. Although this option invokes the unrealistic concept of transferring fractions of persons from one state to another (instead of using fractional probabilities of transferring whole persons) it offers advantages of simplicity in designing and running the programme and requires only one run, instead of many, for each separate set of premises. It is recognized however that situations representing small populations and very low yields may in time require stochastic adaptation.

The frequency of display of interim results, including prevalence, incidence, distribution of states, and the performance of the tests, may be determined by the user.

Final results consists of:

1. A year-by-year display of the frequency distribution of pathological states, expressed both in relation to the 10,000 persons born, and as a prevalence related ($\times 10^{-3}$) to those still alive.

2. An accumulated statement of procedure performance in terms of tests offered, refused, positive, and negative; similar to that provided in the interim displays.

3. An accumulated statement of all transfers into and out of every pathological state; similar to the interim transfer statements.

4. A statement for every procedure, of the actual pathological states of those detected as 'positive'.

THE SYSTEM IN USE

The programme is a substantial one comprising approximately 650 Fortran statements. Running time is also substantial if a large programme of runs is needed, approximately eight minutes per run on a KDF9, thirty-five minutes on a Univac 418 III without floating point hardware, or two minutes per run with the hardware. The system includes an optional return loop so that several policy options (for example) can be run as a single job, and this is useful once the background iterations are completed.

The main problems in use are:

1. The relative complexity of the necessary declarations, and of the process of their adjustment (usually one run at a time, here) until a reasonable fit with available mortality and prevalence data has been achieved.

2. The complexity of the outputs, which usually require a separate graph-drawing operation for a reasonably clear view of results of interest. However, the points of interest differ for different screening problems and it is not easy to devise simplified numerical displays or automatic graphical outputs which would be universally useful.

The system requires the development of some skill and a certain amount of practice before it can be used effectively. This in itself is a salutary reminder of the fallibility of initial essays. Too often these are mounted upon an inadequate analysis of available resources, or resource requirements, or the consequences of factual and theoretical uncertainties, and feed-back may require a wait of ten years instead of five minutes. However, the uses and limitations of the simulation system described above will be best illustrated in tackling real policy problems, and exercises relating to a series of outstanding questions will be the subject of subsequent papers the first of which follows.

Computer simulations of cervical cytology screening programmes

The effectiveness of cervical cytology screening has never been measured by randomized trial. Available evaluations have only compared the observed mortalities, prevalences, and incidences, at different times and places, and in different groups of women. Unfortunately, because the compared groups were never strictly comparable in all respects apart from treatment, observed differences have proved susceptible to explanations other than the supposed effectiveness of screening. Equally, absence of difference or change can often be attributed to causes other than the non-effectiveness of a programme. Even when a programme is considered to have caused a measured change, this may be attributed to concomitant combinations of better treatment, earlier diagnosis of clinical disease, detection of occult invasive cancer, and reduction in the number of cervices at risk, through excision or through hysterectomy for unrelated causes. Thus, even when a programme is effective, the mechanism may not be through the removal

of pre-invasive lesions, and it is difficult to guide a screening policy on the basis of the results.

Specifically, if an effect were mediated mainly through the early excision of otherwise progressive invasive lesions, then maximum benefit for minimum harm would be obtained by screening those ages where the incidence of invasive disease began to rise quite sharply, and by repeating the test frequently. On the other hand, if *in-situ* lesions were always progressive, but seldom very rapidly progressive, and if no invasive lesions appeared without a prolonged *in-situ* stage, then the deployment of resources would be quite different. The optimal age would be early, immediately following the onset of the majority of *in-situ* lesions, and a very limited number of repetitions would suffice.

Because of the difficulties of interpreting non-experimental facts, cervical cytology screening proposals and programmes have been justified mainly on theoretical grounds. Over-reliance upon theoretical arguments was responsible for the absence of trials at a time when they would have been possible, but also for subsequent failures to measure the effectiveness of many established programmes, or to try to validate the theoretical basis. The main theoretical plank, and the main source of contention, concerns the natural history of the various pathological states recognized in the cervix uteri.

All levels of opinion are available. Many gynaecologists and cytologists believe carcinoma *in-situ* never regresses to normal and, given time, will usually progress to invasive disease. Many believe that dysplasia of the cervical epithelium is an earlier stage with the same non-regressive, and ultimately progressive, characteristics. Others believe that dysplasia may regress and some postulate distinct regressive and progressive varieties. The same picture has been postulated for carcinoma *in-situ* and some workers regard the lesion as usually regressive and only exceptionally progressive. The 'two-type' hypothesis can be resolved into a continuous spectrum of liability to regression or progression and the ratio of one type to the other may be seen as varying with age or with other circumstances.

The ignorance on which these variations are based springs from the method of diagnosis. Carcinoma *in-situ*, in particular, is defined topographically and its accurate diagnosis demands complete excision. This destroys the lesion and prevents accurate repeated classification of the pathological state, which is of course a necessity for following the natural history directly. Repeated cytological examination

is non-destructive but error-prone. Colpomicroscopy may sharpen the possibilities of continuous surveillance but the great majority of natural history assessments have been based upon observations which simply were not capable of distinguishing one possibility from another. Very many conclusions on natural history have been based illegitimately upon prevalance data. Prevalence estimates are not dimensioned in time and are not capable of commenting upon natural history hypotheses, which are so dimensioned.

Even time-dimensioned epidemiological data offer serious difficulties of interpretation, as instanced in the British Columbia Study (1). This is one of the very few investigations to produce age-specific estimates both of prevalence and incidence of carcinoma *in-situ*. The age-specific incidence estimates are based upon repeat tests after measured intervals, in women previously negative, and in series they permit an estimate of accumulated incidence of onset, by age. This proves to be far in excess of directly measured prevalences by age. The gap is not closed by supposing that carcinoma *in-situ* converts to invasive cancer at rates comparable with those observed elsewhere, or in British Columbia. The most direct interpretation is that carcinoma *in-situ* is a reversible condition, especially in younger women. However, there are two alternative mechanisms which could explain part or the whole of the gap. The first is an artefact which would arise if the test had a high negative-error rate. If some of the cases on which incidence rates were based were missed old cases rather than genuine new cases, the incidence curve would be lifted and the prevalence curve depressed. The second possible explanation is that the women on whom the prevalence estimates were based belonged to a different cohort from those on whom the incidence estimates were based. A substantial change in sexual behaviour during recent decades could have resulted in differences between the two cohorts and part of the gap could be explained in this way. The results published to date do not permit clear distinctions between these possibilities; nor do they allow us to estimate their relative importance in explaining the gap.

Rational planning demands more than a knowledge of the natural history of carcinoma *in-situ*. Information about negative-error rates at different ages is necessary if we are to plan a level of insurance against mistakes. A knowledge of the false positive error rates in normal women is also necessary if we wish to maximize the ratio of cervices profitably coned to those unnecessarily removed. We should know something of the probable attendance rates at different ages if

we hope to deploy our resources effectively. We should know something of the effectiveness of 'normal' diagnostic and therapeutic procedures at different ages and stages if we wish to maximize the additional benefits of providing screening over and above the traditional procedures. Variations in the attendance responses of women in different pathological (or normal) states are a further prerequisite for knowing how to present a campaign.

All these are necessary to a rational prediction of the effects of a policy choice related either to alternative deployments of a fixed resource, or proposals for increasing the total resource invested. Yet all are uncertain to a greater or lesser degree and it is difficult on intuitive grounds to say which are the critical or important uncertainties, the most urgent to be resolved, and which are less important or trivial. It is particuarly difficult to make these predictions quantitatively and to do so in any publicly verifiable sense.

Objectives of the present study

The present study explores the value of applying computer simulation methods to these problems. The system used (see the first part of this paper) is a general simulation system capable of mimicking many screening procedures, provided only that the necessary premises of the simulation are specified exactly. They include:

1. A statement of the life-table of the population to be screened.
2. The names of the pathological states recognized.
3. A statement of the transition frequencies, per cent per annum, according to age and duration, between selected pairs of named pathological states.
4. The names of the screening procedures to be applied, in sequence, the groups to which they are to be applied, the percentage of positives expected for each pathological state, and the new pathological states to which patients will be transferred when they are found to be either positive or negative to the test.
5. The ages at which each test is to be offered, and the percentage attendance expected at each age.

Uncertain premises are treated by calling for a series of computer runs, the premises being reset at different values within the envisaged limits of uncertainty. Combinations of ranges of values for different factors are presented. The usual practice in simulation procedures, and the one which will be followed here, is to adjust the initial

conditions (1–3 above) until reasonable mimickry of the 'at rest' background is obtained and then to experiment with the procedural variables (4 and 5 above). Different sets of background premises may be capable of explaining the 'at rest' observations with equal facility, and it is then necessary to explore the effects of the procedural variations against a variety of initial backgrounds.

Background conditions

I. Life-table data were obtained from the Registrar-General's Reports for England and Wales in 1970. The l_x data relating to 10,000 females born, at five-year intervals, were punched into a card. These figures have been used without variation in the whole of the simulation exercise which follows.

II. The list of pathological states selected was:

1. 'Normal.'
2. 'Reverted normal.'
3. 'Dysplasia, regressive type.'
4. 'Dysplasia, progressive type.'
5. 'Carcinoma *in-situ*/young type.'
6. 'Carcinoma *in-situ*/older type.'
7. 'Occult invasive disease.'
8. 'Clinical invasive disease, early.'
9. 'Clinical invasive disease, late.'
10. 'Coned normal.'
11. 'Coned dysplasia.'
12. 'Coned carcinoma *in-situ*.'
13. 'Coned occult invasive.'
14. 'Treated early invasive.'
15. 'Treated late invasive.'
16. 'Hysterectomy, not for carcinoma.'

.
.
.

25. 'Dead of this disease.'
26. 'Dead from other cause.'

This list was also used without variation in the examinations which follow. This is without prejudice to the possibilities that 'reverted normal' is indistinguishable from 'normal', or 'carcinoma *in-situ*/young type' from 'carcinoma *in-situ*/older type' (for example), be-

cause it is possible later to discard particular pathological types by not effecting any transfers into them. The list was intended to be comprehensive for a range of simulation purposes and to supply a standard set of names of which subsets could be used in particular runs.

III. The specifications of the natural history were developed iteratively by making a series of proposals about transfer rates between pathological types, computing results, comparing the results with available observations, and readjusting the specifications for another run. The basic observations used for this fitting process were mainly age-specific prevalence data available from various studies (for example, British Columbia) and mortality and onset data available from other sources (for example, Cancer Registries, Registrar-General's Data, etc.). It was stated earlier that these data do not distinguish critically between the different natural history hypotheses, and as expected, several alternative specifications of natural history proved capable of mimicking observed results. There were in addition many specifications *not* capable of producing these results. The practical problem was to construct a small number of competent natural histories which would in effect cover the range of possibilities envisaged by most pathologists and cytologists. Two model natural history specifications were eventually chosen and they have been labelled respectively the 'dynamic-A' and the 'progressive-A' specifications.

Parts of the two specifications are common, particularly those parts concerned with the progression of invasive disease, and the main differences between them concern the transition frequencies of non-invasive lesions.

COMMON ELEMENTS

All the non-clinical conditions (1–7 above) were allocated transfer rates to 'hysterectomy, not for carcinoma' (type 16). These transfers began at age 20 for 'normals'. The initial transfer rate was 0·05 per cent per annum, changing to 0·5, 1·0, 0·5, 0·3, and 0·1 at ages 30, 40, 50, 60, and 70 respectively. Transfers from the other pre-clinical states (2, 3, 4, 5, 6, 7) were expressed in terms of the duration of the current state but were designed to produce a similar curve in terms of age. However, in sympathy with the known association between the risk of malignant disease and the occurrence of other pathologies which could lead to hysterectomy, the rates were set at about $\times 1{\cdot}5$ the rate for normals.

The hysterectomy rate specifications were constructed iteratively to

mimic a 'mid-Atlantic' population construed from various sources. The effect was to produce a hysterectomy prevalence (per hundred living) under background conditions, of 0·1, 1·0, 6·3, 14·7, 18·7, 21·0, 21·8, 22·7, at ages 20, 30, . . . 90, respectively.

The effect is to bring the denominator of cervices at risk of cancer into a rapid decline, several years before that which would have occurred if the life-table sequence had operated alone. The period of rapid fall occurs in the age interval 40–60 when there is a high rate for carcinoma of cervix, and a high mortality. This amplifies a point which has been made more than once, namely that recent changes in the frequency of hysterectomy, particularly in North America, are capable of producing quite rapid declines in the occurrence and mortality of carcinoma of cervix when they are measured against a count of heads, rather than against a count of cervices.

Transfer rates from clinical invasive disease (types 8 and 9) to treated clinical disease (types 14 and 15) were set at 50 per cent per annum starting in the first year, in both natural histories. Progressions of types 14 and 15 to type 25 (dead of this disease) were likewise common to both specifications. For treated early disease the mortality was set initially at 1·0 per cent per annum rising by steps to a maximum of 5 per cent per annum after eight years, where it remained steady for another seven years and then disappeared. For treated late disease the mortality started at 5 per cent per annum, rose to 10 per cent per annum in the third year, declined to 5 per cent per annum after five years and to zero after ten years. The second of these two specifications kills approximately 49 per cent of those who do not die from something else first; the first will kill about 32 per cent. In addition, a proportion of 'late clinical invasive' were set to die without being treated at all. Final mortalities depend upon other elements in the simulation but under background conditions these model processes resulted in a final case-fatality of 42·9 per cent in the context of the progressive natural history specification, and 39·8 per cent in the presence of the dynamic natural history.

Several pathological types (10–13) are attained only through screening procedures and not through components of the natural history. However, they can be allocated natural exit rates. Accordingly, small recurrence rates for treated carcinoma *in-situ* and a small rate of invasive progression both for treated carcinoma *in-situ* and treated occult invasive disease (0·05 and 0·5 per cent per annum respectively, after an interval) were specified in common for both natural histories.

DIFFERING ELEMENTS

The progressive and dynamic natural histories differed in their usage of the available pathological classes.

The progressive specification treated 'reverted normal', 'dysplasia, regressive type', and 'carcinoma *in-situ*/young type' (types 2, 3, 5) as redundant, and no entries to these types were specified. Transfer probabilities were specified for 'normal' to 'dysplasia' (4), 'normal' to 'carcinoma *in-situ*' (6), and 'dysplasia' to 'carcinoma *in-situ*'; also from 'carcinoma *in-situ*' to 'occult invasive disease', from 'occult' to 'clinical invasive disease', and directly from 'carcinoma *in-situ*' to 'clinical invasive disease'. The only 'reverse' movement allowed was from 'dysplasia' to 'normal'. A set of transfer rates was established by iteration, in order to produce a set of prevalences, incidences, and mortalities which fitted as near as possible a pattern reflecting available published data. Transfers from 'normal' to 'carcinoma *in-situ*' began at age 20, at 0·02 per cent per annum, changing to 0·04, 0·05, 0·04, and 0·02 at 25, 30, 70, and 80 years respectively. Transfers from 'normal' to 'dysplasia' began, and reached peak values, at earlier ages but were of comparable extent. 'Dysplasia' progressed to 'carcinoma *in-situ*' or regressed to 'normal' in a 1:2 ratio at rates between 4 and 10 per cent per annum according to the duration of the disease. 'Carcinoma *in-situ*' was made capable of rapid progression in some cases, transferring to 'occult invasive disease' at 4 per cent per annum during the second year of its existence and subsequently 8, 16, and 8 per cent per annum in the fourth, sixth, and fifteenth years. Direct progression of 'carcinoma *in-situ*' to 'clinical invasive disease' (at 10 per cent per annum), omitting the occult stage, did not begin until the eighth year.

These progression rates differ somewhat from notions held by some pathologists to the extent that the 'carcinoma *in-situ*' onsets continued to quite advanced ages and progression of 'carcinoma *in-situ*' to 'occult invasive disease' was rather more rapid than is sometimes believed to be the case. These features arose in iteration simply because onsets limited (for example) to women under 30 and latent periods of ten to fifteen years, when constrained by non-reversibility, proved to be incapable of mimicking available observations on prevalence incidence and mortality. This, of course, is a reflection of the incompatibilities between incidence and prevalence estimates already noted in the British Columbia data.

The dynamic natural history specification was also derived by

iteration. The main difference was that 'carcinoma *in-situ*' was allowed to regress. The full list of pathological states was used, covering two types of 'dysplasia' and two types of 'carcinoma *in-situ*', to reflect a hypothetical range of different natural history types. Direct transfers were specified from 'normal' (type 1) to each of the abnormal classes which had not yet come to treatment (types 3–9) together with transfers back to 'reverted normal' (type 2) from either of the 'dysplasia' or either of the 'carcinoma *in-situ*' (types 3, 4, 5, 6). Both kinds of 'dysplasia' could either progress or regress but type 3 regressed more often than it progressed and type 4 the reverse; these transfer rates were high, between 10 and 30 per cent per annum, and began in the first year. 'Carcinoma *in-situ*/young type' reverted to 'normal' at 15 per cent per annum from the second year and did not begin to progress to 'invasive disease' until the fifth year. 'Carcinoma *in-situ*/older type' also had a regression rate (10 per cent per annum) but not until a later stage and it began in the second year to progress to 'occult invasive disease'. The initial transfer rate of 10 per cent per annum rose later to 15 per cent and finally to 20 per cent per annum. There were in addition a number of direct skips from 'carcinoma *in-situ*' to 'clinical invasive disease' after a longer duration, fifteen years. Direct transfers from 'normal' to 'clinical invasive disease' (types 8 and 9) were also specified; the implications of this will be discussed later.

ATTENDANCE LOADINGS

It has been shown repeatedly that women in classes with high risks for 'carcinoma *in-situ*' or for 'invasive carcinoma', attend screening clinics less often than those in other groups. The simulation system permits allowance for this by introducing percentage attendance loadings for different pathological types. These values are taken to be relative to 'normal' attendance figures (to be declared later) for the particular test at the particular age. When this facility was employed the values used were 90 per cent attendance for pathological types 2, 3, 4; 80 per cent attendance for types 5, 6, 7; 70 per cent attendance for types 7, 8, 9.

Procedural variations

THE SCREENING TEST

Throughout this simulation only one procedure was on test, namely cervical scrape. In practice, various combinations of cervical scrape

and self-administered vaginal aspiration are possible, with different response expectations for each test in different groups of women, and different sensitivities for each in different pathological states. In the first instance, however, it was decided to explore only the consequences of the single test, but with a range of sensitivities and acceptabilities.

The procedure was ascribed different sensitivities for different pathological states; sensitivity was defined as the percentage classified as 'positive' as opposed to 'negative'. The simulation demands, as practice demands, a binary classification, and intermediate states (unsatisfactory, doubtful, etc.) were not allowed for. It was assumed for simulation purposes that doubtful and unsatisfactory tests will be reinvestigated to the point of resolution and the procedure specifications therefore refer to sequences of this kind, when appropriate, and not necessarily to single specimens.

The procedure was declared to be applicable to types 1–9, that is to all untreated states, both invasive and non-invasive. In each case it was decided that a negative result would be followed by no change of state but that a positive result would result in conversion to the appropriate treated state (type 10 to type 15).

Different sensitivities were ascribed to the test when applied to different pathological states. For the purposes of the simulation exercise the 'percentage positive' values ascribed to 'normal' and 'reverted normal' were 0·1 per cent.

This is an arbitrary value difficult to support from hard data; this point will be recalled when results are presented. Values of 60, 75, 80, 90, and 70 per cent were ascribed respectively to 'dysplasia', 'carcinoma *in-situ*', 'occult invasive', 'early clinical invasive', and 'late clinical invasive disease', respectively. These values are also arbitrary, but a rising percentage with increasing extent with a reversal of this trend for advanced disease, is well documented; values around 75 per cent sensitivity, for 'carcinoma *in-situ*', can be argued from several sources and are unlikely to be far in error. These values were used for the range of simulations which follow. The system will, of course, permit explorations of varying assumptions about sensitivity if they subsequently seem desirable.

POLICY

Policy options were mimicked by offering the test at different ages, and in numbers ranging from zero to twenty tests per woman. The

zero offers were used for base-line purposes to enable subsequent results to be expressed in terms of deaths and illnesses saved. The upper limit of twenty was based on the consideration that twenty tests per woman in England and Wales (say at ages 20, 23, . . . by three-year intervals . . . to 77) would, if the offer were taken up, require about 5×10^6 tests in England and Wales each year. The present number provided is between 1 and 2×10^6, so that an offer on the scale would be beyond the available and immediately foreseeable resources. On the grounds that simulations should span foreseeable realities, many of the simulations were carried out using sequences of five or ten tests per woman.

The different acceptance rates expected at different ages provided another dimension to the exploration. A range of contingencies and consequences was explored.

Uniformly, throughout the simulation, the pathological processes were started at age 16 and terminated at age 95.

Results

BACKGROUND CONDITIONS

The two natural histories, 'dynamic-A' and 'progressive-A', operating in the absence of any screening procedure, produced the results expressed in Figs. 1, 2, and 3. Fig. 1 describes chiefly the pattern of mortality for causes other than cancer of cervix, and of hysterectomy for causes other than cancer, and it respresents a cohort of 10,000 live-born females followed between the ages of 16 and 85. It represents the process in the presence of the progressive natural history, but differences between it and the dynamic natural history are scarcely noticeable on this scale. Fig. 2 (progressive) and Fig. 3 (dynamic) demonstrate on a larger scale the age-specific prevalences of carcinoma *in-situ*, of invasive but untreated cancer, of treated cancer, and deaths from cervical cancer. In these figures, each element except for the deaths is represented as a prevalence per thousand living; the deaths are a simple accumulation 'per 1,000 born'. Each element is superimposed upon the previous element so that the upper boundary represents the total accumulation of persons dead from cancer, or surviving with cancer or treated cancer, or with carcinoma *in-situ*. Both figures give in addition the curve of accumulated onsets of carcinoma *in-situ*, integrated over successive age periods and representing the total number of women who ever had the condition.

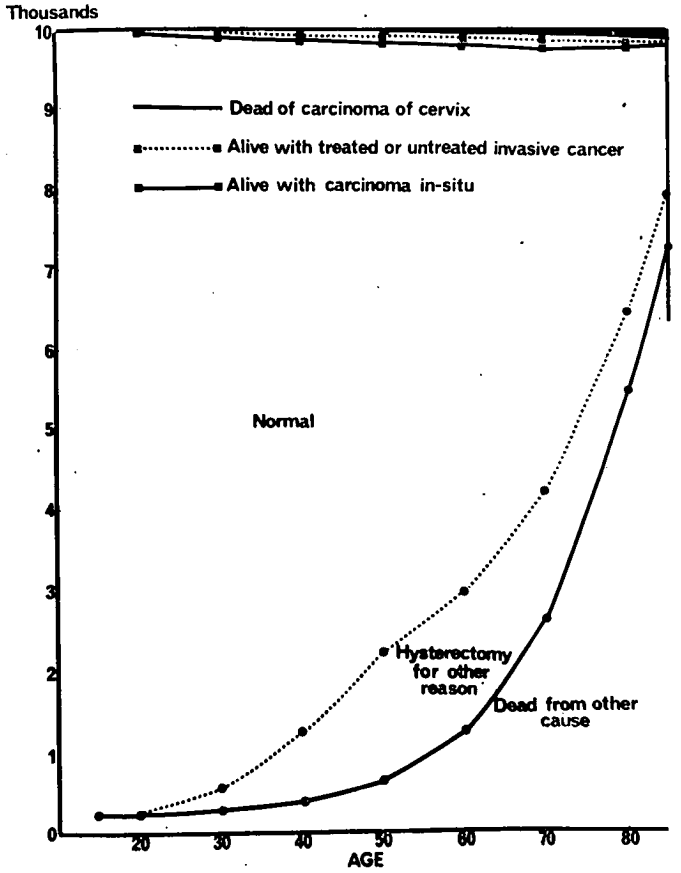


Fig. 1. Distribution of pathological types by age in cohort of 10,000 women.

Fig. 4 demonstrates a similar composite picture drawn from real life. The prevalence of carcinoma *in-situ* and of preclinical cancer, and the accumulated incidence of carcinoma *in-situ* are derived from the British Columbia study (1). The superimposed data for clinical carcinoma were obtained from Cancer Registry data in the Birmingham Hospital Region. The division between fatal and non-fatal clinical cancer was obtained by subtraction of national mortalities for England and Wales from the Birmingham-derived upper boundary.

Figs. 2, 3, and 4 each contain distortions. The model results are

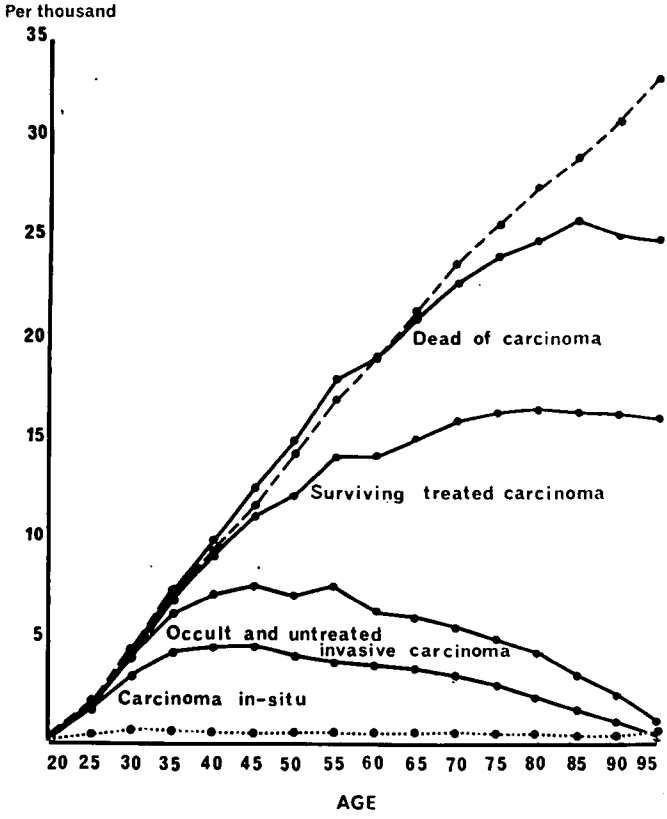


Fig. 2. Accumulated age-specific prevalence and incidence of carcinoma of the cervix: progressive natural history. Each solid line represents prevalence (10^{-3}), by age of all states named below the line. Intervals between solid lines represent the prevalence of the named state. Broken lines represent incidence (10^{-3} per annum) of carcinoma *in-situ* by year of age (lower line) and accumulated by year of age (upper line). Figs. 3 and 4 are similar.

inaccurate in the upper age-groups, the upper boundary of total prevalence becoming too high through the inconsistent method of adding the deaths. The curve of accumulated incidence cannot be accurately compared with total prevalence because it fails to allow for differential mortalities in those who developed carcinoma *in-situ*. However, the curves are reasonably accurate up to age 75 as attested by the close correspondence in the progressive model between the total prevalence and the accumulated incidence of onset of carcinoma

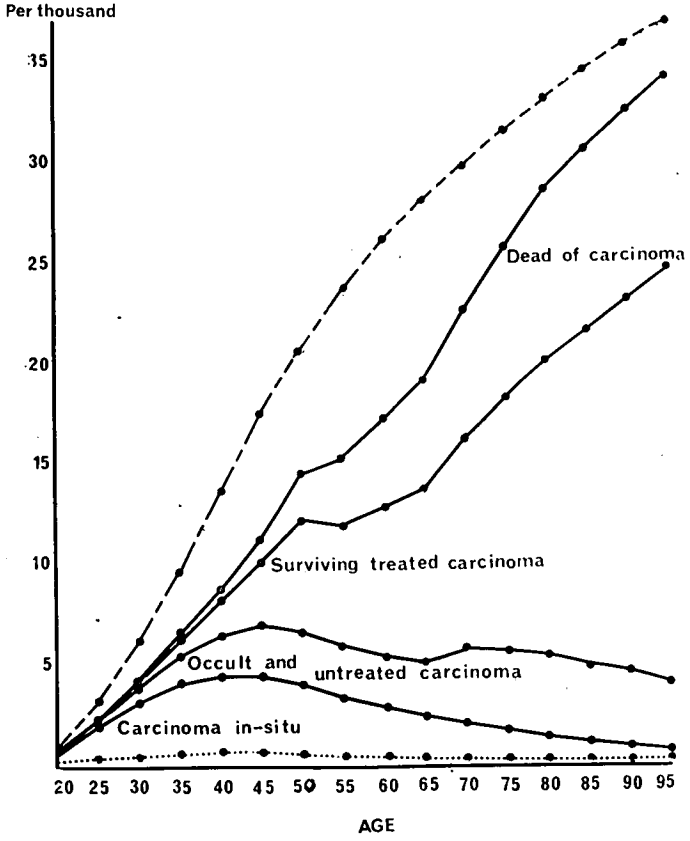


Fig. 3. Accumulated age-specific prevalence and incidence of carcinoma of the cervix: dynamic natural history.

in-situ. The wide gap in the dynamic model in the age interval 30–75, between the accumulated incidence of carcinoma *in-situ* and the total prevalence of cervical pathology, reflects the premise that a proportion of carcinoma *in-situ* revert to normal: on this showing, about one-third of all cases. However, the dynamic natural history hypothesis permits direct transfer to invasive cancer without the intermediate step of carcinoma *in-situ* and if all the cancers were supposed to have passed through this stage a substantially higher rate of reversion would be implied.

The premise of the direct step to invasive cancer was designed to

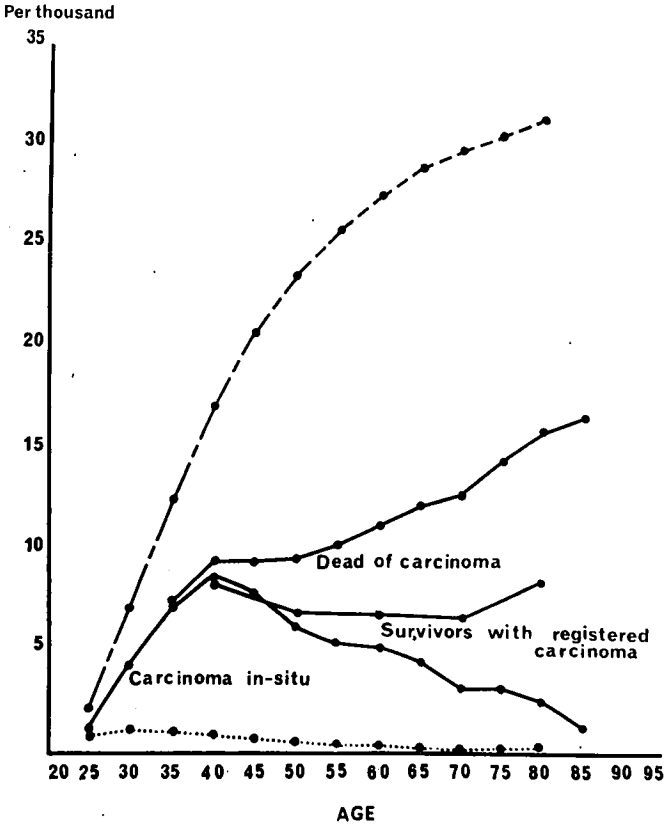


Fig. 4. Accumulated age-specific prevalence and incidence of carcinoma of the cervix: composite observations. In this figure the interval labelled 'survivors with registered carcinoma' is inferred indirectly by difference. In addition the prevalence of untreated (pre-registration) carcinoma is not known and not included. If it were included it would raise the level of total prevalence (upper solid line), narrow the gap between it and accumulated incidence, and improve resemblance to Fig. 3.

represent an operational situation rather than to postulate an explicit pathological mechanism, and to express the possibility that in a practical cytology programme a proportion of cases could make such a leap without there being any real possibility of intervening.

The composite pattern of Fig. 4 is also subject to distortion, not least for the reason that the data belong to different communities. The width of the interval between the total prevalence of cervical pathology and the curve of accumulated onsets of carcinoma *in-situ*

is uncertain; we can only say that it appears to be a wide gap and probably exists in fact. Alternative explanations for this gap have already been discussed and it is not possible to draw precise conclusions from comparisons between the forms of the various curves, although the gaps at the tops of Figs. 3 and 4 suggest strongly that in real life we may be dealing with the dynamic rather than a simple progressive situation.

The model results represent a compromise between the usual formulations of the natural history and the constraints of available data. A comprehensive reconciliation was not in the event possible, but reasonable simulations were achieved with respect to the accumulated incidence and prevalence of pre-clinical disease, and total mortality. Within these boundaries the models were a little too generous in their proportions of survivors with cancer and ungenerous in their division of pre-clinical disease into *in-situ* lesions on the one hand, and pre-clinical invasive lesions on the other.

THE CRITERIA OF SUCCESS OF SCREENING PROGRAMMES

The progressive natural history generated 170 dysplasias, 262 *in-situ* carcinomas, 222 invasive carcinomas, and 89 deaths. The dynamic natural history generated 455 dysplasias, 309 *in-situ* carcinomas, 265 invasive carcinomas, and 96 deaths. In each case the deaths are a subset of the invasive carcinomas, and the invasive carcinomas represent all first progressions beyond the *in-situ* stage. In the progressive, but not in the dynamic natural history, all the invasive cancers started as '*in-situ*', but in the dynamic natural history only 131 of the invasive cancers passed through the *in-situ* stage, so that the total initial onsets of any form of carcinoma came to 443, compared with the 262 of the progressive model.

Against these alternative backgrounds it would not be appropriate to compare the performance of different screening programmes in terms of prevention or detection of dysplasias or carcinomas *in-situ*. This applies in real life as well as in simulation and there is a need for an objective representing a true output. The only proper comparisons are those based upon the prevention of more advanced states, or of necessary treatments, or of deaths, or of various combinations. Age at death would be a contributory factor in choosing between programmes giving equal absolute savings.

A series of simulation experiments showed that the ratios between alternative criteria were inconstant in different environments. In

every case it was found that the screening programmes saved more deaths than invasions and a great part of their effect was through the detection and saving of early invasive disease. However, the ratio between deaths saved and invasions saved varied with the age at which a test was offered, with the natural history premises, and with the number of tests offered per woman. A selection of results is given in Table 1. The conclusion from these experiments must be that no single criterion will necessarily prove adequate for all circumstances, and where alternative programmes show marginal differences in results on the basis of one criterion, then other criteria must be examined. However, for an initial exploration, the most satisfactory index of success is probably the number of deaths saved, expressed as a percentage of deaths occurring under background conditions.

Table 1. *Ratio of deaths saved to invasive cancers saved, in different circumstances.*

	Progressive natural history	*Dynamic natural history*
20 tests per woman	1·08	1·67
10 tests per woman	1·20	1·70
5 tests per woman	1·40	1·73
Single test at 25 years	1·42	1·31
Single test at 35 years	1·59	1·51
Single test at 45 years	1·72	1·71
Single test at 55 years	1·63	1·46

In the context of considering alternative redeployments of a fixed resource, numbers of deaths saved can also be expressed in relation to the numbers of tests carried out, that is as 'numbers of tests/deaths saved'. Where unnecessary biopsy is seen as a harmful result to be minimized, the ratio 'positive tests/deaths saved' can also be used. This, however, has the disadvantage of depending upon an arbitrarily attached false positive rate and must be used with circumspection.

EFFECTS OF SINGLE TESTS

Single tests were simulated from age 25 to age 65 at five-year intervals, the series being repeated for each natural history. Results, in terms of numbers of positives per 1,000 tests, and in terms of percentages of deaths saved, are given in Table 2. The circumstances of these simulations, which assume total attendance on the exact birthday, are highly unrealistic, but the results serve to show that at most ages the progressive natural history is more susceptible to intervention than is the

dynamic natural history, but that for either natural history the most fruitful deployment of single tests would be at about 40 years.

Table 2. *Results of single tests: yields per 1,000 tests and percentages of deaths saved.*

	Age at test								
Natural history	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*
Progressive									
Yield per 1,000	3·7	6·2	8·0	9·0	9·4	9·3	9·1	8·9	8·4
Deaths saved (percentage)	9·0	18·0	23·6	25·8	24·7	22·5	19·1	15·7	12·4
Dynamic									
Yield per 1,000	3·8	5·3	7·0	8·0	8·6	8·5	8·0	7·6	7·2
Deaths saved (percentage)	9·4	14·6	17·7	19·8	18·8	16·7	12·5	10·4	7·3

Alternative criteria confirm the general pattern and Table 3 displays results expressed in terms of 'number of tests/deaths saved' and 'number of tests/invasive carcinomas saved'.

Table 3. *Results of single tests: tests performed per deaths saved, and per invasive cancer saved.*

	Age at test								
Natural history	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*
Progressive									
Per death saved	1,218	602	444	390	383	394	440	499	587
Per cancer saved	696	356	284	254	263	272	287	304	322
Dynamic									
Per death saved	1,083	688	550	475	471	497	628	707	933
Per cancer saved	513	317	302	282	292	331	419	505	726

The ratio 'positive tests/deaths saved' is of interest from the patient's point of view in that it relates the risks of unnecessary and necessary biopsy; variations according to age and natural history premises are given in Table 4. Pathological findings at simulated biopsy reflect the arbitrary assignation of a false positive rate of 0·1 per cent in normals, and it is necessary to consider the consequences of errors in this assignation. For example, the single test at age 40 with a progressive natural history yielded 81 'positives' of which 9 were normal, 16 dysplasias, 35 carcinomas *in-situ*, 14 occult, and 7 clinical invasive disease. Easson (1973) quotes a distribution of positives containing 63 normals, 174 dysplasias, 1,005 carcinomas *in-situ*, and 266 invasive lesions. Table 4 therefore supplies an alternative ratio B in which the numerator contains only half of the normals and the dysplasias detected in the model runs. The ratios are generally lower and the lowest

ratios occur at somewhat younger ages than for the initial ratio, but the essential point from this table is that reasonable ratios, between 3:1 and 4:1 are obtainable for either natural history at any age between 30 and 50, and with false positive rates ranging between those of the simulation and those quoted by Easson (2). The ratio 'all cancers detected/deaths saved' is given as ratio *C* of Table 4.

Table 4. *Results of single tests: 'positives' per death saved.*

	Age at test								
Natural history	*25*	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*
Progressive									
A	4·5	3·8	3·6	3·5	3·6	3·6	4·0	4·4	4·9
B	3·1	2·8	2·9	3·0	3·1	3·2	3·5	3·9	4·3
C	1·6	1·9	2·3	2·4	2·6	2·6	2·9	3·3	3·7
Dynamic									
A	4·1	3·6	3·8	3·8	4·1	4·2	5·0	5·0	6·7
B	3·0	2·8	3·1	3·2	3·5	3·7	4·3	4·6	5·7
C	1·9	2·0	2·4	2·6	2·9	3·1	3·6	3·8	4·7

KEY

Ratio A is computed as (coned normal+coned dysplasia+coned carcinoma*in-situ*+coned occult+clinical)/deaths saved.

Ratio B is computed similarly but only half of 'coned normal+coned dysplasia' is included in the numerator.

Ratio C excludes coned normal and coned dysplasia from the numerator and refers to (all carcinomas detected)/deaths saved.

The effect of different ages-at-test upon the ages at which deaths are saved is illustrated in Table 5. A single test at age 30 is compared with a single test at age 50, and with no test at all. The test at age 50 produced a greater over-all accumulated reduction in mortality but

Table 5. *Results of single tests: age at test and age at death. Accumulated deaths from cancer of the cervix per 10,000 females born, for the progressive natural history.*

	Age at test											
Test	*30*	*35*	*40*	*45*	*50*	*55*	*60*	*65*	*70*	*75*	*80*	*85*
None	0·2	1·7	6·2	14·5	25·5	37·5	49·1	59·7	68·9	76·6	82·5	86·2
Age 30	0·2	1·0	2·5	6·2	13·4	23·2	34·0	44·3	53·4	61·1	67·0	70·8
Age 50	0·2	1·7	6·2	14·5	25·5	35·2	41·1	45·7	51·0	57·2	62·6	66·3

the single test at age 30 produced its effect earlier and succeeded in saving half of the deaths which would have occurred before the age of 50. In terms of life-years saved, or other age-weighted valuations of life, the test at age 30 would probably show advantages.

The findings of the simulated single tests, taken together, confirm

the prior expectation that on all criteria the dynamic natural history offers a less tractable situation than does the progressive one, but that for both natural histories useful results can be expected in women submitting themselves to single tests anywhere in the age range 30–50. A single test between these ages should be capable of reducing mortality by 15–25 per cent among the women who attend, provided that the true natural history is somewhere in the range presented by the two model processes explored.

RESULTS OF MULTIPLE TESTS

Results so far presented suggest that multiple tests might best be deployed at ages where yields of positives and the percentage death salvage rates are greatest. However, the pressure to concentrate multiple tests in high-yield, high-result areas is countered by a need to separate tests so that the effectiveness of the second of a pair is not completely overshadowed by the cleaning-up effect of the test before it. Optimal intervals between tests will depend both upon the natural history and upon the negative error rate of the procedure used and optimal intervals between offers of tests will depend to some extent upon the percentage uptake.

The problems of numbers, timing, and spacing was first explored, for each of the natural histories, by simulating 20 or 10 or 5 tests over the age-range 19–76 years. Thus, the most concentrated regime conducted tests for ages 19–76 by three-year intervals; the next conducted tests from ages 22 to 76 by six-year intervals: the last offered tests from ages 25 to 73 by twelve-year intervals. The results were related in turn to the effects of optimally placed single tests at age 40. As with the single test experiments it was first assumed, unrealistically, that every woman attended. Results are given in Table 6 in terms of percentages of deaths saved and the number of tests carried out per death saved. This table also displays the marginal costs of each level of investment, compared with the lesser level, in terms of additional tests per additional death saved and in terms of additional positive tests per additional death saved.

It can be seen that at these levels of decision-making, involving questions of the escalation or redeployment of resources, the complexities and uncertainties of intuitive prediction begin to mount alarmingly, even without considering the uncertainties of test sensitivities or of the response of patients to offers. The effectiveness of different regimes, in terms of percentages of deaths saved, differ

according to the natural history in an inconstant manner. For a single test the progressive natural history was 30 per cent more tractable than the dynamic but this rose to 47 per cent when five tests were employed and fell again to 39 per cent and 30 per cent for ten and twenty tests respectively.

Table 6. *Results of multiple tests: percentages of deaths saved and tests per deaths saved.*

	Number of tests			
Natural history	*1 (age 40)*	*5*	*10*	*20*
Progressive				
Deaths saved (percentage)	25·8	67·4	88·8	96·6
Tests per death saved	390	648	979	1,818
Marginal tests per extra death saved	390	809	2,024	11,289
Marginal positive tests per extra death saved	3·5	4·6	5·8	16·7
Dynamic				
Deaths saved (percentage)	19·8	45·8	63·5	74·0
Tests per death saved	474	888	1,274	2,212
Marginal tests per extra death saved	474	1,202	2,272	7,937
Marginal positive tests per extra death saved	3·8	5·4	5·5	11·1

From the point of view of the practical situation in Great Britain, the important values are at the intermediate levels of investment, and this is where expectation of benefits in relation to costs seem to depend most importantly on what the natural history may be. These are also the levels where marginal improvements in deaths saved are bought at the cost of rapidly increasing marginal costs in terms of extra tests provided. For the dynamic natural history an arbitrary marginal cost of 1,000 additional tests per additional death saved, is passed at about the fourth test with a total death saving of about 40 per cent; for the progressive natural history the same level is deferred until about the seventh test and with a total death saving better than 70 per cent. Marginal costs to the patient in terms of 'positive tests' per life saved, and in terms of unnecessary biopsies, follow a different curve; very rapid increases in marginal costs do not occur until much higher investments are made.

It can be seen that ignorance of the exact natural history, of some importance when investment is minimal, becomes crucially important as investment rises. These uncertainties constitute a serious disability at the point where the consequences of erroneous decision

really begin to matter. The justification for expenditure upon research to narrow these uncertainties must also arise as the service develops. This is contrary to some traditional and even some prevailing philosophies which tend to separate research from service sequentially as well as in budgetary terms. This progressive pattern of the costs of ignorance, and its consequences in relation to research investment probably extend far beyond the field of planning cervical cytology screening programmes.

RESULTS OF VARYING ACCEPTABILITY OF TESTS

Offers of screening in real life are not uniformally taken up. At one conceptual extreme a population may be divided into attenders and non-attenders. The total salvage is then the mean of the salvage figures of the two groups (non-salvage in one case) weighted according to the relative size of the groups. If an unbiased 60 per cent of the population attends a five-test course and gives a yield of 67·4 per cent deaths saved within that group, then the total population result amounts to 40·4 per cent of deaths saved (67·4 × 0·6).

Table 7. *Effects of varying acceptability of tests: percentages of deaths saved.*

	Percentage attendance (unweighted)					
Natural history	*100*	*90*	*80*	*60*	*40*	*
Progressive						
20 tests	96·6	95·5	93·3	86·5	71·9	(38·6)
10 tests	88·8	84·3	79·8	66·3	49·4	(35·5)
5 tests	67·4	61·8	55·1	42·7	29·2	(26·9)
Dynamic						
20 tests	74·0	71·9	69·8	62·5	52·1	(29·6)
10 tests	63·5	60·4	56·2	46·9	35·4	(25·4)
5 tests	45·8	42·7	38·5	30·1	21·9	(18·3)

* This column of figures is computed as 40 per cent of the death salvage of the 100 per cent attendance column.

At the other conceptual extreme, the population is uniform in its behaviour, but presents for examination with a fixed probability of less than 100 per cent. A series of model experiments was designed to explore this situation and the results are given in Table 7. Attendance rates of 100, 90, 80, 60, and 40 per cent were examined. The results show that for small investment (up to five tests) the salvage fell almost in proportion to the falling attendance rate, so that all the population attending on 40 per cent of occasions, gave only a

little greater salvage than 40 per cent of the population attending on all occasions. However, with higher investments (ten or twenty offers) the effect did not fall as steeply as the attendance rate. For the progressive natural history and a twenty-test offer the salvage for all the population attending 40 per cent of the time, was 1·86 times better than for 40 per cent of the population attending all sessions. For the dynamic natural history the factor was 1·76. With small investments, evidently, the pattern of deployment is not of any great importance, but as the provision of service increases there are convincing grounds for spending resources to encourage the systematic deployment of testing facilities across the entire population.

Table 8. *Effects of varying acceptability of tests: percentages of deaths saved.*

	Percentage attendance (weighted)					
Natural history	*100*	*90*	*80*	*60*	*40*	*
Progressive						
20 tests	93·3	89·9	87·6	77·5	62·9	(75·2)
10 tests	72·5	73·0	68·5	56·2	40·4	(31·0)
5 tests	53·9	49·4	44·9	33·7	23·6	(21·6)
Dynamic						
20 tests	68·8	65·6	63·5	56·2	44·8	(27·5)
10 tests	55·2	52·1	47·9	39·6	29·2	(22·1)
5 tests	37·5	34·4	31·2	25·0	17·7	(15·0)

* This column of figures is computed as 40 per cent of the death salvage of the 100 per cent attendance column.

All the model investigations described so far have assumed uniform attendances according to age and to pathological state. An investigation was next carried out to examine the effects of weighting attendances according to the pathological state, in sympathy with a situation known to occur in real life. Basic attendances were weighted by minus 10 per cent for dysplasia, minus 20 per cent for carcinoma *in-situ*, and minus 30 per cent for invasive cancer with respect to attendances in normal persons. Table 8 gives results of experiments carried out under these conditions in a manner otherwise comparable with Table 7. The effect of the weightings was a further substantial drop in the rate of death salvage accompanied only by a very small drop in the numbers of tests performed (less than 0·1 per cent) and a consequent rise in the cost of saving a death. The marginal costs of escalation, under these conditions, also rise steeply.

It is found in practice that women invited to be examined attend with a frequency which diminishes with age, and that the opportunities for adding a cytology test to attendances associated with pregnancy or with family planning consultations also diminish with age. In an attempt to explore the general consequences of this decline, attendance rates were assigned to different ages in a linearly descending manner according to the formula: percentage attendance = 110 − age. Thus the expected attendance rate would be 90 per cent at age 20, and 40 per cent at age 70. This gradient was applied to the five-, ten-, and twenty- test offers, with and without the loadings according to pathological state. Over-all uptakes were in the range of 64–66 per cent and the results are displayed in Table 9.

Table 9. *Effects of varying acceptability of tests: Percentage of deaths saved and tests per death saved in presence of linear age graduation of acceptance.*

	Number of tests		
	5	*10*	*20*
WITHOUT ATTENDANCE WEIGHTING			
Progressive natural history			
Deaths saved (percentage)	43·8	68·5	86·5
Tests per death saved	649	822	1,345
Dynamic natural history			
Deaths saved (percentage)	31·2	49·0	62·5
Tests per death saved	847	1,070	1,732
WITH ATTENDANCE WEIGHTING			
Progressive natural history			
Deaths saved (percentage)	34·8	57·3	78·6
Tests per death saved	816	982	1,479
Dynamic natural history			
Deaths saved (percentage)	26·0	41·7	56·2
Tests per death saved	1,016	1,257	1,924

These results demonstrate that the combination of an age-dependent declining attendance, a selective loading against the attendance of women at special risk, and the contingency of a dynamic rather than a progressive natural history, set serious constraints upon levels of anticipated effectiveness. A simulated ten-test offer (a large investment despite the costing 'benefits' of the incomplete uptake) did not in these conditions reduce mortality by more than 42 per cent. Even this figure is optimistic to the extent that it supposes uniform behaviour within the attendance rules laid down and not, as is probably the case in real life, a population consisting partly of women who will not attend in any circumstance and others who will attend more often

than average. The salvage rate of 42 per cent (last line in Table 9) was achieved at the high cost of 1,257 tests per death saved, and the marginal cost of increasing to a ten-test from a five-test offer was 1,659 additional tests per additional death saved. An increase from a ten-test to a twenty-test offer would cost 3,831 additional tests per additional death saved and would still not raise the death salvage rate beyond 56·2 per cent.

Costs of this order are very different from those of a single test uniformly applied at an optimal age against the backgrounds of a supposedly progressive natural history. The cost of ambitious escalations of cytology programmes could prove quite incommensurate with the additional benefits obtained. The results of these experiments offer a salutory reminder of the unfortunate operational consequences of accumulated unfavourable contingencies and of the need to use realistic models for predicting the results of real-life situations.

Conclusions

The main conclusions of this simulation exercise are:

1. In the presence of a rudimentary cervical cytology programme the consequences of ignorance concerning the natural history (especially) and concerning the probable patterns of attendance, are not too serious. Almost any kind of deployment of small resources will take the easy pickings at modest cost, the first 10 per cent, say, of the mortality from carcinoma of cervix. The best deployment is around age 40 whether in terms of the death salvage rate, or the number of tests needed to save a death, or the ratio of biopsies to deaths saved.

2. As investment increases, each increment of result is more expensive than the last. Moreover, the cost of ignorance mounts to a point where different appreciations of natural history produce widely discrepant estimates of the deaths likely to be saved and the costs of saving them. The marginal costs of any increased salvage mount rapidly between the limits of five-test and ten-test offers and the discrepancies between estimates based on different natural history appreciations are particularly wide at this point. Proposals to extend the services once they have passed the five-test level justify substantial expenditures upon research to determine more accurately the true natural history of the disease. In operational terms increments at these levels also justify substantial expenditures aimed at systematiz-

ing the deployment of the available resources and their delivery to women who would not otherwise attend.

3. The mounting costs of each increment of salvage will sooner or later limit the extent to which the disease can be controlled. The difficulties of achieving high levels of control have probably been underestimated in the past. Sooner or later a point will be reached where the resources are better spent elsewhere.

The scale of increasing cost, and of diminishing benefit for each increment of cost, is continuous, and it must be clearly recognized that there is no natural limit to expenditure based upon the biology of the disease. More investment will always continue to generate larger results in terms of deaths and cancers saved, but arguments for deploying the resources to alternative preventive and curative procedures will strengthen at an even greater rate. Decisions about appropriate expenditure limits cannot possibly be determined except in the light of the costings, anticipated benefits, and feasibilities of alternative usages and in the light of progressive refinement of the benefit predictions of cytology programmes, and ultimately of results.

REFERENCES

1. FIDLER, H. K., BOYES, D. A., and WORTH, A J. (1968). 'Cervical cancer detections in British Columbia', *J. Obstet. Gynaec. Br. Commonw.* **75**, 392–404.
2. EASSON, E. C. (ed.) (1973). *Cancer of the Uterine Cervix* (London: Saunders).

3

Towards the evaluation of more options for management

An approach to resource allocation in the reorganized National Health Service

J. R. Ashford

MA, Ph D, FBCS

Professor of Statistics and Joint Director of the Institute of Biometry and Community Medicine and the Exeter Community Health Services Computer Project

G. Ferster

MS, Ph D

Head of Economics Group, Numerical Sciences Division, Institute of Biometry and Community Medicine

Diane M. Makuc

MSPH

Biostatistician, Numerical Sciences Division, Institute of Biometry and Community Medicine

An approach to resource allocation in the reorganized National Health Service

Summary. *This essay is concerned with resource allocation, which must be a main concern of management in the reorganized NHS. The basic approach involves the modification and extension of the mathematical programming techniques which have been used with success in industry and commerce during recent years to cope with the particular problems of the health services. Concepts are developed concerning the classification of the population served by the health care system, the sectors of the health services, the types of case dealt with by particular sectors, the alternative procedures available for particular types of case, the associated benefits and costs, and the resources required. It is suggested that the main opportunity for the exercise of the management function will arise in the situation in which alternative procedures involving different inputs of resources are available to provide care for the same type of case. In this context, the benefits and costs will both depend upon the proportion of cases dealt with by each alternative and it is suggested that a policy should be adopted which leads to the greatest possible benefit consistent with the constraints upon resources, case mix, and costs to which the system is subject. The assessment of the benefits of health care is discussed and it is suggested that three main types can be distinguished, relating respectively to the direct effect on the health of the patient, the provision of validation services and the indirect social advantages. Three main contributions to the costs of health care are identified corresponding to the health care system proper, the patient and his family, and other public or private agencies. The need to consider not only over-all benefits and costs but also the relative*

contributions of each component is emphasized. The application of the proposed methods is illustrated by an example from the field of maternity care

INTRODUCTION

The National Health Service was established more than twenty years ago, with the object of providing a universal and comprehensive health care system in the United Kingdom. The original form of organization was tripartite, with separate branches responsible for the provision of the family doctor, hospital, and public health services. This structure reflects the general tradition of medical practice in the UK and emerged at the inception of the NHS as the fruit of a compromise between the various professional and public interests involved. In the light of two decades of practical experience it is generally accepted that the performance of the system has been broadly satisfactory. During recent years, however, a very rapid escalation of expenditure, resulting largely from rising public expectations and the growth of the 'scientific' element in medical practice, has emphasized the need to review the structure of the system in the light of current conditions. The original and perhaps somewhat idealistic emphasis upon broad principles and social equality has given way to a more practical concern about the efficient and effective use of scarce resources. The main shortcomings of the existing system have been identified as being associated with duplication and poor co-ordination between the three branches of the NHS and with the absence of a modern and quantitative approach to management at all levels (6).

After much public discussion extending over several years, the decision has been taken to restructure the NHS and a new form of organization will be set up in March 1974. The stated general aims (8) are 'first, that there should be a fully integrated Health Service in which every aspect of health care can be provided by members of the health care professions and second, that this care should be provided as far as possible locally and with due regard to the health needs of the community as a whole'. Within this new framework, multidisciplinary teams will be responsible for management of personal health services at the district, area, regional, and national levels of organization which are to be established, although the associated personal social services will be managed separately as the responsibility of local

authority social service departments. At each such level one of the main tasks of management will be to allocate resources to needs. This is the kernel of the management function in a wide variety of situations. Although some aspects of the problem in the NHS are similar to some aspects of the problem in commerce and industry, other and important facets of resource allocation in health care are unique to this particular field. For this reason any useful approach to resource allocation in the health services must involve the modification and development of the techniques which are well established in other areas of activity. The present paper presents a view of this problem and is concerned with the construction of an appropriate conceptual framework, as well as with the subsequent computational procedures. The practical application of these ideas is illustrated by an example relating to one major sector of the health services at the district level.

Some features of the reorganized NHS

A modern system of medical care is a complex undertaking. A wide variety of human and physical resources is available to the human population which they serve. Under the new management arrangements, the basic operational unit of organization in the NHS will be the district. A typical district will cover a population of about 250,000, sharing a common set of health service resources which will be managed as a coherent whole. At the primary level of care there will be about a hundred general medical practitioners, assisted by a team of health visitors and home nurses, and about fifty general dental practitioners. Hospital services for the district will be provided under the aegis of some fifty consultant specialists, practising from and using the resources of the hospitals in the district and assisted by about three times their number of more junior medical staff. Each district will include at least one hospital at which care is given for 'acute' medical and surgical cases, together with hospitals for the mentally ill and mentally handicapped and for providing less intensive medical and surgical care. At any one time some 550 persons will require acute hospital in-patient care, some 600 will require in-patient care for mental illness and 300 will require in-patient care for mental handicaps. Considerably larger numbers of patients will receive hospital out-patient services or domiciliary care by hospital doctors. There will be some 1,500 hospital nurses and about the same number

of hospital ancillary staff. Laboratory and other professional services will be provided by a staff of about one hundred scientists and technicians.

For management purposes districts will be grouped in areas, whose boundaries are to be coterminous with those of the new local authorities which are to be set up in 1974. The area health authority (AHA) is to be the lowest level of statutory authority, with full planning and operational responsibilities. Areas will be grouped in regions and the AHA will be accountable to the regional health authority (RHA) for planning services, establishing priorities, and the allocation of resources to operational services. Special arrangements will be made to cover medical and dental teaching carried out in universities. In England the highest level of authority will be the Department of Health and Social Security, which will be responsible for establishing national policies and priorities and for determining the kind, scale, and balance of services to be provided by the NHS.

The management of a health care facility is subject to a number of important and characteristic constraints which do not apply to most other organizations of similar size and complexity. In the first place, the system exists to serve individual patients, either by attempting to provide a 'cure' for a 'disease' or by attempting to ameliorate the effect of symptoms or conditions from which the patient is suffering. Thus the performance of the system must be judged solely in terms of its contribution to the improvement of the health and well-being of many individual patients. This contribution is the only relevant measure of the output of the system and must be clearly differentiated from the items of service provided, which may (or may not) lead to improvements in the health and well-being of the patients concerned. In this context it must be emphasized that there are at the present time very substantial difficulties in obtaining reliable estimates of the effects on health of particular common medical procedures (5). The position is further complicated by the fact that medical care is only one of a very wide variety of factors which can contribute to better health (3). Even when the effect of health care on the individual patient is established, it is necessary to place some value on the resulting benefit to the patient and to society in general. No completely satisfactory answer is available. Fuchs (12) has suggested that three different kinds of output flow from health services: 'health', 'validation services', and 'other consumer services'. The first category is perhaps the most obvious, but nevertheless very difficult to assess in monetary

terms. 'Validation services' cover certification, for insurance or other purposes, whilst 'other consumer services' include the 'social' content of institutional care and for certain types of patient (such as those suffering from mental illness) may be the most important.

The position as regards the measurement of the input of resources to the system is in some respects more satisfactory, although three separate components of costs, associated directly with the health care system, with the patient's own contribution or with other public agencies, must be considered. In countries where the patient pays for health care by 'item of service', accounting procedures allow the health service resources consumed by an individual to be identified and costed. However, the NHS is financed largely from public funds and the existing accounting practices are not designed for this purpose. The calculation of costs or the assessment of resources utilized for an individual episode of illness may therefore present considerable practical difficulties. The lack of any coherent set of conventions for equating capital and revenue expenditures in the public sector of the economy is also an obstacle to the calculation of the consolidated costs to the health care system. Apart from these direct costs, access to the NHS must involve costs to the patient. Thus the need to take a day off work to see a doctor for ten minutes in a hospital out-patient clinic or to wait perhaps many months for 'non-emergency' surgery clearly affects both the patient and the national economy. Furthermore, in situations in which the contribution of the health care system to the social conditions of the patient is important, other public agencies (such as the local authority) or private charities may offer alternative forms of care. Different methods of providing care for the same condition may involve a change in the balance of costs between the health services, the patient, and other public or private agencies. It is customary to set objectives for management in business and commerce in terms of profit, as measured by the excess of the output or sales over the input or costs. In the field of health care the calculation of an equivalent quantity in a situation in which both input and output may be regarded as having three such disparate components presents manifest difficulties.

A third major feature of health care is that decisions about the use of resources are taken by doctors, who in general work as individuals and are not subject to the restraints of a hierarchical structure. This well-established tradition is embodied in the principle of 'full clinical freedom', which is regarded by the British medical profession as an

essential basis of good practice. According to this principle each doctor need take account only of the interests of his particular patient when he decides what form of care is required. In practice, however, most experienced doctors accept that critical resources are limited both by financial and other constraints and that patients must in effect compete for access to hospital beds, the services of nurses, or indeed for their own professional care. Thus, the basic decisions about the way resources are used are taken on the 'shop-floor', rather than in the 'board-room'. Management can exercise control only indirectly, by determining how much and what kind of resources should be made available and by setting ground rules to govern competition between different potential users. Even in this context the freedom of action of the managers is limited to the extent that, if adequate finances are available, many of the important resources, such as doctors or hospital beds, can be increased only marginally in the short term.

In spite of the prime role of the doctor in the allocation of resources and the existence of free exchange of information within the medical profession, evidence from many parts of the world suggests that there may be very substantial variations in the pattern of care which particular doctors are prepared to recommend or to accept for patients suffering from the same condition. In the UK these variations are reflected in systematic differences in the use of resources between different parts of the country. For example, the average length of hospital in-patient stay in acute specialities in 1968 in England and Wales varied from 9·4 days in the Oxford region to 14·0 days in the Liverpool region (7). This suggests that, for a particular type of case, alternative methods of treatment, involving different inputs of resources or 'streams of care' (10) exist and are medically acceptable, although not necessarily regarded by all doctors as being equivalent in every respect. When patterns of resource utilization are compared, due account must be taken of the effect of differences in resource availability; facilities not available cannot be used, whilst facilities which are available tend to be used to full capacity, if no direct financial penalty is involved (16).

When the provision of medical care for a particular population is considered, it is necessary to differentiate clearly between terms such as 'demand', 'needs' ('perceived' or 'not perceived'), and 'usage'. In the context of a system such as the NHS, the classical interpretation in economic theory of the term 'demand' as involving both willing-

ness and ability (by the patient) to pay for the particular service is not applicable. Apart from the initial decision to consult a doctor, the patient is not normally qualified to exercise any judgement concerning the nature of his needs or the quality of the service provided and must be guided by his medical adviser. The concept of 'market forces' which underlies much conventional economic theory does not apply. Even when 'willingness to pay' is essential to obtain access to health care (and this is not the case in a system such as the NHS), it would be unwise to equate the market price of medical care with the resulting benefit. Nor would it be reasonable to assume that the health care system is managed in such a way as to maximize the profits of the undertaking, as represented by the difference between income and expenditure. The traditionally used Pareto criterion in economic welfare may not be appropriate. The concept of 'need' is imprecise and involves a continuum, rather than a clearcut threshold in the context of medical care. The identification of need and the provision of resources to meet need is a subtle problem, to the solution of which professional knowledge must be coupled with the stimulus which, in the modern world, is provided by pressure groups. This is an issue of which each level of health services management must take account, with appropriate guidance from higher levels and with the specific object of providing resources to meet needs. From the point of view of day-to-day management, arguments about the precise definitions of need are academic, since the system must cope with demands, which will reflect not only the state of health of the patients, but also public and professional attitudes to health care.

The structure of a typical health care system recognizes the diversity of the various tasks which must be performed, in the sense that the services can be divided into a number of 'sectors'. Examples are the services for maternity, children, the elderly, the mentally ill, and the mentally handicapped. Each such sector has a characteristic function to perform, but in general is not completely self-contained, in the sense that the same resources can be in theory (and often are in practice) shared between different sectors. Within any one sector there are, of course, a variety of different 'types' of case and for each such type there may be alternative streams of care.

The use made of different sectors of the health services varies widely within a population, depending upon factors such as age, sex, social class, and environment (1). Thus, a population may be divided on the basis of personal characteristics and attributes into 'classes', in

such a way that variation in the usage of particular sectors within a class is small in comparison with variation between classes. The existence of a suitable system of classification and a knowledge of the number of persons belonging to each class enables the numbers of cases requiring care from particular sectors of the system to be estimated (2).

Concepts and model

As in all fields of application in which quantitative methods of management are applied, the first step towards the understanding of the real-life situation is the development of a set of basic concepts which represent the essential features. This will generally involve a process of simplification and approximation, but is a necessary preliminary to the construction of a mathematical model to embody the basic relationships between the concepts. If the concepts and the model are appropriate, the effects of departures from the existing conditions may be estimated by the use of appropriate mathematical techniques. In this way the results of particular decisions may be assessed and light may be thrown on the best means of achieving objectives.

Any well-directed health care system should be centred on the patient and the basic concept is that of the *population* of patients. Members of a population may differ in many respects, including their use of the health care system. Thus, a population containing n patients is divided into r *classes*, $C_1, C_2, \ldots, C_r$ on the basis of attributes associated with health service use and we suppose that class C_h contains n_h subjects, where

$$\sum_{h=1}^{r} n_h = n.$$

Care is provided by health services which are organized in s *sectors*, $S_1, S_2, \ldots, S_s$, each of which covers a particular general type of care. In general, a subject from any class may receive care from any sector. Patients requiring care present themselves to sectors as *cases*. During any given period of time a single patient may receive care from several sectors, or several times from a single sector. The classes are chosen in such a way that variations within classes in the usage of particular sectors are small in comparison with variations between classes. Knowledge of the numbers n_h within each class, coupled with information about the usage of different sectors by patients from the var-

ious classes then enables the total number of cases, M_i, referred to any particular sector, S_i ($i = 1, 2,..., s$), in any defined period of time to be predicted. The quantities M_i will depend upon the size of the population, the length of the period, the classification system, the number, n_h, in each class, C_h, and the prevailing pattern of health services usage and will in general be random variables. For the purposes of this essay, we suppose that the variation of the distribution of M_i is negligible and that we can effectively regard M_i as m_i, a fixed number.

The cases dealt with by a particular sector, S_i, may be subdivided into a_i different *types*, T_{i1}, $T_{i2},..., T_{ia_i}$, depending upon nature and severity. If the proportion of cases of type T_{ij} is denoted P_{ij}, the number of cases of this type in the given time period will be $r_{ij} = m_i p_{ij}$. We suppose that for a typical type of case, T_{ij} ($i = 1, 2,..., s$; $j = 1, 2,..., a_i$), there exist q_{ij} ($q_{ij} = 1, 2,...$) *alternative procedures*, $P_{ij}^{(l)}$ ($l = 1, 2,..., q_{ij}$), for providing care of a medically acceptable standard. Each such alternative (or the single acceptable procedure if there is only one) calls for a characteristic set of *inputs of resources* or *streams of care*. The resources available to the system are denoted R_1, $R_2,..., R_u$. In general, this list must include all resources provided by the NHS from the three existing branches (such as doctor time, nurse time, hospital beds, laboratory tests), together with resources provided by public services outside the NHS (such as social work), and resources provided by the patient and his family (such as the costs of travelling to obtain access to care and facilities for domiciliary care). The number of units of resource R_k required for a case of type T_{ij} under procedure $P_{ij}^{(l)}$ is denoted $X_{ijk}^{(l)}$. The quantities, $X_{ijk}^{(l)}$, are strictly random variables but we suppose that the variation of the distribution of $X_{ij}^{(l)}$ is negligible and we can effectively take $X_{ijk}^{(l)}$ as $x_{ijk}^{(l)}$, a fixed number.

As far as the management of the medical care system is concerned the numbers of cases, r_{ij}, of type T_{ij} will depend solely on length of the time interval and the total size and characteristics of the population and are not subject to managerial control. However, for cases of a given type we suppose that alternative streams of care are available and (subject to the approval of the doctors concerned) there is a possibility of varying the numbers $r_{ij}^{(l)}$ of cases assigned to procedure $P_{ij}^{(l)}$, subject to the total

$$r_{ij} = \sum_{l=1}^{q_{ij}} r_{ij}^{(l)}$$

being fixed. The task of the managers is to allocate the numbers $r_{ij}^{(l)}$ in such a way that in some sense the 'best possible' use of resources is obtained.

The number of units, t_k, of resource R_k may be expressed as the sum over all sectors, types of case and procedures of the product of the numbers of cases, $r_{ij}^{(l)}$, and the resources per case, $x_{ijk}^{(l)}$ in the form,

$$t_k = \sum_{i=1}^{s} \sum_{j=1}^{a_i} \sum_{l=1}^{q_{ij}} r_{ij}^{(l)} x_{ijk}^{(l)}. \qquad [1]$$

However in any real-life system of care the resources available are limited. These constraints can be expressed in terms of limits between which t_k must lie. The constraint involving a maximum utilization reflects the fact that resources such as hospital beds and domiciliary nursing time have finite limit. The lower bound to resource utilization reflects two aspects of management. In the first place it may be decided as a matter of policy that the use of certain resources should not fall below a certain (and normally significant) level. For example, it may be decided that a new hospital must be fully used, although it might be more expensive to run and provide no better outcome for the patient than an existing older hospital. [This is a more common situation than perhaps some of the advocates of new hospital building may realize.] Secondly, there may be some critical threshold of utilization for a particular resource below which the existence of that resource is not viable. Thus a domiciliary midwife service with such a small staff that cover cannot be provided for staff holidays and sickness could not provide an effective service.

The input to the system is the aggregate of the resources consumed, which consists of the sum of the quantities, t_k, representing the numbers of units of each resource. It is clearly desirable to measure the consumption of each resource on a common basis and running or revenue cost is an obvious common denominator, due account being taken of the need to 'service' any associated capital charges. In order to make the necessary calculations the unit cost, c_k, of resource R_k must be determined. In general, it may be expected that c_k will depend in some way upon the over-all rate of utilization, t_k. A possible simple form of the relationship would involve an expression for the over-all cost of the resource as a linear function of the rate of utilization of the form,

$$c_k t_k = \alpha_k + \beta_k t_k, \qquad [2]$$

where α_k can be termed the 'fixed' cost of the resource and β_k the 'variable' cost per unit. If this is the case, the total cost of the system is the sum taken over all resources of the product of the number of units required and the cost per unit,

$$C = \sum_{k=1}^{u} c_k t_k = \sum_{k=1}^{u} (\alpha_k + \beta_k t_k)$$

$$= \sum_{k=1}^{u} \alpha_k + \sum_{k=1}^{u} \beta_k \left[\sum_{i=1}^{s} \sum_{j=1}^{a_i} \sum_{l=1}^{q_{ij}} r_{ij}^{(l)} x_{ijk}^{(l)} \right], \qquad [3]$$

a linear function of $r_{ij}^{(l)}$. However, in some situations the simple cost structure [2] may not apply and if this is so the total cost may not be expressible in this linear form. In the calculation of expression [3] for the total costs, it may be desirable to subdivide the resources into three groups, corresponding to the health services, the patient and his family, and other public or private agencies.

In order to derive a criterion for the choice of the numbers $r_{ij}^{(l)}$ it is necessary to embody the basic aims of the system in terms of an appropriate objective function. Ideally we may suppose that there exists a 'benefit', $b_{ij}^{(l)}$ enjoyed by a case of type T_{ij} receiving procedure $P_{ij}^{(l)}$. The benefit may involve three components, corresponding to improved health, validation services, and other consumer services. Provided a common set of standards is applied throughout the system the units of measurement of benefit are arbitrary. Within a given sector this means that any convenient and appropriate unit of measurement can be chosen. However, because of the diversity of activities and aims in different sectors the only realistic common standard for calculations involving the whole system is likely to be financial advantage. On the assumption that each case dealt with is regarded as equally important, the output of the system may be expressed as the total of all the benefits taken over all sectors, all types of cases, and all procedures.

$$B = \sum_{i=1}^{s} \sum_{j=1}^{a_i} \sum_{l=1}^{q_{ij}} r_{ij}^{(l)} b_{ij}^{(l)}, \qquad [4]$$

which is also a linear function of $r_{ij}^{(l)}$. Different numbers of cases, $r_{ij}^{(l)}$, in the various streams will lead to different values of B.

Application of the model

A rational basis for resource allocation in these circumstances would involve choosing the $r_{ij}^{(l)}$ in such a way as to maximize the total benefit

B, subject to defined constraints on the total costs C (or on the three separate components of cost), on the resource availability and on the total numbers of cases of each type in each sector. This is a problem in mathematical programming. If the objective function B and the constraints are linear functions of $r_{ij}^{(l)}$ the techniques of linear programming would be appropriate. An analysis of this kind could be applied at any level of aggregation between the lowest (district) and highest (national) levels for particular types of case, for particular sectors or for the whole system. However, in practice there are in the present state of knowledge various specific problems (some potentially very exacting), which may be summarized as follows:

1. The sectors of the health services must be defined. This is to some extent a matter of semantics and there is general agreement about the appropriate classification of the major part of the activity.

2. For each sector so defined the main types of case must be specified. This must involve a detailed consideration of the nature of the activity within the sector, with the participation of the doctors and other staff involved. It is likely that an effective classification would be operational rather than based upon a diagnosis assigned after the completion of the episode.

3. The numbers of cases, r_{ij}, of each type T_{ij} in the population concerned must be determined. In general, this information may be obtained by a retrospective analysis of the care provided for the same population during a previous period, updated if necessary to take account of known patterns of change. If past records are not available or are inadequate, estimates based upon information about population size and structure must be used in conjunction with observational studies in other similar populations.

4. For each type of case, T_{ij} the alternative procedures, $P_{ij}^{(l)}$, for providing an acceptable programme of care must be determined. This depends essentially on medical practice and is best undertaken with the participation of medical and other staff whose opinions and experience embrace the whole spectrum of possibilities. Having identified $P_{ij}^{(l)}$, the appropriate stream of care must be specified in terms of the inputs $x_{ijk}^{(l)}$, of all the resources, R_k, concerned. This information is best obtained by the analysis of representative case-histories, supplemented where required by work study analyses of particular situations.

[The above steps 1–4 are concerned with matters that are essentially the same throughout the whole health care system. An appropriate structure for any one general population is likely to be applicable to any other general population. Standardization within the whole system would be desirable, to provide a basis for monitoring and control and a framework for the provision of information about the performance of different parts.]

5. The resources $R_1, R_2, \ldots, R_u$ available to each particular sector of the health care system must be identified. Constraints on the use of each such resource must be determined. This will involve an operational analysis of all the constituents of the system and an assessment of possible alternative uses of particular resources (such as buildings or nursing time). Consideration must also be given to the resources under the management of the social services departments of local authorities and to those provided by the patients and their families.

6. The unit costs, c_k, of the various resources, R_k, under all viable levels of operation, t_k, must be determined. Information of this nature is not normally accessible directly from the standard NHS accounts. However, when taken in conjunction with historical records of work done, the standard accounts may provide a basis for the estimation of unit costs at the historical levels of utilization, but special costing studies may be required. If large changes in utilization are contemplated, a simulation of the operation of a particular facility or set of resources may also be necessary.

[Steps 5 and 6 are concerned with the characteristics of the particular part of the system under study. Variations exist in resource provision between different parts of the country and each part must be considered separately, although standardization of costing techniques would be desirable.]

7. The benefits, $b_{ij}^{(l)}$, to cases of type T_{ij} treated by each of the alternative procedures $P_{ij}^{(l)}$, must be identified. At the present time there is not sufficient knowledge about the relative effectiveness of even the most common health care procedures for this to be possible on a general basis. The first steps are most likely to be made in terms of comparisons of benefit within the same type of case and subsequently within the same sector of the health services. Comparisons between sectors are likely to introduce the most intractable problems. The well-known difficulty of dealing on an *a-priori* basis with competing

claims for hospital beds between consultant specialists in, say, general surgery and general medicine provides an excellent example. Moreover, recent studies of the effectiveness of current health care procedures (5) suggest that for some 'popular' treatments the benefit may not be a positive quantity, whatever scale of measurement is chosen.

In the current state of knowledge the use of an objective function of the form [4] must be regarded as speculative, if all sectors of the health service are to be considered simultaneously. Within particular sectors, however, special detailed studies of health care delivery have been carried out and offer the possibility of applying the approach, if only on an experimental basis. If the managers do not wish to grasp the nettle of allocating quantitative benefits to procedures, the objectives can be shifted from maximizing the benefit within a fixed set of costs to the more limited goal of minimizing the total cost (or one component of the cost) subject to defined 'quality of care' thresholds being achieved. This again is a mathematical programming problem, the object being to determine the values of $r_{ij}^{(t)}$ which lead to a minimum cost, as given by equation [3] and subject to the given constraints on resource utilization. Under this approach it must be tacitly assumed that each alternative procedure relating to a case of a given type is of an equally acceptable standard of effectiveness. To some extent the imposition of constraints on resource utilization or on the numbers of cases receiving particular streams of care might serve to take some account of the 'quality' issue. However, cost minimization will always be an unsatisfactory substitute for an approach in which the possibility that the benefits of alternative procedures may differ is acknowledged explicitly.

We have pointed out that resources relating to health care may come from other public services and from the patient. The question of whether to include all three contributions is likely to be of critical importance, particularly if cost minimization is the prime objective. The appropriate weight to be given to family costs involves an element of political judgement, but it is clear that the costs of publicly supported activities, such as social work, should not be ignored merely because they do not happen to be included within the health care system as at present constituted. The provision of care for the elderly or for the mentally ill are important sectors of the system for which substantial contributions from outside the NHS might be involved in

alternative policies. Considerations of this type underline some of the difficulties associated with the decision to separate the health and social services for managerial purposes. However the unification of the NHS under a single managerial structure will improve the flexibility of the disposal of resources that belong to the health services proper.

The mathematical programming approach to resource allocation may be used for a variety of purposes. In the first place the technique may be applied to the day-to-day use of an existing set of resources, at, say, the district or higher levels in the NHS. The optimization process, whether carried out with the object of benefit maximization or cost minimization, then serves to produce an operational policy in terms of the numbers of cases to be dealt with by the various alternative procedures. Decisions about individual cases must of course still be left to the doctor. In order to ensure that such individual decisions conform to the over-all operational policy close participation by medical staff is required. In addition to pointing to the optimal 'mix' of streams of care, the mathematical programming technique also permits the critical restraints to be identified and in the cost minimization context the 'shadow prices' of particular resources forming bottlenecks to be determined. The approach may also be used for planning purposes, in order to explore radical changes in policy and as a guide to the effects of new expenditure and to the results of particular developments of the health care system.

An example of maternity care in a district

As an illustration of our approach we consider the management of a single sector of the system, the maternity services, in a geographical area which will become a district of the NHS. The maternity sector is particularly suitable in this respect, since a wide range of policies for the provision of care is available and all three of the present branches of the health service are involved. The starting point of the study was a comprehensive analysis of the existing maternity care system, which has been described in detail elsewhere (10). The district is centred on a city with a population of about 100,000 and also includes a number of small country and seaside towns situated in a thinly populated hinterland. The total population is about 300,000 and the number of births taking place annually is about 3,700. The boundaries of the district lie between 10 and 30 miles from the city. At

the present time two local authorities are responsible for this area, of which one (city) covers the city and the second (county) covers a large rural region of which the remainder of the district is a part.

THE FACILITIES OF THE SYSTEM

The facilities of the maternity care system are summarized in Table 1. There is one consultant unit (CU) situated in the city, with just under 50 beds. About the same number of beds is available in general practitioner units (GPUs). There are 5 hospitals with GP maternity beds, of which the largest (with 24 beds) is situated in the city about one mile from the CU. The remaining GPUs are situated in small towns; two have 10 beds and two have 4 beds. Only one of the GPUs (city) is solely concerned with maternity care; the four remote GPUs and the CU are each part of a hospital with other acute or long-stay facilities. Both the city and county local authorities operate a domiciliary midwifery service, which in 1970 was responsible for about 13 per cent of all local deliveries. In the city the midwives are employed solely on maternity care, but in the county midwifery forms only part of the duties (on average about 25 per cent) of the district nurse midwives.

THE TYPES OF CASE AND THE ALTERNATIVE PROCEDURES

The first stage in the analysis of the existing situation was to develop an appropriate classification of the types of case with which the maternity sector must deal. Following consultations with medical and nursing staff and a study of hospital and local authority records of obstetric care during the calendar year 1970, the major factor to be identified was intervention during the delivery process. Mothers who were delivered without intervention rarely required the nursing care and other facilities that are available only in hospital for more than three days after delivery. By the same token, deliveries with intervention required additional facilities during the delivery process and post-natal care over a longer period. Within each of the two types of case there are, of course, variations in the medical requirements; for example, deliveries by Caesarean section called for a higher level of care than other types of intervention. However, the differences within the two classes were relatively small compared to differences between the classes. Further elaboration of the classification system would be possible, but for the present purposes the use of just

Table 1. *Facilities for maternity care in the district in 1970*

Place of confinement (procedure)	*Distance from CU*	*Type of hospital*	*No. of beds*		*Relevant complement of staff*		*No. of maternal deliveries*
			All	*Maternity*	*Medical*	*Para-medical*	
Consultant unit (CU)							
City	—	Acute and long-stay	257	49	3 consultants 2 other senior staff 2 junior staff	10 midwives 12 nurses 5 pupil midwives 8 auxiliaries	1,501
General practitioner unit (GPU)							
City (GPU 1)	1 mile	GP maternity	24	24		7 midwives 2 nurses 4 auxiliaries	911
Rural (GPU 2)	22 miles west	Acute	17	4		3 midwives 1 nurse 1 auxiliary	98
Rural (GPU 3)	14 miles north	Acute	50	10		6 midwives 1 nurse 2 auxiliaries	356
Rural (GPU 4)	17 miles east	Mainly long-stay	119	10		7 midwives 1 nurse 2 auxiliaries	270
Rural (GPU 5)	16 miles south-east	Acute	34	4		3 midwives 3 auxiliaries	90
	All GPUs		244	52			1,725
Domiciliary (local authority)							
City	LA		—	—		9 midwives 6 pupil midwives	213
County	LA		—	—		11 midwives (equivalent)	254
	Both LA		—	—			467
Total confinements							3,693

two classes (with and without intervention at delivery) is considered adequate.

A wide range of alternative procedures for maternity care is available. In almost every case, medical care starts with the confirmation of pregnancy by the GP, who decides on the basis of medical and obstetric history and social circumstances whether to refer the mother to a consultant obstetrician. For mothers not referred to a consultant, a decision is taken about whether the confinement should take place in the GPU or at home, an appropriate booking is made and a programme of antenatal care is organized. Mothers referred to a hospital are booked for delivery in the CU at the discretion of the consultant. Periodic antenatal assessment by GPs and midwives brings to attention medical or obstretric problems and patients developing complications are referred to the CU. About one-quarter of the beds at the CU are assigned to antenatal care and expectant mothers spend variable periods of rest and treatment in hospital before labour begins. As a general rule, delivery takes place within 24 hours of the time of admission to hospital of a mother in labour. The period spent in hospital after delivery is variable, depending upon the medical and obstetric situation, the availability of beds and other factors. If necessary, post-natal care for patients discharged from hospital is provided by the domiciliary midwives. Under the existing system, changes in the place of booking are often made during the course of the pregnancy. However, transfers during labour from home or the remote GPUs to the CU are rare.

From this description, it is apparent that the patterns of maternity care can be classified by the varying inputs of resources provided throughout the period of pregnancy and terminating at an arbitrary time (taken in this analysis as six weeks) after delivery. Ferster and Pethybridge (10) have shown that the set of inputs which an expectant mother receives can be identified as a 'stream of care', and that this concept is sufficiently robust to accommodate virtually all pregnancies. Thus, the resources allocated to a particular stream can be identified in terms of physical units and as a result of special costing studies can be translated into monetary terms. By the same token, the benefits associated with particular inputs of resources can be assessed. For the purposes of this study the various streams may be classified on the basis of place of delivery in terms of eight procedures; home (city), home (county), GPU 1, GPU2, GPU 3, GPU 4, GPU 5, and CU. Transfers from one procedure to another during pregnancy have

been ignored for this purpose. The problem of resource allocation consists essentially of choosing appropriate numbers of patients with and without intervention at delivery to be assigned to each of the eight procedures, subject to given total numbers of cases of each type. Each possible combination will make a particular set of demands on the available resources and will be associated with particular sets of costs and particular sets of benefits.

It is a characteristic feature of the process of delivery that positive information about whether or not intervention will be required is available only after delivery has been completed. Although there has been some success in identifying patients who may be 'at risk', no 'booking policy' can possibly achieve absolute certainty in this respect. In an ideal situation as regards the provision of care for cases requiring intervention, all such cases would be delivered in the CU. To achieve this objective it would be necessary to ensure that all deliveries, including the majority for which no intervention is required should also take place at the CU. Even if this situation were considered to be desirable, the resources currently available are inadequate to cope with all deliveries. Thus, in terms of the general formulation of the resource allocation model given above, it is realistic to assert that intervention may take place for mothers delivered other than at the CU and each procedure covers both types of case.

When the effect on costs and benefits of varying the numbers of cases dealt with by the various procedures is considered, some assumption must be made about the corresponding proportions of interventions. For the purposes of the present analysis we suppose that the 1970 figures will apply. This is probably a simplification of the real-life situation, since it is likely that the proportion of interventions for deliveries under any particular procedure will vary at least to some extent in terms of the proportion of cases dealt with by that procedure. Furthermore, improved knowledge and a more efficient use of the information available are likely to improve the accuracy of the prediction of the need for intervention.

CRITICAL RESOURCES AND CONSTRAINTS ON THEIR USE

As part of the study of the 1970 maternity care system, an analysis was made of the resources used by each type of case under each of the eight procedures. To a considerable extent the necessary information was available in obstetric and other hospital and local authority

records. However, certain important issues called for special work study analysis (15). The critical resources and the constraints on their use are summarized in Table 2. The major constraints are hospital beds, labour wards, and domiciliary nursing and these reflect associated inputs of resources, including hospital nursing and other staff

Table 2. *Critical resources and constraints on their use*

Resource		*Units of measurement*	*Consumption during 1970*	*Constraints*	
				Max. available	*Min. viable*
CU	beds	days	13,400	14,200	11,400
GPU 1			4,800	7,900	4,400
GPU 2			600	1,300	500
GPU 3			1,900	3,300	1,200
GPU 4			1,900	3,300	1,200
GPU 5			700	1,300	500
CU	labour wards	hours	2,700	3,300	na
GPU 1			1,700	3,300	na
GPU 2			200	1,600	na
GPU 3			700	3,300	na
GPU 4			500	3,300	na
GPU 5			200	1,600	na
County	domiciliary midwives	hours	14,100	20,100	na
City			9,800	16,000	na

and facilities. Both maximum and minimum levels for particular resources are considered, the latter being determined on the basis of consultations with the relevant staff as the threshold below which continued existence of the resource would not be viable under existing operational policies (as far as they can be defined). The maximum levels of usage of particular resources are long-term averages and provision has been made for short-term fluctuations due to seasonal and random effects, on a scale agreed in discussion with the appropriate medical and other staff. Account is also taken of the use of hospital beds and other facilities for antenatal care.

Other constraints on the way in which the system can be operated must also be considered. In the first place the total numbers of cases involving intervention and non-intervention have been fixed at the 1970 levels as 441 and 976 respectively for the city and 790 and 1,486 for the county. Changes in birth-rates, in population structure, or in medical practice at delivery would affect these totals. It has also been assumed that the use of the CU will continue at least at the 1970 levels

so that lower limits for the intervention and non-intervention cases dealt with by the CU were set up respectively as 353 and 265 for the city and 504 and 379 for the county confinements.

COSTS

Three broad types of cost have been identified in relation to medical care. In the context of resource allocation in the maternity services it is only necessary to consider the elements of cost which vary between the different procedures. This means that consideration can be restricted to costs to the medical services and to costs to the patient and her family. Costs to other public or private authorities are unlikely to vary significantly from one procedure to another and have therefore been ignored.

As far as the health services costs are concerned, the identification of the consumption of resources associated with each procedure means that the costs per case may be calculated as the sum of the costs of each set of resources. Thus the costing problem consists essentially of the estimation of the unit costs associated with each separate health services resource. At the present time the bases used for the presentation of financial accounts differ between the three branches of the health services. For all three branches, moreover, further information is required to estimate the unit costs of particular major resources (11). A study was carried out of the position in 1970 and the costs of the system are summarized in Table 3. The figures presented in this table cover the GP (Supplemental Maternity Fee), the local authority (antenatal and post-natal care and delivery where appropriate) and the hospital costs which vary from one procedure to another. They exclude costs such as dental care which do not differ between procedures.

Table 3 shows that, as expected, the deliveries which involved intervention cost more on average than the remainder. This difference reflects longer average periods of institutional confinement and greater inputs of resources in hospital. The difference is greatest at the CU, where most of the mothers who required intervention were confined. However, for deliveries without intervention the average cost at the CU was lower than at any of the GPUs. Amongst the GPUs, the lowest average cost per case was at the largest (city) GPU and the highest at the two smallest GPUs. The costs per case at the most expensive GPU were 70 per cent higher than at the CU.

The costs considered in this study reflect two main components,

which may be termed 'fixed' and 'variable'. The variable cost component may be assessed in terms of the additional cost associated with one further delivery at the 1970 levels of throughput. The corresponding figures for each type of case and for the various places of confinement are also shown in Table 3. Differences between the procedures are considerably smaller than for the average costs. Furthermore, the costs of one additional delivery are considerably lower than the average costs per case. This is a reflection of the fact that the major part of the total is concerned with items such as staff, which would not be affected by one further case in a year.

In many applications of optimization techniques by mathematical programming it is customary to regard the costs per additional case

Table 3. *Revenue costs on maternity care per average case for 1970 system (based on levels of use of resources in 1970)*

Place of confinement	*Intervention during delivery*	*No. of deliveries*	*Cost per delivery (£)* *Total**	*Next additional† case*	*Family cost‡*
City (CU)	No	644	90	8	20
	Yes	857	116	11	20
City (GPU 1)	No	782	93	16	23
	Yes	129	103	17	15
Rural (GPU 2)	No	95	136	19	21
	Yes	3	169	21	15
Rural (GPU 3)	No	189	110	15	21
	Yes	167	138	20	15
Rural (GPU 4)	No	229	111	16	15
	Yes	41	123	17	15
Rural (GPU 5)	No	75	145	16	15
	Yes	15	157	18	15
City domiciliary	No	204	94	16	18
	Yes	9	94	16	18
County domiciliary	No	244	77	16	18
	Yes	10	77	16	18

* The total costs include all items which vary from stream to stream (ie GP supplemental maternity fee, ante- and post-natal care, and delivery by local authority and institutional care where appropriate).

† These are the costs of handling one further case, given the levels of usage which prevailed in 1970. They cover solely those items which can be attributed to the particular mother, and exclude items such as staff, running costs of institutions and administration. Because of the differences in the GP supplemental fee between deliveries in the CU and elsewhere, a sum of £8 has been added to the GPU and domiciliary deliveries.

‡ Family costs for cases at the CU and GPU 1 have been calculated as a weighted average of costs for city and county cases.

as 'marginal costs' and to assume that differences in costs between different sets of inputs may be measured in terms of the corresponding differences in the totals of the marginal costs. This is perfectly satisfactory in situations in which only marginal changes in the system are contemplated. If, however, we wish to consider major changes in resource allocation in the context of the maternity services the concept of marginal costs is less helpful. For example, if the number of confinements in one of the GPUs were to be doubled, the total increase in cost would be unlikely to be merely the cost per additional case at the existing level of throughput multiplied by the number of additional cases.

In order to deal adequately with the situation in which major change is contemplated it is necessary to determine the structure of the total costs of using a resource as a function of the number of units used. This involves a consideration of the reaction of the medical and lay management to changes in the workload in particular parts of the system. In the present analysis we have regarded the total costs as a linear function of the utilization of the resource of the form given in equation [2]. The coefficients α_k and β_k have been assessed on the basis of the 1970 situation. This is probably satisfactory for the present analysis, although a detailed empirical study of the cost-utilization relationship would be desirable.

The costs to the patient and her family of each of the procedures have been estimated by means of a special study of a sample of about 300 confinements in the city (9). Briefly, the method of inquiry was by means of an interview with the mother at which all measurable costs associated with the confinement which might vary between the different procedures were identified. The 'family costs' per delivery are also shown in Table 3. There are, as expected, some differences between the various places of confinement. There are also variations in terms of the place of residence of the mother, which reflect differences in transport costs.

In the analysis of the benefits associated with alternative policies, the costs of the system have been included as constraints. As part of a separate study of the same data Ashford *et al.* (4) have shown that substantial reductions of health services costs are possible under a range of alternative policies. The health services costs have therefore been constrained at less than £350,000, which is 90 per cent of the 1970 situation. As far as the family costs are concerned, a basic National Insurance payment of £25 is made to the mother for each

liveborn child, whatever the place of delivery. In the analysis the family costs which vary from one procedure to another have been constrained to less than £75,900, which represents 105 per cent of the 1970 figure. However, this upper limit is still £16,400 below the amount received in the form of maternity grants.

BENEFITS

In order to apply the procedures outlined in the previous section, an essential prerequisite is to make specific assumptions about the benefits associated with each procedure for each type of case. This is in itself a very substantial problem, if reliable and objective quantitative estimates are to be made and further research is required to ascertain certain essential items of information. However, in order to obtain a reasonable basis for an analysis of resource allocation it was decided to proceed in terms of subjective estimates of relative benefits. The authors have consulted what is in their opinion a representative selection of the medical and nursing staff directly involved with the care of the mother and the newborn child. The main feature of the opinions expressed was a fundamental disagreement about the relative merits of different procedures. Three main points of view emerged, which may be attributed respectively to the consultant obstetrician, the GP, and the domiciliary midwife. Rather than to attempt to produce an artificial consensus from such differing components, it was decided to calculate the benefits of the various procedures on three different bases, although a common underlying structure has been adopted, as indicated in Table 4.

The benefits considered in this table are those which arise from the health care system, either directly in terms of the improvement in health of the mother and child or indirectly in terms of the contribution of the system to the 'hotel' or other social care. Benefits associated with validation services apply equally to all procedures and have therefore been ignored. The standard of reference used in these calculations is the benefit enjoyed as the result of a non-intervention delivery at an institution, excluding subsequent in-patient stay. In other words, it is assumed that for a 'natural process delivery' in an institution, the care provided results in a benefit, in comparison with the hypothetical situation in which no health care is available. Table 4(*a*) is concerned with benefits additional to this standard. For deliveries taking place at home, it is assumed that a benefit, *f*, exists in relation to the possibly lower average risk of cross-infection and to the more

favourable psychological effect of home confinement on the mother, the child, and her family. Confinements taking place at institutions receive hotel benefits at a rate of g units per day of stay. However, for non-intervention deliveries we have assumed that these hotel benefits are cancelled out by the medical and psychological disadvantages of staying in hospital after more than five days of in-patient stay. The benefits of medical care for non-intervention deliveries have been taken as constant for all places of confinement. On general grounds, it is to be expected that the main effects of medical care are likely to be felt for deliveries which require (and presumably receive) intervention. The amount of any such benefit is taken as at least as great in a remote GPU as at home, and at least as great in a central GPU as in a remote GPU (because of easier access to the CU) and at least as great in a CU as in a central GPU.

Different values are taken by the benefit parameters for the three

Table 4. *Excess over 'standard' benefit in terms of type of case and procedure*

(a) General structure

Type of case	*Place of confinement (procedure)*			
	Home	*Remote GPU*	*Central GPU*	*CU*
Intervention	$x+f$	$y+dg$	$z+dg$	$t+dg$
Non-intervention	f	$dg,\ d \leq 5$ $5g,\ d > 5$	$dg,\ d \leq 5$ $5g,\ d > 5$	$dg,\ d \leq 5$ $5g,\ d > 5$

KEY

$\left.\begin{matrix} x \\ y \\ z \\ t \end{matrix}\right\}$ = Excess benefit associated with effect of health services on $\left\{\begin{matrix} \text{home} \\ \text{remote GPU} \\ \text{central GPU} \\ \text{CU} \end{matrix}\right.$ deliveries with intervention at

f = Excess benefit associated with home deliveries in terms of psychological and other effects on mother, child and family.

g = Excess benefit associated with each day of institutional stay (applies for first five days only for non-intervention deliveries).

d = Number of days of institutional stay.

(b) Range of values of benefit used in the calculations (arbitrary units)

Parameter/attitude	x	y	z	t	f	g
I 'Consultant obstetrician'	5	5	10	30	1	1
II 'General practitioner'	5	15	20	25	2	1
III 'Domiciliary midwife'	5	5	10	25	5	1

attitudes and the figures used in the calculations are shown in Table 4(*b*). The scale of measurement is in each case arbitrary ('preference units') and the differing attitudes are reflected in the values assigned to the various parameters. On the assumption that there is no question that an acceptable maternity care system must ensure that all confinements receive medical care of some sort, the benefits considered in the resource allocation models are the excess benefits over the standard as defined above. If there were a serious possibility that some confinements might take place without medical care it would of course be necessary to work with absolute rather than relative values. It will be noted that we are concerned in this analysis solely with benefits to the patient. In particular, we have ignored the undoubted benefits in terms of expertise and job satisfaction to the GP and the domiciliary midwife which result from deliveries in GP hospitals and at home. Whilst the managers must undoubtedly take note of issues of this type, we feel unable at this stage to determine the appropriate weighting with respect to direct benefits to the patient.

On the basis of Table 4, the benefit enjoyed by each type of confinement under each procedure may be calculated. The total benefit associated with the system may then be derived in terms of the sum of the products of the number of cases of each type dealt with under each procedure and the corresponding benefit. The numbers of cases are then adjusted by appropriate mathematical techniques in order to obtain the greatest possible total benefit consistent with the constraints on the system.

MATHEMATICAL PROGRAMMING

Having outlined the structure of the analysis, mathematical programming techniques have been applied to determine the flow of patients which leads to the greatest total benefit, subject to the various constraints on resources, costs and resource utilization defined above and in accordance with the pattern of medical care prevailing in 1970. In this study we have not considered the effect of variations in the pattern of medical care associated with particular procedures, such as the reduction in the length of post-natal stay in hospital. In order to accommodate the particular cost structure used in the analysis, integer programming was used. For each of the three attitudes a sequence of analyses was carried out. In the first place, a feasible solution was determined. If the resulting costs were substantially below the limit set in the original linear programme, the analysis was re-

peated with a lower set of cost constraints. The change in the definition of the problem resulted in a new optimal solution in terms of benefits. This process was repeated in a series of iterations which culminated in the solutions set out in Table 5. In this way, we have reacted in much the same way as we would expect an active district management team to meet a changing real-life situation.

Table 5. *Policies for resource allocation*

Place of confinement (procedure)	*Intervention during delivery*	*No. of deliveries according to attitude regarding benefits*		
		Consultant	*GP*	*Domiciliary midwife*
City (CU)	No	767	767	654
	Yes	982	982	1,069
City (GPU 1)	No	1,287	1,287	717
	Yes	209	209	117
Rural (GPU 2)	No	209	209	0
	Yes	6	6	0
Rural (GPU 3)	No	0	0	0
	Yes	0	0	0
Rural (GPU 4)	No	0	0	0
	Yes	0	0	0
Rural (GPU 5)	No	154	154	0
	Yes	32	32	0
City domiciliary	No	0	0	416
	Yes	0	0	17
County domiciliary	No	45	45	675
	Yes	2	2	28
NHS cost	(% of 1970)	89	89	84
Excess benefits	(% of 1970)	111	106	110
Family cost	(% of 1970)	104	104	101

Reference to Table 5 shows that the sets of benefits attributed to the consultant and the GP lead to identical policies. In either case, two of the four remote GPUs would be closed and institutional care would be concentrated mainly at the CU and the city GPU. The two remote GPUs which remain open are both small (4-bed) units for which the fixed element of cost is low in comparison with the two larger remote GPUs. Both the CU and the city GPU would be run at virtually full capacity. There would be no domiciliary deliveries in the city and only a relatively small number in the county. The domiciliary midwives would, however, be concerned with both antenatal care and

post-natal care for patients delivered in hospital. Both solutions lead to a cost reduction of 11 per cent in comparison with the 1970 situation. The relevant (subjective) benefits are increased by 11 per cent on the basis of consultant attitudes and by 6 per cent on the basis of GP attitudes.

The expression of benefit in terms of the attitude attributed to the domiciliary midwife leads to a radically different solution. The proportion of home confinements is greatly expanded in both city and county. The city GPU (which involves a relatively low average cost) would remain open whilst all of the four remote GPUs would be closed. In comparison with the 1970 situation, the policy would result in a shift in confinements from GPUs to the domiciliary service. It is interesting that although costs would be reduced by as much as 16 per cent, subjective benefits in terms of the attitudes of the domiciliary midwife would increase by 10 per cent.

The closure of GPU 3 under each set of assumptions reflects the relatively high proportion of intervention deliveries, which were 'transferred' to the CU, where the benefit is greater, by the programme. The existence of identical solutions for the consultant and GP attitudes reflects the cost constraints imposed by the iterative procedure. If costs were allowed to attain, say, the 1970 level, different policies would emerge. Family costs under the GPU and consultant attitudes are 4 per cent greater than in 1970, in comparison with an increase of 1 per cent for the midwife solution.

Whatever weight is given to the subjective opinions attributed to the three classes of staff, it is clear that substantial savings can be made without violating any of the basic constraints or without changing existing (1970) standards of practice. Shortening the length of stay in hospital for non-intervention cases would produce further savings in cost and exploratory analyses of the basic data have been carried out under a range of assumptions about length of in-patient stay. In this way the mathematical programming technique may be used to explore the effect of varying procedures as well as different uses of resources according to existing procedures.

Comment

The procedures with which we have been concerned in this paper are characteristic of the techniques of operational research. In essence, we have examined a complex real-life situation, extracted what we

consider to be the essential features by means of a detailed analysis of the way in which the system operates, constructed a conceptual model to provide a simplified representation of the existing patterns of care and examined by the use of mathematical programming the consequences in terms of attributes of the system (including benefits, costs, and resource utilization) of variations in the parameters of the model which correspond to different management policies in real life. This is essentially an exercise in simulation and in this sense has close affinities with the essay in which Knox (14) describes a simulation system for screening procedures. An analysis of different policies on an empirical basis, which would not be practicable because of operational and financial constraints, is replaced by a 'paper exercise' based upon a mathematical model. In neither case, however, can this paper exercise be described as a 'back-of-the-envelope' calculation, since the complication of the basic situation demands the use of a digital computer to carry out the necessary calculations. When studies of this kind are undertaken, it must always be borne in mind that the results are applicable in the real world only to the extent that the concepts and model are realistic. The validity of the assumptions must be tested at every opportunity and particular attention must be given to the possible need for elaboration of the model to cope with specific detailed facets of the situation.

In our approach, we have concentrated on choosing a management policy which leads to the greatest benefit possible in the circumstances with which the health services are faced. We have produced quantitative estimates of benefit on the basis of subjective opinions expressed by the experts in maternity care. In this respect, we are using the same 'raw material' as is currently available to the policy-makers, who make decisions in the light of subjective opinions expressed in committees or by pressure groups. What we have done is to attempt to quantify the strength of these expert opinions and on this basis to apply quantitative methods in combination with objective data concerning procedures, resources, and costs to determine the best management policy. No claim is made that any of the three sets of benefit figures used in the example represents the absolute truth. Indeed, we would emphasize very strongly the need to consider a wide range of opinions in order to determine the extent to which policies are sensitive to different assumptions about benefits.

The example demonstrates very clearly the complications which

arise even in a comparatively simple application. This is not surprising, since the health care system is a complex undertaking characterized by an almost total absence of information about its most critical features. The importance of obtaining accurate basic data and an appropriate conceptual framework concerning important aspects such as the costs of resources cannot be emphasized too strongly. In this respect, there can be no proper substitute for detailed and painstaking fieldwork. Unless there are to be radical changes in basic accounting procedures within the NHS, special work studies and the empirical analysis of operational aspects such as the relationships between the unit costs of critical resources and their current throughput cannot be avoided. These are fields in which the basic methodology is not yet firmly established and much remains to be done in developing professional expertise.

In the long run, consideration must be given to the allocation of resources between different sectors of the health services simultaneously. In our example we were unable to take account of the 'opportunity costs' of alternative uses of resources such as remote GPUs and, ideally, allowance should have been made for this aspect in our calculations. Moreover, there are strong *a-priori* grounds for the view that a major inadequacy of the current procedures for health services management relates to the difficulty of allocating resources *between* rather than *within* the different sectors on a rational basis. Furthermore, under a system such as the NHS the balance between the contributions of the health services, the family, and other agencies is of critical importance and some basic principles to provide political guidelines to the desirable relative contributions at different family income levels is required.

In the present paper we have given little attention to the mathematical aspects of the processes involved in the optimization of benefits. In our example we have applied conventional linear programming methods. However, there are clear indications that the use of non-linear objective functions and constraints might improve the adequacy of the mathematical models. Further study of the application of separable programming techniques and other numerical methods is required.

Finally, the use of our mathematical models involves the development of separate systems of classification for patients, for types of case, for medical care procedures, for resources, and for costs. These systems relate directly to the management of the health services and

if established would provide an effective framework for the collection of information for management purposes. It is clear that existing information systems such as Hospital Activity Analysis can never meet this requirement and that a new approach to the collection of information about health care is required.

ACKNOWLEDGEMENTS

The authors are obliged to the staff of the maternity and other sectors of the health services at all levels for their advice and their willingness to assist with the provision of information.

REFERENCES

1. ASHFORD, J. R., and PEARSON, N. G. (1970). 'Who uses the health services and why?' *Jl R. Statist. Soc., A*, **133**, 295–357.
2. —— (1972). 'Patient contacts in general practice in the National Health Service', *The Statistician*, **21**, 265–89.
3. —— READ, K. L. Q., and RILEY, V. C. (1973). 'An analysis of variations in perinatal mortality amongst local authorities in England and Wales', *Internat. J. Epidemiol.* **2**, no. 1.
4. —— FERSTER, G., and PETHYBRIDGE, R. J. (1973). 'Information requirements for management at the district level in the reorganised NHS', *Commun. Med.* **26**, 307–17.
5. COCHRANE, A. L. (1972). *Effectiveness and Efficiency in the Health Services*, Rock Carling Lecture (Oxford University Press for the Nuffield Provincial Hospitals Trust).
6. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1971*a*). *National Health Service Reorganisation: Consultative Document* (London: HMSO).
7. —— (1971*b*). *Digest of Health Statistics* (London: HMSO).
8. —— (1972). *Management Arrangements for the Reorganised National Health Service* (London: HMSO).
9. FERSTER, G., and COOPER, J. (1974). 'Family costs in maternity care' (to be published).
10. —— and PETHYBRIDGE, R. J. (1974). 'A local maternity care system under review', *Internat. J. Hlth Serv.* (to be published).
11. —— —— (1973). 'The costs of a local maternity care system', *Hosp. and Hlth Serv. Rev.* (to be published).
12. FUCHS, V. R. (1972). 'An economist's view of health and medical care', in Fuchs, V. R. (ed.), *Essays in the Economics of Health and Medical Care*, chap. 1 (New York: Columbia University Press for National Bureau of Economic Research).
13. KLARMAN, H. E. (1965). *The Economics of Health* (New York: Columbia University Press).
14. KNOX, E. G. (1973). 'A simulation system for screening procedures' (to be published).

15. MAKUC, D. M., and FERSTER, G. (1973). 'Distribution of variable costs for institutionalised maternity care' (to be published).

16. RO, K.-K. (1972). 'Patient characteristics, hospital characteristics and hospital use', in Fuchs, V. R. (ed.), *Essays in the Economics of Health and Medical Care*, chap. 5 (New York: Columbia University Press for National Bureau of Economic Research).

4

Yet the basis of it all is information and that depends on systems

Information systems in the unified Health Service

M.R. Alderson
MD, MB, BS, FFCM
Professor of Medical Information Science, University of Southampton

Information systems in the unified Health Service

Summary. *This paper begins with a brief introduction on the requirement for information systems in the unified Health Service, and the need for co-ordination of these systems within a region. There is then review of some of the documents on the reorganization of the NHS, pulling together the various points that have been made about the need for management information and the responsibilities for collecting, handling, and using such information at the different levels in the service. It is a background to the remainder of the paper, and will be common knowledge to many readers. The next section discusses some of the main classes of data required for management and planning in the Health Service. The third section describes the present location of data collection, with respect to capture within the present three branches of the NHS. The Appendix should be referred to in conjunction with this section ; it lists the returns currently forwarded to the DHSS. This provides an indication of the information that could automatically be channelled through the district information units, and collated at regional level for transmission to the DHSS, and back to area and district levels. By channelling these returns through a regional unit it should be possible to provide a collating and processing service, so that comparative data can be fed back to districts and areas. The data sent back to districts should show the performance of an individual district against the area and regional pooled material, whilst the data sent to area should show the performance of the constituent districts, the pooled area, and pooled regional material. This is followed by comment on some changes in the handling of*

information that should be considered upon reorganization, with further emphasis upon the need for co-ordination; particular reference is made to: environmental data, liaison with departments of social service, occupational health data, school health statistics, sickness absence, and vital statistics.

There is then a section on computers and information handling; this briefly describes the current situation in the country, and the anticipated changes over the next few years. This is followed by the most important section of the report which puts forward suggestions for the regional organization of information handling, post April 1974. This section advocates a responsibility at regional level for co-ordination of information systems; also the main responsibility for the development and organization of such systems, the statistical treatment of data, and the provision of interpretation of this material. Systems development must, however, be closely tailored to users' requirements: users at district, area, regional, and national level. It is suggested that at area there will be no responsibility for data collection, and negligible responsibility for data processing; however, the staff located at area will be one of the main categories of users, both of data sent directly from districts, or via a regional information unit. At district level there is an important role in supervising the capture of data; it is anticipated that there will be minimal processing at this level. The DMT will form one of the main users of information, the bulk of which will have been processed by the regional information unit (for any districts in which locally based computer projects are being developed there will be collection of much more extensive data and a considerable capacity for data processing: this may set a precedent for the long-term development of information systems and deployment of computers). The section then discusses the general flow of information in the reorganized NHS, and is accompanied by a chart which shows in a highly schematic way how the medical information will flow from activity level through the districts, between the different authorities within a region, and up to central government. Also on the chart is an indication of the other sets of data that should be drawn into the general picture when assembling the profiles of the districts, and co-ordinating responsibility for care in the total health and welfare fields. The section ends with brief note on the staffing requirements for handling information and a comment on the need for an 'information systems co-ordinating committee'.

INTRODUCTION

The coming unification of the NHS presents the opportunity to review the handling of information in the NHS; it should stimulate the introduction of comprehensive systems which bring together information that currently is handled in many different and discrete ways throughout the segments of the NHS. The first main section discusses the proposals for reorganization, with particular emphasis being placed upon the functions at regional, area, and district level that require adequate information; the role of the community physician (CP) is discussed in this context. The next section expands on the classes of information required to fulfil the management needs; this is followed by a description of the currently available data, and the location of data collection. The paper then indicates how information is presently handled in the NHS, and suggests ways in which handling could be modified either by different approaches in collating and processing the information, or by marginal extensions of the range of material collected. This section identifies some of the major gaps in the currently available information, and comments on long-term solutions to these problems. A brief section deals with the computing power that is likely to be available in the coming few years in the NHS. Finally suggestions are made about the long-term development of information systems in a unified Health Service.

The proposals about long-term development would require agreement at district, area and regional level if they were to be implemented; some of the suggestions could not be implemented unless there was co-ordination and willing co-operation between all the districts within a region. Some of the suggestions entail the changes in the collection and transmission of information through to the Department of Health and Social Security (DHSS), and the Office of Population Censuses and Surveys (OPCS). Some of the suggested changes are fairly minor and could probably be negotiated without any great difficulty with the central authority. The opportunity should not be lost to review now the statistical returns called for by the DHSS, and relate these to the information requirements of region, area, and district. This point is discussed in the final section.

Reorganization of the National Health Service

FUNCTIONS AND INFORMATION REQUIREMENTS AT REGIONAL, AREA, AND DISTRICT LEVEL

The ideal arrangement for a management information system can only be finalized if one has a clear picture of the proposed organization and functions of different components of the unified Health Service. Gradually documents are coming out which throw light upon this; some of these documents have major gaps, and there is some conflict between the detail presented in the different reports. The White Paper on National Health Service reorganization (10) suggested that the management structure at area level will only be worked out when the Shadow Authorities are established in 1973, though assessments can be carried out in advance. The DHSS report *Management Arrangements for the Reorganized National Health Service* (the management study report) (13) provided some guidance, though acknowledging that this is subject to change with further experience. As a preliminary to the consideration of the information requirements, a personal view of the possible organization and functions of segments of the NHS are spelt out in the following paragraphs. The structure and general functions of the components are only brought in, where they are related to the use of information.

The White Paper suggested that the central department will have a responsibility for the standardization and preparation of national statistics. Obviously within the White Paper there was not the opportunity to argue out the need for standardization, versus the need for local flexibility. It is important to relate the feedback requirements at local level with the requirements for national statistics; if in fact the national system is calling for data that are difficult to collect locally or are of no use locally, and could be replaced by slightly different sets of information, this should be carefully examined in order to try and meet these two rather different requirements. This point receives further attention in a later section (p. 125).

There will be a regional health authority (RHA) whose task will be strategic planning, co-ordination, and supervision, with some executive role. The planning and definition of priorities will be based on a review of the needs that have been identified by the area health authorities (AHA). The RHA will not specifically be responsible for collecting the data on needs, but will examine the papers drafted upon such work at a subsidiary level. The RHA, having agreed area plans,

will have a responsibility for monitoring the performance of the areas against their objectives within the specified programmes. This will require a system for data collection at district level, and examination of these performance data at regional level. The RHA will be responsible for the provision of the blood transfusion service, and this is a service that in itself can benefit from the clear presentation of information about projected needs, as well as the exploration of the use of computers for optimizing donor recall. It is suggested that the regions should carry out epidemiological studies: though what these should be has not been spelt out (the use of such studies will be described in further detail [p. 107]). It is suggested that the RHA may provide computer facilities, and be responsible for the compilation and processing of statistics; this presumably reflects the scarcity of trained staff, and the need to rely on such staff and expensive capital equipment based at regional level.

The management study remarks that there will be a team of officers accountable to the RHA and responsible amongst other things for monitoring AHA's performance; it does not indicate how this will be done, though it mentions that the team might consist amongst others of a medical officer. Also it does not specifically identify the medical officer, or any of the other members of the team, as having a particular responsibility for monitoring AHA's performance.

The AHA will be responsible for planning and developing services (in consultation with the matching local authority and with the RHA). It will regularly and systematically appraise the quality of existing services in the districts and assess unmet needs, comparing the situation with national standards of care and identifying opportunities for improvement. It is suggested that this approach will encourage comparisons between alternative methods of care (for example, home versus hospital care). The AHA will set up a family practitioner committee (FPC) to administer the contracts of individual practitioners (this FPC will have a responsibility for family doctors, dentists, pharmacists, ophthalmic opticians, and dispensing opticians; thus the title FPC is slightly misleading). The FPCs will deal direct with the central department, whilst the AHA will have a responsibility for the planning and development of health centres, the approval of practitioners' own proposals for providing premises, plans for services in new towns and development areas, and general arrangements for nursing and other skilled staff employed by the AHA or local

authority to work with family doctors whether in health centres or elsewhere. At area level the chief officers including a medical officer will form an area team of officers (ATO). It is not clear what part such a medical officer or other members of the ATO will play in the handling or use of information.

There will not be a separate formal authority below the areas, but the districts will be the natural community for the planning and delivery of comprehensive care. In particular the management study suggests that the district teams would be accountable for operational services either through the area team or directly to the AHA. The White Paper stressed that supervision of good planning entails the monitoring of performance to ensure that planned standards of service and efficiency are being achieved. It suggests that performance can be monitored: by the collection and analysis of regular statistical information; by special reports and inquiries; by periodic visits and contacts between the staff of the DHSS and districts; by systematic visiting such as by the ATO; and by self-critical observation and analysis of practice by the professions. It points out that there is a need to seek methods of obtaining improved information and more effective measurement of needs and of performance, and that such measurement is exceptionally difficult. The White Paper indicated that the specialist in community medicine will assess the needs for health services, evaluate the effectiveness of existing services, and plan the best use of health resources. The White Paper did not indicate clearly whether this work will be done by a CP at area or at district level, though the management study put this activity at district level. The management study indicated that the district management team (DMT) will include a district community physician (DCP) who is accountable for the assessment of community health needs, and for the surveillance of the district-wide provision of health care. It is stressed that the DMT would be a consensus group, with a chairman appointed or elected. It is emphasized that the DCP has important roles: to identify opportunities for improvement in the provision of services; in co-ordination of patient group teams; and the co-ordination of health and social services. In a recent paper Burbridge (9) has suggested that the DMT will be concerned with organizing operational policies and developing district plans, rather than day-to-day management whilst the area authority is to co-ordinate health care with local authority planning.

AHAs will set up community health councils, to represent the

views of the consumer. These councils will have the power to secure information. Amongst other things they may study statistics on the volume and type of complaints about a service or institution, using this as a measure of public satisfaction (though the AHA will have a separate complaint machinery). A new authority will price prescriptions for the whole of England, and will be a joint health authority responsible to the Secretary of State. This will presumably be located in one place in the country, and therefore all prescriptions will have to be transmitted to this authority.

In the White Paper chapter referring to collaboration with local government, there is the comment that management services (such as computers, O & M and work study, and information services) are some of the facilities that could be shared between the NHS and local government for day-to-day operation. It is important also to note that in the same paragraph there is reference to the need for social service staff to be made available by the local authority to the health authority, and that local authorities must have professional advice in order to carry out their statutory functions in the personal social services, education, environmental health, and housing fields.

What is not quite clear from the White Paper is how the regional, area, and district staff will work in conjunction in devising integrated systems such as are required for the information handling within a region. Information handling is a complex problem at regional level because of the need to provide a service for districts, and also handle the considerable volume of data that will have to be transmitted from a regional information unit to the relevant central departments (DHSS, OPCS, DES, and possibly others). There is, of course, need for standardization of data capture and coding, and there must therefore be some mechanism for introducing and controlling this standardization. It is particularly with the introduction of computing systems that standardization can be re-examined, and some of the benefits of standardized systems become apparent.

THE WORK OF THE COMMUNITY PHYSICIAN (SPECIALIST IN COMMUNITY MEDICINE)

The White Paper touched on the functions of the CP, particularly at district level. It is appropriate to refer in somewhat greater depth, however, to the way in which such individuals may function in the unified Health Service, because it is probable that a major responsibility for handling of information will be placed with the CP. The

Report of the Working Party on Medical Administrators (the Hunter Report [11]) says that the DCP will have the functions (as a specialist in community medicine) of building up and maintaining a health profile of the district through the collection and analysis of relevant statistical and other data, and he should contribute to the health information input of the area. He would participate with colleagues in research into the aetiology of disease, and its association with social and environmental factors. He would make available to his consultant colleagues his knowledge of the district and his expertise in the organization of health care, and be particularly concerned with developing services or activities for the promotion of health. As a member of the DMT he would take part in the preparation of suggestions for the AHA on the development of area plans and policies, and participate in their implementation within the district.

The community medicine specialist at area level is thought by the Hunter Committee to have a particular responsibility for the provision of health information. The report does not clearly distinguish here between the responsibility of the DCP and ACP in establishing collection and even analysis at district level, and the duty of the CP at area to use information provided by the districts. In planning at area level the CP will be responsible for continuously reviewing existing services, and advising how they can best be developed. He should also study the outcome of particular health programmes, and with the help of all concerned, identify the lessons to be learnt for the planning and development of future services. In relation to management the community medicine specialist's tasks will be to monitor and evaluate the operation of all health services in the area, and to promote improvements in the organization and delivery of services within available resources. The report then goes on to emphasize that the health information system will be used to monitor services in the district and area levels and the CP must take a lead in building up systematic development of indices to measure the performance of the health services. This will require co-operation between the CP and clinicians in the collection and interpretation of data. It is emphasized that the CP will in many areas have to start virtually from scratch in building up the health information required by area and district management on the functioning of general practice. The Hunter Committee felt that the CP should marshal information on general practice, work in hospitals, and local authority health departments in a form suitable for decision-making; he would also have a special responsibility for

assessing the outcome of medical care services generally, and for initiating local studies to throw more light upon critical areas. For all these functions the report suggests the CP may be operating in an area or part of an area; it is not clear how much of this work should be carried out at district level. In addition the CP will be responsible for giving advice and assistance to local authority departments particularly for social services and education, and to voluntary and other bodies. He will also be responsible for the provision of medical services required by local government authorities with responsibility for environmental hygiene and communicable disease control.

The CP will be a member of the group responsible at regional level for planning, and for the determination of objectives and policies. He will be concerned with the co-ordination and monitoring of the medical services provided by the various AHAs in the region, and will be involved in discussions on resource allocation. The chief administrative medical officer (CAMO) and his staff will have a special responsibility for building up and maintaining health information services for the region. This is in addition to having functions with regard to postgraduate medical education, provision of medical expertise in building projects, encouraging and promoting clinical research and research into the operation of the medical services, and the provision and arranging for the provision of specialist advice. The Hunter Report notes that the regional and area authorities will need comprehensive information systems, but their main emphasis seems to be on the development of units at regional level. They comment that there will be a need to bring together existing information systems, the current growth of technical capabilities in the field of information engineering, and the needs of a more comprehensive administrative system. This will create a demand for both capital investment and skilled staff. The report suggests that at regional level there should be a senior medically qualified member of the CAMO's staff, experienced in identifying the information requirements for health service evaluation, management, and planning; he should also be knowledgeable in the use of modern data-processing facilities for these purposes. This person will be a key member of the multi-professional team directing the information services unit, and will bear the particular responsibility for seeing health information requirements are met in full. It is suggested that the regional staff should advise the areas on the development of information systems, whilst at area level there will be need for doctors with specialized knowledge of information needs and

data interpretation, but not the same degree of involvement with the design and operation of information systems.

The Scottish Home and Health Department Report (17) *Doctors in an Integrated Health Service* puts the functions of the specialist in community medicine slightly differently. The report suggests that these are to investigate and assess the needs of the population so that the priorities may be established for the promotion of health, the prevention of disease, and the provision of medical care. In addition the specialist will be concerned with co-ordinating medical expertise so that policies which are in accord with medical needs can be presented to the AHA. The specialist in community medicine would provide for his clinical colleagues data about population needs, would assist in the evaluation of each division's activities and the managerial options open to it, and would provide a link not only with the area administration but also with the various services providing information and other support for divisions. They have a responsibility for assembling and interpreting data relating to health service needs, population trends, and the effectiveness of existing services. In collaboration with those responsible for providing services he will ensure that resources are being used to the best effect whether in hospital, health centre, or elsewhere and will play a part in the planning and initiation of new services. He should assist each clinical division to assess plan and organize its work in relation to the needs of the population, and help to provide a communicating link with other parts of the administration including the information unit, the health education unit, and other divisions. In this way by the use of modern epidemiological techniques and management skills, the effectiveness of clinical work could be enhanced. The report says that community medicine specialists will be applying techniques known collectively as 'research and intelligence'. They will be using quantitative methods in planning, organizing, and evaluating services and this work will be greatly helped by close collaboration with data collection and processing services, and with non-medical experts in operational research. The report agrees that every area authority will need access to an information service under a senior information processing officer. In addition there will be the need to undertake population surveillance, with a further objective of the creation of a unified patient record system. They recognize the need for development of new forms of analytical techniques capable of obtaining rapidly and accurately information which would enable the best use to be made of modern managerial

skills. The Scottish Report considers that at area level there should be a community medicine specialist available to work with each division in the Cogwheel structure. He should work along with the medical organization, and be closely identified with a specific part of it, assisting in the formulation of problems, in the interpretation of data, in co-ordinating the activities of different parts, and in preparing advice and reports for the AHA. The report advocates the organization of general practice divisions, within the district medical committee structure alongside the other medical and service divisions; but it does not advocate the formation of a division of community medicine, but sees community medicine functioning as part of the servicing system for each division.

It is appropriate to mention the report by Naylor (16) *Organisation of Area Health Services.* In this he shows seven charts of the possible organization of an area health board. He produces separate charts for single and multi-district areas, with and without district committees, and with and without a chief executive officer at district or area level. It is interesting to note that in each of his charts he has a health care research and evaluation activity as one of the supportive and advisory services at area level. He indicates that this embraces epidemiological research, analysis of health care needs, and the evaluation of medical services. This activity is functionally related to the chief medical officer at area level. Separately on the charts are additional supportive services entitled management services (embracing work study, management analysis, and systems analysis), together with a further general administrative support including patients' records, management information with financial information, office services, and data processing. These two latter supportive services are shown reporting to the chief administrative officer, or chief executive officer. It is interesting to note that his organizational chart does not indicate how information services will be provided at the district level, though in the text of his report he indicates that each of the Cogwheel divisions should have an attached CP. The Management Study Report shows at district and area level information under the district administrator (DA), but at region 'Information Services' under the RMO, whilst management services are under the RA. The logic of this is obscure, and Ashford (5) has criticized the proposal that information provision should be the responsibility of the DA.

These official reports do not spell out in any detail the specific information requirements, the way in which data should be captured,

the location of units for handling information in the unified Health Service, nor the staffing required. The following sections of this paper describe in greater detail the arrangements that could be made within a unified Health Service for processing the currently available data.

Management requirements from information systems

It is appropriate to begin with a definition of an information system. Benjamin (6) defined this as a system that collects basic information from all segments of an organization and stores these centrally for use for previously specified purposes. The system must involve carefully organized capture, editing, storage, updating, manipulation, and retrieval of the material. The system may store numerical data or textual information, or act as a reference to sub-files of retrievable information, and this information may exist in a wide variety of formats (as documents, punch cards, magnetic tapes, etc.). The system should be so planned that it provides periodic reports at varying intervals in time according to a preset schedule; in addition the system should be so organized that it can handle *ad-hoc* inquiries. Where an *ad-hoc* inquiry cannot be answered by specific manipulation and retrieval of stored data the collection of additional information may be instituted on a permanent or temporary basis, and the latter may be by way of a specific survey. Wherever possible the stored data should be used as a sampling frame, or for presentation of background information to illuminate the results of the special survey.

It is suggested that the information system should assist in management of the NHS, at all levels from health care planning teams within districts to strategic planning at regional level. To play an effective part in this range of work requires the production and presentation of rather different sets of output. For some of the management functions the information system should be used for routine surveillance of the functioning of the NHS. This should be by periodic presentation of tabular data on a variety of aspects of the NHS. The previous section described how a number of recent reports have emphasized the need for such work: HRC (73)3 again stressed the importance of a monitoring system (linked to budgetary control) in achieving accountability between statutory authorities. The only indication of the types of information required was in the Hunter Report which commented that at area level there will be a requirement for informa-

tion on: the demographic character of the area; the health needs as indicated by morbidity and mortality data; physical resources; manpower; demands made on services; performance of the service both qualitative and quantitative; the social, and educational services; housing; environmental hygiene; communicable disease, and voluntary services. In relation to this string of information, it comments that much of the data will need to be analysed on a district basis as well as an area basis. The community medicine specialist at district level will be involved both in its collection and in its interpretation.

CLASSES OF DATA REQUIRED FOR MANAGEMENT

Trends in demand for services. Information should be available to gauge the change in demand for health services. Simple indices of this change are in themselves useful, even if the complex issues underlying such change cannot be clarified. Ideally the system should present data on new attendances at family practice, fresh requests for use of ancillary staff in the community (such as requests for health visitor support or district nursing), direct access investigations requested by family practitioners, new referrals to out-patients, numbers of emergency/booked/planned/waiting-list admissions to hospital, and workload of service departments such as pathology, X-ray, etc., in the hospital.

Trends in bottle-necks. At periodic intervals there should be presentation of material on the queues and waiting-lists for various services such as: numbers of patients waiting deployment of resources in the community (such as patients requiring but not receiving routine visiting from the district nurse or other supporting services); the numbers of patients waiting and the mean waiting-time for urgent and non-urgent appointments at out-patients; the waiting-lists for in-patients (preferably tabulated by individual consultant and diagnostic category with the waiting-list converted from a count of patients to an estimate of bed-days required by the patients on the waiting-list); delay time for complex investigations such as cardiac catheterization, respiratory function tests, etc.; the number of patients in acute beds in hospital awaiting transfer to long-stay care in chronic hospitals, admission to sheltered accommodation in the community, or discharge to their own or relatives' homes upon mobilization of community resources.

Use of resources. Ultimately it would be advisable to have a profile that indicates the complete deployment of staff and physical facilities, not just as inventory of resources but in relation to services provided. For instance there should be information about the work of the family practitioner and the ancillary staff in the community; this should not be just a count of items of service performed, but should relate the workload to specific patient-care groups. Similarly one requires information which quantifies the use of out-patient facilities, in-patient facilities, and service units in hospital. The collection and presentation of this information on use of resources should be problem orientated (so that one can look at, for example, the total contribution made by all aspects of the Health Service in the care of the expectant mother, and not just the number of admissions to maternity homes, or attendances at clinic for expectant mothers).

The material on use of resources will be most usefully presented in three rather different ways; first within a district it may be used to indicate the distribution of resources between different treatment problems (for instance contrast the bed and staffing allocation for the care of children versus the care of the elderly, or the mentally ill). Secondly the information may be used at district level to present clinicians caring for a particular category of patients with comparisons of the use of resources in their own district against data for the other districts in the region (for instance one may look at variation in lengths of stay for particular conditions between consultants in one district and from district to district). Thirdly at area and regional level there will be interest in looking at comparisons between districts, pooled data for the areas and the region as a whole, and any relevant information that is published or directly available from the DHSS.

Outcome of care. The examination of use of resources will become much more meaningful when it is possible to relate variation in practice to outcome of care. The information system will have to devise means of acquiring useful indicators of outcome of care; judicious manipulation of the data may extend the currently available case fatality rate to indices of complication rates, recurrence rates, and readmission rates. At its simplest this can be achieved by bringing together readmissions of individual patients to a hospital or group, using the unit number. Accumulation of extended patient files at a regional level raises a number of important issues. Much more difficult to obtain will be indices of recovery, re-ablement, and patient

satisfaction; some of this material will have to be collected by special studies; which cannot be replicated in every region for even the major patient care groups, let alone specific diagnosis. Examination of the trends in the outcome of care will be required, and these trends within a district must be compared with data for the other districts in the region (and with any relevant published material for the country).

Monitoring of innovations in health care. It is essential that any new developments in medical care (whether these are gradual evolution in the way of delivery of health care, or more major change due to innovation and the introduction of fresh programmes of care) are monitored to demonstrate the impact of the change. For instance with preventive schemes such as immunization and cervical cytology it is essential that statistics are presented on the uptake of these procedures, identifying the subgroups of the population who do not participate in the programmes, and contrasting the subsequent morbidity and mortality in the protected and non-protected. With the introduction of administrative or medical changes in the care of patients, such as the intensive use of out-patients, day-surgery, day-hospitals and short-stay, it is important to monitor such things as time on the waiting-list, number of operations performed, bed occupancy, length of stay, turnover interval for in-patients, and outcome of care.

***Ad-hoc* data collection.** Many problems will be handled at relatively short notice by the management teams at district, area, and regional level; the presentation of periodic data as indicated in the five sections above may be irrelevant to many local issues. This may be because the specific tabulation relevant to the problem in question had not been produced; such queries require the facility to retrieve immediately a relevant tabulation from the information system. Other queries will require the special collection of material missing from the information system. This information may be available within the record system at unit or hospital level, and merely require staff to abstract the data and collate it with that available from the information system. A number of the issues will, however, require special surveys; the scope and complexity of such a survey will depend partly upon the issue in question, but also upon the time and facilities available to answer the query.

INFORMATION REQUIREMENTS FOR PLANNING

Another aspect of the function of the information system is to assist with planning. The term planning is used by different people to mean very different things; it can be defined as the process culminating in decisions regarding the future provision of the correct balance of domiciliary, out-patient, and in-patient facilities for the investigation treatment, and care of all 'perceived health needs' of the community. The planning process involves a lengthy time-scale and is particularly concerned with planning for the health needs of the next generation. Excluded from this definition is the examination of current resources and current demands resulting in the short-term reallocation of these resources in order to meet more effectively these demands 'tomorrow'. This latter process is part of the day-to-day functions of management and has been covered in the previous section.

The simplest way to plan is to identify the current workload, quantify the resources used to meet this workload, and relate this to the population in the catchment area; in this way 'norms for provision of care' can be derived. These norms can then be applied to the projected population figures in order to obtain anticipated requirements. Such an approach is likely to perpetuate the fault in the current system. Bispham, Holland, and Stringer (7) have pointed out that the studies of Barrow-in-Furness and Teesside were not able to establish objective criteria of need for medical care; the ten-year hospital plan for England and Wales used the findings from these studies and thus the recommended bed ratios were based on current demand. These authors point out that this is a very inadequate measure of need. In order to produce a more precise approach one requires detailed information about the diseases currently affecting the population, and on to this one needs to build estimates of: trends in the incidence of disease; possible changes in the attitudes of the population to health and health care; future variations in delivery of care; impending changes in therapy; and the effect that any new therapy is likely to have on the prognosis of disease. HRC(73)3 emphasizes that plans for district services should be based on a consideration of the needs of the population for health care of various kinds.

The routine data, especially the data that have been discussed in the previous section of this paper, are somewhat limited. If the data can provide a clear picture of the functioning of the major (ie costly) segment of the NHS this is of some help. It is, however, a rather

limited picture upon which to base long-term planning. There is also the need to have some prediction of the two rather different types of change which can occur in the delivery of medical care. First are minor alterations and the gradual shift due to the steady evolution of medical care; second are major changes resulting from innovation and advances in knowledge or technology. Examination of trends in patterns of working and discussion with the deliverers of medical care can provide estimates of change in practice. It is here, however, that the techniques of operational research can come into play; providing one knows the current demand and the way in which this is being met, one can build a model of the health care system, simulate change in the system, and observe in the model the repercussions that alteration in the deployment of resources has upon workload and throughput.

A report of a WHO expert committee (1971) agreed that there are at least five steps that may be recognized in such a planning process; these are first a situational analysis, then formulation of alternative tactical approaches, a decision phase, discussion and implementation, and finally evaluation. It is little use just studying data (even if these are accurate data) that demonstrate the current demand for health services; even less use is the examination of data that solely refers to demand for a particular segment of the health services. In order to build up a conceptual model of the functioning of the NHS, one really requires to know how many people are out and about in the community suffering from incipient or overt disease that have not contacted the NHS. One then requires to know what proportion of patients contact their family doctor, and for all those who do contact the family doctor, how he deals with their particular problem. Does he rely on his clinical judgement or use direct access investigations, does he prescribe treatment, does he continue to see the patient at intervals, or does he refer the patient to the hospital services? For those patients referred to out-patients one wants to know at first attendance the proportion in whom a clinical diagnosis is made; have investigations; re-attend out-patients, and how many times are they brought back to out-patients; are put on the waiting-list for admission. For those patients admitted to hospital one wants to know what investigations and what operative procedures are carried out, and what is the total length of stay. For all these alternative forms of care one requires information about outcome; both hard data on mortality, recurrence, or complications, and soft data about return to

active life, and satisfaction of both the patient and his family with the care provided. Such information would enable examination of the total resources currently involved in providing health care, and the costs entailed in the provision of such care. A routine information system would not collect precise and accurate particulars about the total uses of the NHS, let alone actually tell why some patients with raised blood pressure go to their doctor, and why others do not. No information system could tell why some family doctors treat the patient without investigation, why some carry out investigations themselves or make use of direct access information, and why some refer their patients to out-patients. However, unless we know the answer to these questions we cannot confidently comment on the potential for change in the health care system.

Matthew (15) has discussed the problem of extrapolating local findings to a wider population. He pointed out that it is unlikely that studies of samples of the population in every district of the country, and on a wide range of conditions will ever be practicable; he questions whether the prevalence of disease in one population may be deduced from that in another. He suggests that further research is needed to check on this, and to determine what other characteristics should be recorded in an attempt to make such extrapolation possible. Many health planners now dream of having a single index of the health status of a community that would serve as a guide to the requirements for a range of health problems. It is extremely unlikely that a single index can be derived, as it can in the economic field: where a monetary unit can be obtained from a number of highly diverse components. It has been suggested (WHO, 1971) that, to be satisfactory, an index of health would have to meet the following requirements: availability, completeness of coverage, quality (ie not varying with time and place), universality, ease of calculation, acceptance, reproducibility, specificity, sensitivity, and validity.

Location of data collection

Currently, the NHS collects a multitude of information about morbidity, workload, staffing, facilities, and cost. The bulk of these data are collected at the level of activity within the three branches of the NHS, and it is convenient to begin a consideration of handling of information as it now exists with the NHS in three divisions. This section deals with location of data capture and comments upon the

transmission of statistical returns to the DHSS. The Appendix lists all the returns routinely sent to the DHSS, and briefly indicates their content.

EXECUTIVE COUNCILS

Though a considerable volume of workload is carried out by the family practitioner, the dentist, the pharmacist, and the optician, the amount of information about workload collected at individual patient care levels is slight apart from the dental system for collecting item of service information for payment purposes. The executive councils are responsible for initiating a series of returns relating to staffing and cost, which they transmit directly to the DHSS.

HOSPITALS

In the hospital service information is collected on items of work at out-patients, and carried by the service units together with simple information on use of beds (SH3 returns) together with the more detailed information provided by HAA. Apart from the HAA system which theoretically should be a by-product of the use of a standard front sheet in each case folder, the collection of data is a form-filling exercise carried out for statistical purposes. The Appendix lists the other returns on workload, including the mental health statistics and data from a range of service departments. There are an extensive series of returns on staffing, and others dealing with psychiatric facilities. A summary of unit costs are submitted at the end of the financial year. The information is collated at hospital and group level and transmitted to the regional board and DHSS. Some of the staffing returns will be completed by the regional board, where this is the employing authority.

LOCAL HEALTH AUTHORITIES

Local health authorities collect information on items of service provided by their staff (district nurses, health visitors, district midwives, clinic nurses, and clinic doctors). This information is again initiated at the activity level and collated at clinics serving localities before being transmitted to the local authority headquarters (in counties the information may be handled at divisional level or level of delegated authorities). Consolidated returns on workload, together with returns on staffing, cost, and facilities are sent to the DHSS. The individual returns are listed in the Appendix.

MORBIDITY DATA

There are five main sets of morbidity data currently collected in this country, in addition to the medical information that becomes available automatically through HAA. These relate to abortions, cancer, congenital abnormalities, infectious disease, and sickness absence. These sets of data are used at local, regional, and national level to monitor trends in incidence and prevalence of disease in the community, and a brief comment follows about the current mechanisms for collecting the basic data.

Since 27 April 1968 regulations have made notification of abortion, carried out for various legal reasons, a statutory obligation upon the operator. A return has to be sent within seven days of the procedure direct to the DHSS. This material is neither collated at hospital or regional level, and strict precautions are taken about the handling of the material at central level due to its confidential nature. Consolidated returns appear in the regular publications of the OPCS and DHSS.

Since the introduction of the NHS there has been a steady spread of the voluntary scheme for cancer registration. All fourteen boards in the country now participate, but may operate slightly differently schemes for the organization of data capture. Basically the onus is on local staff to identify patients with malignant disease, and complete a form containing the minimal particulars required by the national scheme; a number of regions collect extended data. The material is collated at regional level, and transmitted to the OPCS. Some boards handle the data at regional level on a computer and transmit the required information on computer-compatible tapes to the OPCS. A simplified scheme is now in operation in Wessex whereby cancer registration is run as a by-product and adjunct to the collection of basic HAA.

In 1964 a voluntary scheme for the notification of congenital abnormalities was introduced by the then Ministry of Health. There is no standard notification form used at local level, but most local health authorities have added a section to the birth notification form, upon which the person present at delivery may complete particulars regarding identifiable congenital abnormalities. These data are collated in the public health departments, and a consolidated return is completed and forwarded to the DHSS monthly. This scheme does not automatically pick up abnormalities that come to light through subsequent investigation of children, and will be particularly likely to

miss mild and internal abnormalities that do not create an acute medical problem at birth.

Since 1899 practitioners have been required to complete a standard notification form for persons under their care who develop any of a specified number of infectious diseases. In 1968 the list of diseases included in the scheme was altered (Public Health [Infectious Diseases] Regulations, SI 1366). The returns are forwarded to the medical officer of health of the appropriate sanitary authority (ie rural district or county borough); the MOH will use this to initiate local steps in the monitoring and control of infectious disease. Weekly consolidated returns are forwarded to the DHSS by the MOH.

With the introduction of the national scheme for social security, and associated with the operation of the NHS, practitioners have had a responsibility for completing statutory forms for certifying incapacity for work from sickness. These forms are completed by the doctor caring for the individual patient; the patient has to add additional particulars to the form and transmit this to the local Social Security office (a branch of the DHSS). A simple running total of new claims for sickness absence is kept by the local offices each week; once a month a count is made of the number of cases where there is a current medical certificate. For a sample of new claimants (formerly 5 per cent, and since 1969 $2\frac{1}{2}$ per cent) a detailed return is completed at the local office and forwarded to the data-processing branch of the DHSS at Newcastle either at the end of a spell of sickness, or at the end of the sickness absence year. A voluntary scheme is operated by some local offices, which notify the hospital authorities when there has been a dramatic rise in new claims from one week to another; this scheme is particularly useful in identifying the acute problem caused by a major flu epidemic, and provides a good warning of anticipated pressure on the NHS.

VITAL STATISTICS

In this country the system for collecting and processing vital statistics is quite separate from the collection of NHS statistics. The parents have a responsibility for registering births with the local registrar, marriages conducted in churches have to be registered by the officiating priest with the local registrar; whilst deaths are certified by the attendant practitioner or the coroner, and the information is then handled by the registrar. In any local health authority there may be a number of sub-districts which have their own registrar, and a super-

intendent registrar (at the level equivalent to the local health authority). The superintendent registrar is responsible for forwarding returns direct to the Registrar-General.

The two links with the NHS are in relation to births and deaths; the attendant at birth is responsible for sending a notification of birth within thirty-six hours to the medical officer of health; the MOH then provides a list of these notified births to the registrar, who uses this to check subsequent registrations. The system for handling the birth notification requires review in relation to other documentation completed about the time of delivery of a child. At the other end of life the medical profession are responsible for providing certificates of cause of death, apart from those deaths notified to the coroner, and this cause of death certificate is incorporated in the particulars recorded by the local registrar.

There are laws and statutory regulations governing the availability of the information in the vital statistics registers; under certain circumstances the Registrar-General is able to provide information to the health services. The MOH automatically is provided with a copy of the death notification completed for each death in his area. Also information from vital registration is used in the cancer registration scheme; copies of death certificates which mention malignant disease in Part 1 or Part 2 are transmitted to the appropriate regional registry covering the area in which the deceased resided. Apart from this arrangements for provision of information has been by special agreement, such as for the Oxford Record Linkage Study where information on all births, marriages, and deaths occurring to persons resident in the area of the study are transmitted to the project.

Suggested changes in the unified Health Service

Any changes here are tentative, and require close examination of local data collection systems and local proposals for handling and using information. The following scheme is a skeleton structure proposed for discussion only. It is not suggested that these changes are likely to be easy to implement, or will necessarily be suitable for implementation in every district within a region.

CO-ORDINATION

The main change suggested is that the total flow of information should be moulded into a flexible but unified hierarchical system.

With the bulk of data capture at operational level, this is where detailed information will be available; the processing system should be so geared to handle this, that rapid turnround and presentation of material at local level occurs. Information will usually be aggregated before presentation at area or regional level, whilst grouped data derived from this material will be transmitted to the DHSS. The preceding section indicated the range of data that are collected at the basic activity levels in the district health services; the first part of the Appendix provides a list of all the returns that are currently transmitted to the DHSS. The main challenge is to bring together these sets of data currently handled in isolation: at present the executive council, hospital, and local health authority material is dealt with by different staff in separate fields of work. Even within one branch of the service it is very difficult to relate cost, staffing, workload, and outcome data to the care of different patient care groups. Co-ordination will have to incorporate a measure of control of the development; data collected in the districts will have to be processed by a common computer system without duplication of effort in programming. Variation to meet local information requirements must not vitiate the ability to compare performance between districts.

Time is desperately short in order to set up a co-ordinated information system adequate to translate into practice some of the hopes expressed in the first section (p. 96). That section indicated in a broad brush fashion the different functions at district, area, and regional level and their associated requirements for management information. There is a major short fall between the information indicated in the second section (p. 104) as required for management and planning, and that currently available (p. 110). Extension of the data base is thus required in addition to a major review and co-ordination of current procedures. Alderson (2) has indicated the way in which the data base should gradually extend in order to improve the availability and use of management information in the Health Service.

ENVIRONMENTAL DATA

Another problem at area level will be the handling of information about environmental problems. Some of these data are now handled by the local health authority,particularly by the Public Health Inspectorate working under the MOHs. There is a responsibility for monitoring air pollution, environmental noise, quality of water, sanitation, housing conditions, and food hygiene. Currently this informa-

tion is collected and handled at the level of sanitary authorities (ie down to the level of rural districts). Under reorganization this information will still be handled by the local authority; the DCP, who will also be acting as medical adviser to the local authority, will play a part in monitoring this information. It is important that some of these environmental data are fed into the NHS information system so that account can be taken of them in identifying specific problems in different localities. In addition to the responsibility of the local authority, other agencies currently monitor the environment, particularly with regard to emission from factories; these data can be handled by the Alkali Inspectorate who work quite independently of the Public Health Inspectorate and MOHs. Presumably there will be changes instituted in the near future to co-ordinate monitoring of the environment, as this is of growing concern.

LIAISON WITH DEPARTMENTS OF SOCIAL SERVICE

At district level it is important that there is good communication between the NHS and certain branches of the local authority. Particularly this will relate to the work of the departments of social service who have a major responsibility for child care, care of the mentally ill and mentally handicapped, care of problem families, and care of the elderly (one must also remember the work of the probation service, which currently has not been amalgamated with other social services following the Seebohm Report). There will be the dual problem of translating confidential information required in client/patient care, but also the need to exchange information between the Health Service and the departments of social service for the purposes of management and long-term planning.

OCCUPATIONAL HEALTH

The occupational health services are not to become part of the NHS, but an Employment Medical Advisory Service has been set up. Clinical care of some individuals in employment may still be carried out by the equivalent of the appointed factory doctor, who may presumably be a family practitioner working part-time in industry or a full-time doctor in occupational health. Apart from the problem of exchange of clinical data at the patient level there will be need to transfer information to and from the CP probably at district or area level.

SCHOOL HEALTH SERVICE

The Health Service is to have responsibility for the school health service; this will obviously create considerable problems in the transfer of information with regard to healthy, sick, and handicapped schoolchildren. The Reorganization Bill stresses that ascertainment from the educational point of view will be retained within the local education authority, and such ascertainment is closely bound up with medical advice particularly for some of the categories of physical and mental handicap. Some system of liaison between the family practitioners, the hospital service particularly the paediatric departments, the school health service, and the schools will be required. In addition one must consider the transfer of information through to the departments of social service, who will have a responsibility in the care of special subgroups of the children. There will obviously need to be some system for transferring information about call-up for examination, results of examinations, and recall for subsequent surveillance of those found to be abnormal in any way.

SICKNESS ABSENCE

Information on sickness absence can be used in a crude way to warn the district health services of epidemics of flu (there are warnings for a 150 per cent and 200 per cent rise in sickness absence claims from one week to the next). It will be necessary at some stage to examine the transmission of data on sickness absence within the NHS so that monitoring of the situation can be carried out by the CP; the possibility of using this information as a guide to morbidity in the population, and the efficiency of rehabilitation following major illness requires further exploration. There is a major problem of confidentiality of the medical records, but matching sickness absence data with that relating to spells of in-patient care can provide an index of outcome of medical treatment.

VITAL STATISTICS

Ultimately it would be a great advantage if the material collected through the vital registration system was transmitted to the central government authority via the information unit at district level. Such a change requires very careful examination of the advantages, and the problems in implementing such change. It would be useful to have up-to-date information on births, particularly to provide the denominator for rates of home confinement, perinatal mortality, etc. Data on marriage and death can be of use in correcting patient registers

whilst the mortality data themselves may warrant regular study at district, area, or regional level. The change would, however, require change of the relevant Acts (the Marriage Act 1949, the Registration Service Act 1953, and the Births and Deaths Registration Act 1953). It is mentioned here as an ideal towards which the information system might move; until such a change is implemented, it is suggested that all the data handled by the registrars at district level are made available to the information system, either through direct transfer of copies of the original data, or through transfer back of the data via the central department.

Computers and information handling

REGIONAL ACTIVITIES

Currently the majority of the computer capacity available to regional hospital boards is not used in handling information directly related to 'management' in the NHS. This is because many of the initial applications taken on to their machines were finance and payroll work, for which these machines are particularly suited. However, the amount of medical and other management data being processed on regional hospital board machines is increasing. The report *Using Computers to Improve Health Services* (12) noted that eight RHBs have their own computer installation, four share two computers, and the other two are served by the hospitals computer centre for London. In 1971 the DHSS announced a policy of standardization of RHB configurations (2 August 1971; A/C376/01); this has been followed by a public announcement in 1973 confirming a bulk order for the purchase of standard configurations for the RHBs in the country. (On 15 January 1973 the Secretary of State for Social Services met ICL to mark the occasion.)

The report *Using Computers to Improve Health Services* (12) suggested that the RHA computers on reorganization might be responsible for batch processing services to their health care unit with particular emphasis on commercial and management applications, but also for processing the applications now run by local authorities. Because there will be no other computers available throughout the NHS on reorganization, the transfer of this LHA computer work seems inevitable, on to the RHA configurations as soon as they are able to absorb the volume of work. This has been examined in Wessex, and a report accepted by the Regional Joint Liaison Com-

mittee has proposed that all financial work taken over from the LHA should be run on the regional configuration from April 1974, and that all hospital medical statistics for those hospitals coming into the new region should also be handled on the regional machine. As soon as capacity is available (on acquisition of the standard configuration), the health applications now run by the local authorities throughout the region should be taken on to the regional machine; in the meantime the local authorities are being asked to continue this work on a bureau basis. It is likely that similar schemes will be operating in the other regions in the country. This indicates that the only consistent computing throughout the NHS upon reorganization will be taking place at regional level as and when the standard configurations are acquired and the capacity is available for this work.

DISTRICT ACTIVITIES

The management information that is currently processed on RHB machines is obtained chiefly from the form-filling exercise that is considered at local level a low-priority chore. This means that there are problems with the completeness and accuracy of the data, and some delay with the capture of the material. *Using Computers to Improve Health Services* discusses the current DHSS projects, with funded work occurring at hospital level. This mentioned the three operational medium to large-scale computer projects, and other medium-scale computer installations that were planned. It also commented on the number of small computers that were already in service, some fifty in number, processing a wide variety of laboratory work, operational and research calculations, and data analysis. An appendix gave details of the current computers operating in the NHS. This work is predominantly hospital-based; one DHSS project, the Exeter project, is district-based and is looking at communication between the community services and the hospital. Some further projects are under development, with attempts being made to link local hospitals with regional board configurations.

Wessex RHB have ordered two Modular One computers for installation in the Basingstoke and Southampton districts. It is suggested that such computers will be used to handle out-patient appointments, recall of patients to out-patients, waiting-lists for in-patients, and a day-to-day file of current in-patients; linked to this may be a request and reporting system for some of the service departments including chemical pathology, haematology, bacteriology, and in due course

radiology. A system for scheduling nurse deployment also requires consideration. To someone concerned with information systems, the chief interest of the introduction of such computers is not the fact that they may help to schedule patients through their contacts with the hospital service, but that as a by-product data may be collected about the functioning of the health service. This should prove an economical way of capturing statistical information, with the emphasis on avoidance of duplication of effort in data capture, and the multiple use of all stored items. Because the scheduling system has to treat patients as people (and not just record isolated events such as attendances at out-patients and discharges from hospital beds) it should be possible to accumulate data from which may be derived reattendance rates at out-patients, admission rates from out-patients, and readmission rates to hospital; this information can be classified by consultant, diagnosis, and patient characteristics such as age and sex. With a functioning request and reporting system, it is possible to relate the consultant and patient data to use of the service departments. The nurse scheduling application should provide information about use of nursing staff in relation to workload in different units in the hospital. Thus the system should build up a profile of demand, workload, outcome, and use of staffing and other resources.

The local installations at district level will be associated with the capture, editing, and day-to-day use of this basic information. Careful examination is required of the optimum way of processing and presenting management statistics from the files of historical data that will be accumulated. It is likely that some analyses (such as statistics on referrals to out-patients, waiting-list size, use of beds and theatres, or use of service departments) might be produced at monthly intervals for examination within the relevant divisions of the Cogwheel system. However once one contemplates interrogation of extended files and production of more detailed analyses with comparisons against other districts in a region, it is likely that it will be more efficient to use the regional computer installation for this work. There will therefore be a need for the facility to transfer the basic edited data to regional files; these may have to be handled alongside other material coming from districts without a local computer system.

HEALTH SERVICE COMPUTING IN THE FUTURE

It is likely that the RHA configurations will require to be complemented with development of district facilities. These may vary from

simple terminals, remote job entry terminals, or local computers. This equipment should be linked to the regional configuration, and a communication network developed, so that historical data and annual comparisons can be processed at the regional level. There will be a continued need for processing at regional level of data for local applications, run in batch-processing mode with an acceptable courier service for regular transmission of data for scheduled and *ad-hoc* work. This may be complemented by data transmission links, sending data quickly for overnight processing. Other data transmission links may be developed for immediate access to time-sharing computer installations. In the next few years there will only be the standard regional configurations, and limited development of computing at district level; computers, even medium-sized, are expensive, and the capital programmes of health authorities do not change rapidly, and allow for a sudden expansion in the acquisition of computing. The balance between district and regional computing may shift with the testing and development of fresh applications that radically alter the volume of computing, or the response time required. This agrees with Knox, Morris, and Holland (14) who suggested that computer facilities should be available at all regional and some district levels, and that dispersing health information to area level was neither practical nor desirable.

Earlier, this paper (p. 109) touched on the need to collect data illuminating the functioning of the GP services. This information is required partly to throw light on the incidence and prevalence of disease in the community, but also to enable some examination of the different ways in which family practitioners work. Alderson (1) indicated how a scheduling system for particular groups of patients in family practice might be set up as a justification of the introduction of computing into the primary medical care field. It is not suggested that in the next decade such work should advance from initial studies, to implementation through all practices in a region. If the relatively high volume of information about the functioning of the hospital service is to be complemented by some indication of the health problems and workload carried in general practice, some system is required at least on a sampling basis. A skeleton framework, based on selected practices for whom an age and sex register is held, would enable detailed surveys to be mounted to investigate some of the particular problems faced by the unified Health Service. Only if such work as the establishment of master patient registers are introduced in the future, as

suggested by Bodenham and Wellman (8) might the location of computing require to be reviewed. This is, however, a long-term development (3). It is therefore suggested that current short-term planning should be for the development of information systems linking the district to regional configurations, perhaps with the steady expansion of district facilities whether locally based or communicating with the regional machine. Only when further work has been done on clarifying other suitable applications within the NHS, will there be adequate guidance to define the long-term deployment of computers at regional, area, and district levels.

The location of computing power now and in the coming few years has been spelt out in this section, but this should not be taken to imply that the location of the hardware determines the development of the information system. The capture of the data, particularly when operating batch-processing applications can be carried out quite irrespective of the specific location of the computer (providing the courier service can transmit the data). Only when data transmission through telex, linked terminals, or remote job entry terminals commences does the disposition of the mainframe and peripherals become more important. When on-line capture of data is occurring (with immediate data edit by the computer system) the deployment of equipment must be intimately linked with the operational requirements of the system.

At the present stage of development of information systems the careful control of current data capture, the consideration of extension of the data base, and suggestions regarding use of the data base are the more crucial aspects determining success. It is more important to consider how the information system within a region can develop, striking the appropriate balance between flexibility to meet particular local requirements, and the opposing advantages of standardization. The need for liaison between the complete range of users of the information (at unit, district, area, region, and central government level) is paramount in the successful development of the system; the next section indicates how development may take place in the coming years in order to achieve an information system of use to management at all these levels.

Suggested organization of information handling post April 1974

REGIONAL ACTIVITIES

It is suggested that in each region there should be a medical information unit with responsibility for co-ordinating the development of information systems throughout the region. There are three main reasons for advocating this; first it is essential that these systems are relatively standardized, and this requires co-ordination at regional level. Secondly, the development of fresh systems requires the involvement of staff who are in relatively short supply, and who would not be available at district and area level: particularly systems analysts and statisticians. Thirdly, the section on computing indicated that, for a number of years to come, the main site of data processing will be at regional level. The information unit at regional level should be responsible for the development and organization of the information systems, the statistical treatment of the data handled by the regional system, and for the provision of interpretation of this material. All the systems that are developed should be closely tailored to the users' requirements; this requires close collaboration with users at district, area, regional, and national level over the wide range of uses that have been referred to in the section on management and planning within the unified Health Service.

AREA ACTIVITIES

As discussed in the previous section no processing responsibility is envisaged at area level, once the RHA standard configuration has taken on the work currently done for local health authorities. The area staff are seen as one category of users of information; though some of this information may come through to them directly from their districts, the bulk of this will be provided in digested form via the regional information unit. A lot of the material for routine monitoring at area level will consist of comparative data for all the districts in a given area, against pooled data for the region as a whole. It is suggested that the regional information unit will have a major role to play, partly using the facilities of the regional computer; analyses prepared from the computer may require statistical interpretation in the regional unit prior to their being forwarded to area. A service for data processing should be provided not only for routine information, but also for data specially collected even at unit or district level. In

the next few years there is unlikely to be adequate provision of statistical staff at district or area level to carry out data manipulation, and thus the information unit at region will have to carry out this service role.

DISTRICT ACTIVITIES

For any new application there should be a clear definition of the items to be used in the system, and this can only be drawn up after close consultation with DCPs, who will be aware of the local situation with regard to the problems of data capture. It is suggested that the DCP shall have an important role at district level for supervising the capture of data: as has already been indicated the vast bulk of the material handled in the information systems will be captured at unit or district level, thus requiring supervision and quality control in the district.

Apart from any districts which may be involved in the development of district-based computer projects there will be minimal facilities available for processing information at this level, beyond that required for clerical or some simple mechanical handling of data. It is to be hoped that such clerical and mechanical handling of data can be kept to a minimum, providing an adequate service by the regional processing facilities is available. At district level the staff would supervise the collection of the basic data and play a part in interpreting information gathered locally (even though this may be processed regionally). In addition the staff would play a very active part in formulating and requesting specific analyses from the regional processing unit; they would also be heavily involved in the preparation and dissemination of routine information in relation to specific problems under discussion by the DMT, and in the Cogwheel structure.

Earlier sections (pp. 98 and 99) indicated the information requirements of the DMT, and the role of the DCP as provider of this information. There are two aspects of this work that require emphasis here; first is the need to bring together the routine data currently handled by the three branches of the NHS. Second is the need for close formal links between the DCP and the regional information unit, which will be handling information and controlling the development of new systems on behalf of the DCP. Ashford (4) has stressed that as far as information for management is concerned, the district is the crucial level for organization, due to the need to collect detail on operational matters.

THE FLOW OF INFORMATION

The Appendix indicates the types of health service data that have to be collected and transmitted from district through region to the DHSS. These sets of data will form the core of the routine information; the chart (pp. 134–6) shows the flows of this health service data, and also the flows of other information relevant to examination of the health problems of the community. The chart indicates at district and area level a possible interrelationship between the NHS and other organizations handling relevant data. It should be emphasized that this chart is of the flows of statistical information, and does not cover the direct transfer of information relevant to day-to-day patient care. (Such information will, of course, continue to be transferred directly between those responsible for joint care of individual patients, for example, transfer of information between a doctor working in the school health service, and an educational psychologist.) If the flow of information indicated in the chart is to be implemented, with the development of an hierarchical system funnelling information through district to region and back to all the users there will be a need in the lead-up period to unification to manipulate, to a limited extent, some of the current channels of transfer of information. A considerable impact on the availability of information at district, area, and regional level can be obtained merely by co-ordinating the collection and handling of information that is already (theoretically) available. Having set up such a basic system attention should then be paid to amplifying the extent of the material handled in the information system; progress with this will be relatively slow due to the time required for careful systems study, appraisal of user requirements in relation to resources for information handling, and the development of agreed systems.

It is a fundamental tenet of a successful information system that each item that is collected as a routine should have been included because of clearly identified use. The list in the Appendix is long, but this gives no precise indication of the specific variates included in each return, nor the relationship of these items to (*a*) the district, area, and regional information requirements, and (*b*) the ease of data capture in relation to operational activities at local level. This is an aspect of the information system that requires continual review, and clear thinking. The continued collection of every item must be justified after consideration of the costs of data capture, the accuracy of the data, and the use to which it is put.

STAFFING REQUIREMENTS

Until the general principles outlined in this document are accepted, no specific guidance can be proposed about the numbers of staff that will be involved in handling information. It is appropriate to point out however that already at local authority level the ECs have staff in their registrars' departments handling statistics; in the hospital service there are staff involved in this kind of work in medical records departments, and HMC staff supporting the Cogwheel structure. The LHAs have statistical sections involved in handling their own material. It is suggested that these staff should ultimately be brought together at district level in order to provide a statistical section supporting the CP (p. 99 has indicated the DCPs' responsibility in this field). At area level there will be no staff required for data processing, because this work should be done either at district or regional level; however there will be the need for some assistance for the CP, who will have a particular responsibility for collating and presenting information to the ATO. It would be appropriate to consider the provision of two statistical assistants (perhaps at higher clerical officer level) in the area to support the CP involved in this work. The information unit at regional level will bear an additional burden, when responsible for handling information from the EC and LHA branches after these are brought into the unified Health Service. This work will not add greatly to the coding load, but will add to the processing, and this will have to be carefully reviewed; the main task will be initially systems studies to clarify the optimum way to handle this additional material; perhaps additional staff will be required to assist in the statistical analysis and interpretation of the material.

The Wessex Medical Information Unit has recently collaborated with the Regional Computer Centre in transferring the routine SH3 returns on to the computer. This has necessitated careful checking of the interpretation of definitions of terms at local level, has speeded up input of material, and has provided an automated collation and analysis of these data. A number of the other returns listed in the Appendix require careful consideration; where the clerical chore of collation is appreciable, and comparative analyses are required between districts, the need for automated processing should be reviewed.

THE NEED FOR A CO-ORDINATING COMMITTEE

This section began by emphasizing the need to establish a co-ordinated system for handling information across regions; one step to-

wards this could be the establishment of an 'information systems co-ordinating committee'. It is suggested that this might consist of a CP from each district, the three area CPs who have the particular responsibility for information, together with representatives from the regional information unit (staff in such a unit are likely to have a particular responsibility in the three fields of information handling, statistics, and computing). It would be appropriate to co-opt on to this co-ordinating committee representatives of the treasurer's and personnel departments and an administrator from a DMT, an ATO, and the RTO. These representatives should provide some indication of the user requirements, other than that directly obtained through the CPs. The power to co-opt should of course extend to the practice of bringing into meetings staff concerned with particular aspects of information use that are to be discussed by the committee.

ACKNOWLEDGEMENTS

This paper stems from work in progress in the Wessex Medical Information Unit; a lot of this development relies on the ideas and hard work of my colleagues in the Unit, especially Michael Slattery and Ken Goulding. The Wessex Regional Hospital Board and the DHSS jointly fund the Unit, whilst Dr John Revans, the Senior Administrative Medical Officer, and Mr John Shaw, the Board Secretary, provide continual encouragement for our work. We also rely on the comments and advice of many throughout the region and in the DHSS. My personal secretary, Mrs Sharon Pointer, patiently deals with all the work involved.

REFERENCES

1. ALDERSON, M. R. (1972). 'The Wessex plan for computing', in Abrams, M. E. (ed.), *Spectrum 1971—A conference on medical computing* (London: Butterworth).
2. —— (1973*a*). 'Towards a health information system', *Hlth and Soc. Serv. J.* (in press).
3. —— (1973*b*). *Master Patient Registers.* Wessex Medical Information Unit internal memorandum.
4. ASHFORD, J. R. (1971). 'The uses of information within the health services. The challenges of technical change', in McLachlan, G. (ed.), *Challenges for Change. Essays on the next decade in the National Health Service* (Oxford University Press for the Nuffield Provincial Hospitals Trust).
5. —— FERSTER, G., and PETHYBRIDGE, R. J. (1973). 'Information requirements for district management in the reorganised NHS', *Commun. Med.* **26**, 309–17.

6. BENJAMIN, B. (1971). 'The use of statistics in the management of health and welfare services, in Gatherer, A., and Warren, M. D. (eds), *Management of Health Services* (London: Pergamon).

7. BISPHAM, K., HOLLAND, W. W., and STRINGER, J. (1971). 'Planning for health', *The Hospital*, **67**, 82.

8. BODENHAM, K. E., and WELLMAN, F. (1972). *Foundations for Health Service Management, A Scicon report for the Scottish Home and Health Department on the requirements for a health service information system*, Occasional Hundreds 3 (Oxford University Press for the Nuffield Provincial Hospitals Trust).

9. BURBRIDGE, D. H. D. (1973). 'National Health Service: the philosophy of change (England)', *Hlth Trends*, **5**, 22–23.

10. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1972*a*). *National Health Service Reorganisation: England*, Cmnd 5055 (London: HMSO).

11. —— (1972*b*) *Report of the Working Party on Medical Administrators* (Hunter Report) (London: HMSO).

12. —— (1972*c*). *Using Computers to Improve Health Services: A review for the National Health Service* (London: DHSS).

13. —— (1972*d*). *Management Arrangements for the Reorganised National Health Service* (London: HMSO).

14. KNOX, E. G., MORRIS, J. N., and HOLLAND, W. W. (1972). 'Planning medical information systems in a unified Health Service', *Lancet*, **ii**, 696–700.

15. MATTHEW, G. K. (1971). 'Measuring need and evaluating services', in McLachlan, G. (ed.), *Portfolio for Health. The role and programme of the DHSS in health services research, Problems and Progress in Medical Care*, Sixth Series (Oxford University Press for the Nuffield Provincial Hospitals Trust).

16. NAYLOR, W. M. (1972). *Organisation of Area Health Services* (London: Institute of Health Service Administrators).

17. SCOTTISH HOME AND HEALTH DEPARTMENT (1971). *Doctors in an Integrated Health Service* (Edinburgh: HMSO).

Appendix

List of returns currently forwarded to the DHSS

1. EXECUTIVE COUNCIL RETURNS

The following list identifies the routine returns that are completed by each of the executive councils, and forwarded to the DHSS. The returns from each executive council consist of a collation of information from the various units carrying out work in the executive councils; thus the returns on ophthalmic workload will be consolidated on executive council level for each optician and ophthalmic practitioner providing services in the area.

Medical staff returns

DI 1	Eighty column punch cards giving basic particulars about family doctors.	(amendments quarterly; details annually)
DI 2	Doctors index entry form (to be completed for change of index).	(quarterly)
DI 3	Doctors index amendment control sheet.	(quarterly)
DI 4	Particulars of assistant medical practitioners.	(at 1 October . . .)
DI 5	Doctors approved as trainers, and particulars of trainees.	(at 1 October . . .)
DI 5a	Doctors undergoing training under EC127/66.	(at 1 October . . .)
DI 6	Doctors receiving initial practice allowance.	(at 1 October . . .)
DI 9	'Salary' and 'small-share' partners.	(at 1 October . . .)
DI 12	Principals for whom council became responsible.	(quarterly)
DI 13	Principals for whom council was responsible throughout quarter, but payment changed.	(quarterly)
DI 14	Locums deputizing for doctors.	(at 1 October . . .)

DI 15	Locums filling vacancies, pending appointment.	(at 1 October . . .)

Other staffing returns

SBL 634	Return on staff in posts.	(at 30 June . . .)

Ophthalmic work

SBE 515	General ophthalmic services workload.	(quarterly)
EC 48	General ophthalmic services return of practitioners and premises.	(annual)
EC 48C	Particulars of chemists' drug stores and appliance contractors' premises.	(annual)

Cost returns

AG 501	Cash statement for current period.	(monthly)
AG 501a	General Medical Services supplement to form AG 501.	(quarterly)
AG 501b	Pharmaceutical Services supplement to form AG 501.	(quarterly)
AG 501 (imprest)	Application by executive council for funds.	(monthly)
AG 505	Receipts and payments.	(annual)
AG 505a	Statements of account and analysis of payments.	(annual)
AG 505b	General Medical Services supplementary statements.	(annual)
AG 505c	Summary of losses, etc., written off.	(annual)
AG 512	Estimate for next financial year, and revised estimate for current year.	(annual)
AG 540	Rural practice payments.	(quarterly)

2. HOSPITAL RETURNS

The following list identifies the returns completed for the hospital service and forwarded to the DHSS. Many of these returns are collated at clinic and hospital level; the hospital groups are responsible for producing the annual returns copies of which are forwarded to the regional board, and to the DHSS. Some of the forms relate to staffing, and these will be completed by the appropriate staffing authority which may be the regional board, or the group hospital management committee.

Medical, dental and nursing staff

SBH 57	Hospital medical and dental staff notifications of staff changes in grades of senior registrars and above.	(for each staff change)
SBH 50	Hospital medical and dental staff registrars and below in post.	(at 30 September . . .)

SBH 51	Hospital medical and dental staff vacant posts for registrars and below.	(at 30 September . . .)
SBH 54	Return of administrative medical staff of regional hospital boards and Welsh Hospital Board.	(at 30 September . . .)
SBH 2	Hospital nursing and midwifery staff return.	(at 31 March . . .)
SBH 2b	Senior nursing and midwifery staff with Salmon gradings in post.	(at 31 March . . .)
SBH 155	Hospital nursing staff receiving the 'geriatric lead'.	(at 31 March . . .)
SBH 101	Nursing publicity survey.	(quarterly)

Other staff

SH 7	Regional hospital board senior administrative, architectural, engineering, and administrative nursing staff return.	(at 30 September . . .)
SH 6	Hospital administrative and clerical staff return.	(at 30 September . . .)
SBH 56	Hospital service computer department staff return.	(at 30 September . . .)
SBH 153	Student radiographers.	(at 31 December . . .)
SBH 154	Return of teaching staff and students of hospital schools of physiotherapy.	(at 31 December . . .)
SBH 5	Hospital professional and technical staff return.	(at 30 September . . .)
SH 13 (part 1)	Hospital works and maintenance staff return.	(at 30 September . . .)
SH 13 (part 2)	Hospital domestic portering, catering, and stores staff return.	(at 30 September . . .)
SH 13 (part 3)	Hospital farms, gardens, and ground maintenance laundry and miscellaneous staff return.	(at 30 September . . .)

Workload

SH 3	Hospital return of in-patients, out-patients, and workload in other departments.	(for year ending 31 December . . .)
SBH 1	Hospital return of in-patients, out-patients attendances.	(six monthly)
SBH 179	Chronically sick and disabled persons admitted and resident in hospital.	(six monthly)
T 145	Tuberculosis patients under supervision and new attendances.	(annual)
SBH 140	Cervical cytology; smears examined and number positive.	(six monthly)
SBH 60	Sexually transmitted diseases; new cases.	(quarterly)
SBH 13	Mental health statistics; legal status of patients on admission, leaving and changes.	(monthly)

SBL 626	Numbers of patients awaiting entry to hospital, admitted for temporary residential care or admitted for guardianship.	(annual)
SBH 6	Pathology statistics; workload of laboratories.	(annual)
SBH 62/A	Work measurement in physiotherapy department's patient record sheet.	(six monthly)
SBH 65	Points system of assessing radiologists' workload.	(six monthly)
SBH 19	Hospital eye service workload.	(annual)
SBH 136	Intermittent haemodialysis for chronic renal failure; staff in unit, and patients treated.	(six monthly)
	Mass radiography examinations 10 per cent sample and record of abnormal cases.	(monthly)
NBTS 47	National Blood Transfusion Service workload.	(monthly)
SBH 168	Miners' rehabilitation service; workload	(annual)

Facilities

SBH 112	Return of psychiatric facilities.	(annual)
SH 9	Patients receiving treatment under contractual arrangements; workload by specialty.	(annual)

Costs	Summary of unit costs for financial year.	(annual)

3. LOCAL HEALTH AUTHORITY RETURNS

The following list identifies the returns completed by the local health authorities, and forwarded to the DHSS. The returns are collated by the local health authority from information provided on an ongoing basis by the operational units at local level.

Staff

LHS 27/6	Medical and dental staff in post.	(at 30 September . . .)
LHS 27/8	Nursing staff in post.	(at 30 September . . .)
LHS 27/9	Ancillary and nursing staff employed in support of the nursing services.	(at 30 September . . .)
SBL 620	Ambulance staff (all grades) in post.	(at 30 September . . .)
SBL 624	Miscellaneous staff in post.	(at 30 September . . .)
SBL 653	Attachment of health visitors and home nurses.	(at 30 September . . .)
SBL 605	Medically qualified staff in the local authority service.	(at 30 September . . .)

Workload

LHS 27/1	Number of live and stillbirths, premature live and stillbirths by weight and span of life.	(annual)
LHS 27/2	Number of ante- and post-natal examinations, number of sessions, and clinics.	(annual)
LHS 27/3	Number of cases and type of cases visited by health visitor, home nurse, and attended by domiciliary midwife.	(annual)
LHS 27/5	Number of beds, cases admitted, and duration of stay in mother and baby homes.	(annual)
LHS 27/7	Dental treatment of children, expectant and nursing mothers.	(annual)
SBL 603	Return of smallpox vaccination of persons aged under 16.	(annual)
SBL 607	Vaccination other than smallpox of persons under age 16 completed during year.	(annual)
SBL 611	Visits to recent immigrants during year.	(annual)
SBL 618	Chiropody treatments.	(annual)
SBL 655	Tuberculin test and BCG vaccinations.	(annual)
SBL 667	Family Planning Service; new patients, total attendances, and sessions held.	(annual)

Facilities

LHS 26/4	Number of day-nurseries and places; number of registered child-minders and premises.	(annual)
SBL 630	Mental nursing homes registered.	(at 31 December . . .)
SBL 631	Mental nursing homes authorized to detain patients.	(at 31 December . . .)
SBL 669	Family Planning Services; facilities provided, scope of services, and payments.	(annual)
SBL 685	Registered nursing homes.	(annual)

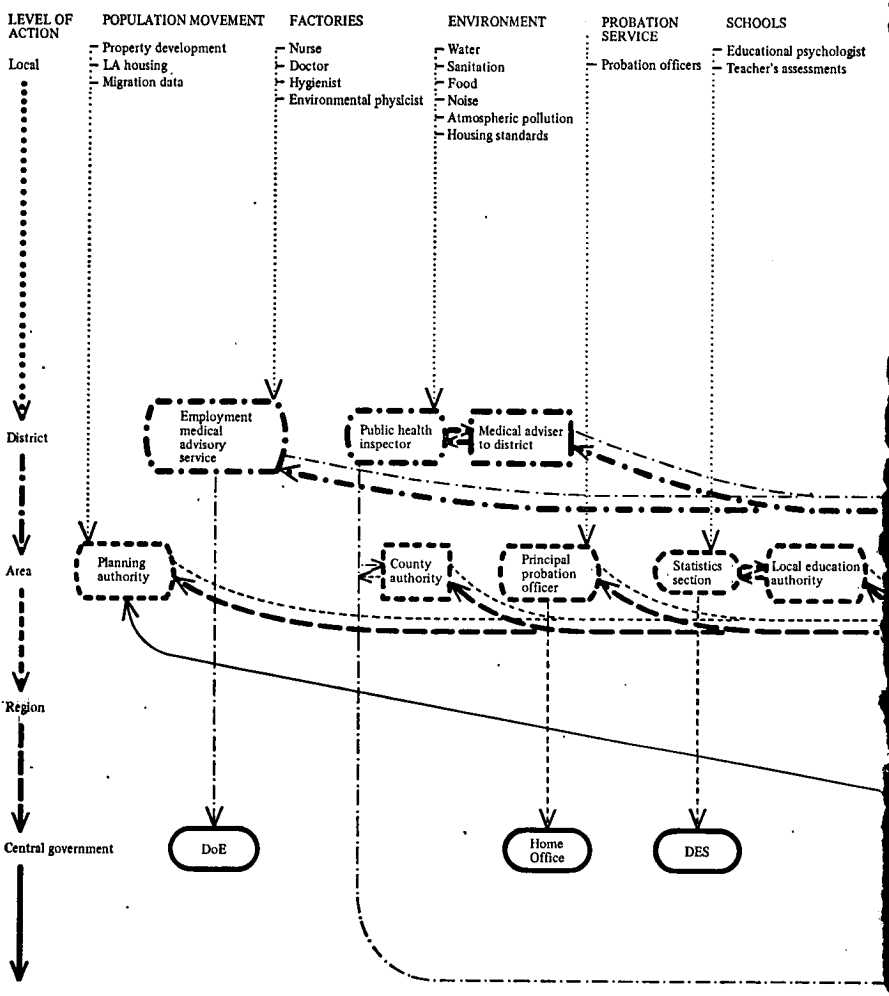
Cost

SBL 637	Ambulance service, cost statement for financial year.	(annual)
AG 213	Return of estimated expenditure.	(annual)
RO 3A/3B	Return of actual expenditure for previous financial year.	(annual)
PCP 1	Three-year programme of local health authority building projects.	(annual)

Other

SBL 640	Return for food poisoning for year, including all *Salmonella* infections including dysentery, paratyphoid, and typhoid.	(annual)

INFORMATION FLOWS IN THE UNIFIED HEALTH SERVICE



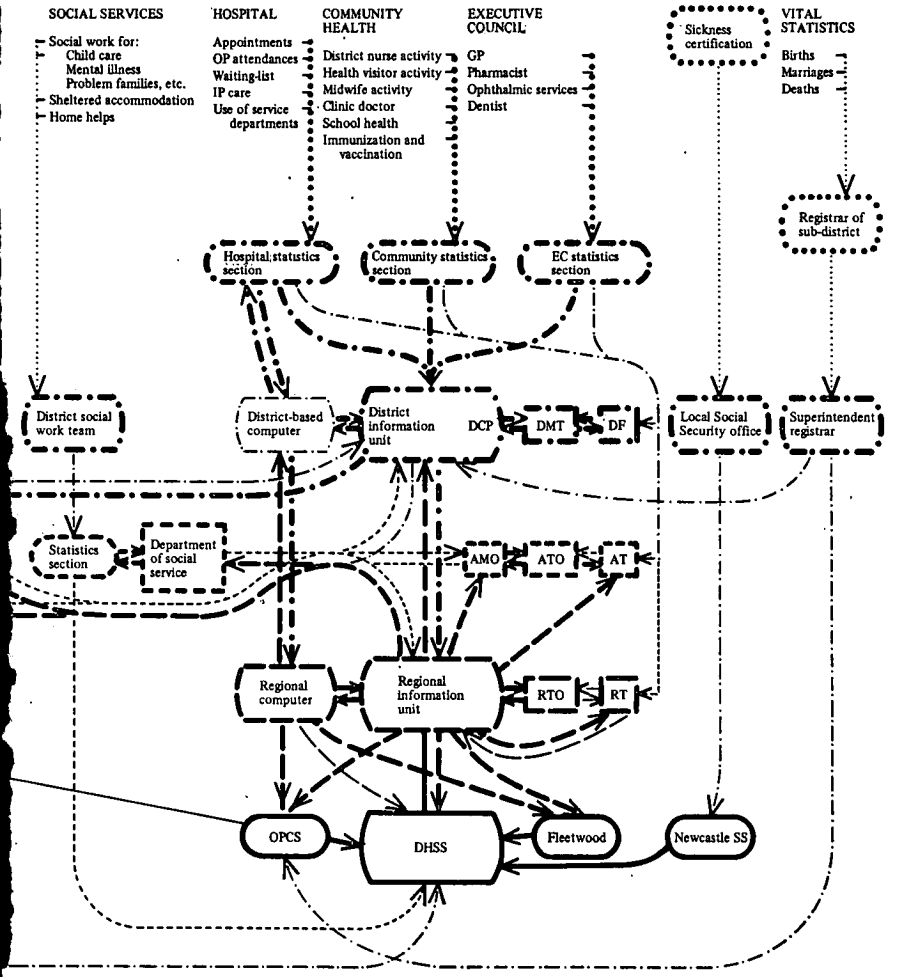

SOCIAL SERVICES
Social work for:
Child care
Mental illness
Problem families, etc.
Sheltered accommodation
Home helps
HOSPITAL
Appointments
OP attendances
Waiting-list
IP care
Use of service departments
COMMUNITY HEALTH
District nurse activity
Health visitor activity
Midwife activity
Clinic doctor
School health
Immunization and vaccination
EXECUTIVE COUNCIL
GP
Pharmacist
Ophthalmic services
Dentist
Sickness certification
VITAL STATISTICS
Births
Marriages
Deaths
Registrar of sub-district
Hospital statistics section
Community statistics section
EC statistics section
District social work team
District-based computer
District information unit
DCP
DMT
DF
Local Social Security office
Superintendent registrar
Statistics section
Department of social service
AMO
ATO
AT
Regional computer
Regional information unit
RTO
RT
OPCS
DHSS
Fleetwood
Newcastle SS

CHART ON INFORMATION FLOWS IN THE UNIFIED HEALTH SERVICE

This chart is a schematic representation of the flows of information from operational units within the districts up to central government level. The main block on the chart with thick lines indicates the flows of 'medical data' from activity level through the district and regional information units. Feedback of information occurs from the regional unit to area and district levels, as well as onward transmission to central government. As well as data on demand, workload, use of resources, and outcome, the health service will collect other sets of data. One of these, financial, is indicated by thin lines from local level to the treasurers at district, area, and region.

The chart also indicates how other data may be collected on demographic, health, and welfare issues; these are relevant as district, area, and regional level in examination of the health care needs of a total population. Transfer of this information into the medical information system is indicated at the top of the chart. It should be emphasized that this is transfer of 'statistical' information, and does not represent the direct communication between workers of different disciplines caring for individual patients or family problems (such as the direct contact between health visitor, school health doctor, and educational psychologist in the assessment and handling of an individual child).

The bulk of the data are captured at operational level within districts, and in order to try and distinguish the flow of information at different levels five types of lines have been used on the chart.

Transfer of data from the local level (at which it is captured) is indicated by: .....

Transfer of information within or from the district is indicated by: -.-.-.-.-.-

Transfer of information at or from area level is indicated by: - - - - - - - - -

Transfer of information at or from regional level is indicated by: — — — — — —

Transfer of information at central government level is indicated by: ———————

A district-based computer is shown on the chart. As explained in the text this is thought to be the most likely site for computers other than at regional level. They will be primarily for applications directed at improved patient care, but capturing as a by-product management information. Implementation of such machines will alter the structure of the regional information system, and the chart shows the alternative pathways with and without a district computer.

5

The engagement of cogwheels needs special gearing

Information for the management of clinical work in hospitals

J. M. Yates
AHA
Medical Information Officer,
United Birmingham Hospitals

Information for the management of clinical work in hospitals

Summary. *The proposals for a new administrative structure for the National Health Service have been accompanied by discussion about the scope of the management function in the health care system. At the same time discussion has taken place on the role of the clinician in the management of the health service. In both these discussions the underlying aim is seen to be the more rational determination of priorities in medical care and the better use and control of resources, and in both cases emphasis is placed on the need for much improved information to support the management process (1, 2, 3, 4, 5).*

This paper examines the role of information in the management of one aspect of the health care system: clinical work in hospitals. An examination is made of the development of management in relation to clinical work in hospitals and of the role of information in that process. An attempt is made to clarify the relationship between information and management and to show the results of the presentation of information to clinicians involved in the management process. Finally, areas where there is a need for further study of the role of information in the management of clinical work in hospitals are proposed.

Management of clinical work in hospitals

During the first twenty years of the NHS two broad categories of medical staff organization existed in hospitals. One consisted of a medical advisory machinery in which all the senior medical staff working in a hospital formed a medical advisory committee. This

committee was the chief form of communication between the medical staff and the hospital authority. In the second form of structure a medical superintendent was responsible for a large proportion of the administration of the hospital and that individual often acted as the link between the medical staff and the hospital authority. In both types of structure the participation of the clinician in the management of the hospital was limited and attempts amongst clinicians to manage their clinical work were not manifest.

The publication in 1967 of the first Cogwheel report (1) marked the first significant interest to be shown in the subject of management of clinical work in hospitals. The report suggested that the existing medical staff organizations were not able to cope adequately with the management of clinical work and had in the past failed to do so. It recommended that allied specialties in a hospital or group of hospitals should be formed into divisions within which clinicians would tackle the problems of management that arise in the clinical field. It was further recommended that representatives of divisions should form a medical executive committee responsible for co-ordinating the work of the divisions and hospital clinical activities. The report suggested a need for greater participation by clinicians in the general management of hospitals and of the management of clinical work in particular. The following four years showed only a limited acceptance of the Working Party's views and in 1971 a second Cogwheel report was published (6). This report looked forward to the universal adoption of systems based on the Cogwheel philosophy and was designed to help forward such movement. The report gave examples of the experiences of some of those hospital groups operating the machinery and offered further guidance on the representative system, methods of co-ordination with other disciplines and services in support of Cogwheel.

Since the publication of the first Cogwheel report considerable discussion has taken place on the management of clinical work in hospitals. Three features of this managerial role are worth considering in further detail.

A NEW ROLE

To many clinicians and other hospital staff the concept of the clinician being involved in the management of hospitals in anything other than an advisory role is new. A clinician's training and the hospital environment are geared to meeting the needs of the individual

patient. The clinician has an ethical, moral, and legal obligation to the individual patient and any other responsibility is seen to be additional to, if not competing with, his prime responsibility. The management of clinical work in hospitals requires the clinician to consider groups of patients and to decide priorities in allocating resources between those groups. Such a task sometimes appears to interfere with the clinician's right to choose what is 'best' for each individual patient and thus the new role is questioned and sometimes rejected.

In the past the managerial role has been attempted by groups of individuals other than clinicians. The clinician has traditionally established a need and then relied upon others to provide the wherewithal for him to meet that need. A situation in which the clinician has had no responsibility for finding or controlling resources is also one where little need is seen for accepting any responsibility or accountability in the use of those resources.

AN IMPRECISE ROLE

Management responsibility at hospital level almost always implies joint or shared responsibility. The clinician is encouraged to take on not only a new role but also one for which he does not have total or defined responsibility. The advice offered to the clinician concerning his new role does not always specify the precise relationship between the medical staff structure and the hospital authority. Doubt exists as to the nature of the relationship between the medical profession and the other disciplines within the hospital. Stress is placed on the clinician's responsibility in the planning of services and the use of resources but aspects of this work are undertaken by divisions, executive committees, other hospital professions, and the hospital authority. If lack of clarity in this field leads to duplication of effort, neglect of important topics or conflict between groups, the clinician may become disillusioned and thus reject the managerial role.

A DIFFICULT ROLE

When the clinician accepts a managerial role he soon discovers that the task is a difficult one. The choice of priorities may affect the progress of his own work or that of some of his colleagues. The financial constraints imposed upon divisions quite often result in there being no reward for effective or efficient behaviour and only too often money saved is used to help those who have mismanaged their affairs and overspent. In his new role the clinician becomes acutely aware

that he is unable to measure and control adequately the outcome of care and treatment given. He wishes to exert control on the outcome but finds that he lacks both the measures and the information that he requires. Decision-making is found to be difficult without adequate information but in seeking information he reveals a new set of problems: collection, accuracy, analysis, and presentation of data. Soon the need for information itself becomes a contending priority for the limited resources available at hospital level, and this is complicated by the fact that it is not an immediately clinical priority.

The role of information for management

The publication of the two Cogwheel reports encouraged clinicians and administrators to consider the management of clinical work in hospitals. At a later date the proposals and discussions concerning the reorganized NHS also stimulated a more general interest in the role of the clinician in the management of the health care system. In both cases it was advocated that good management would not be possible unless the manager was adequately informed. The rest of this paper attempts to examine the role of information and its relationship to the management of clinical work in hospitals. The illustrations are based on experience over four or five years and have been taken from a number of hospitals in different parts of England. To some readers the examples illustrated may seem extremely simple but they do show what can be done by using existing information, and by their very simplicity serve to highlight some of the difficulties in this field.

MANAGEMENT AND INFORMATION

Before looking specifically at the use of information in the management of clinical work in hospitals it is useful to consider on a more general level the relationship between information and management. There have been various attempts to describe the manager's job by management theorists but it has been summarized to include planning, organizing, motivating, and controlling (7). This classical description may be striven for by those working at hospital level but in practice the manager's role is seen to be one of problem-solving and day-to-day decision-making. In either the classical or more practical definition there is acknowledged a need for information to aid the process. A simple illustration of the need for information may be

taken from the clinical field. A GP refers a patient to an orthopaedic out-patient clinic. His letter to the orthopaedic surgeon gives the patient's details, a history of pain in the right hip, and asks the consultant to examine the patient. In forming an opinion the consultant sifts through the information provided by the GP. He examines the patient, takes a history, obtains radiographs, and may request certain pathology examinations. The consultant's decision to operate and implant a hip prosthesis is made with the aid of this information. It is unlikely that such a decision would be made without any information, but there may well have been stages in the process where the consultant would have been prepared to make a decision without obtaining certain information. Information helps to make a better or more soundly based decision. At no stage does information of itself determine the decisions or action to be taken.

In any decision-making process it is not always clear what information is required and there are more often than not problems about the relevance, accuracy, completeness, and timeliness of the information and the need to take into account factors which cannot be precisely described. Ultimately the decision is affected by the manager's judgement and cannot solely be based on information provided.

So far in this paper an attempt to define 'information' has been avoided, but the introduction of the word 'data' requires that a distinction between these two words be made. McRae (8) defines 'data' as any non-random set of symbols and 'information' as data which are used as an input to the decision-making process. Such a distinction becomes helpful when attempting to relate information to management. For the purpose of this paper it is suggested that data presented to the manager, but not used by him, remain data. Data presented to the manager and used become information.

TO PARTICIPATE OR NOT

The simple relationship between information and management described above cannot be immediately related to the management of clinical work in hospitals without considering briefly the nature of the managerial responsibility in a hospital. The management of a hospital is not carried out exclusively by one group of people who are full-time professional managers. A large number of people are needed to participate in the management process for it to become effective, and the amount of time required of these people varies. Some hospital

staff are full-time managers whilst others spend only some part of their time involved in the management process. Some have management responsibilities for the whole of a group of hospitals whilst others have responsibilities for one particular section, ward, department, or function. This somewhat complicated management structure leads to lack of clarity about the respective roles amongst those responsible for management. This inevitably leads to different standards and expectations of management, and there is a wide range between those at one end of the spectrum, who participate fully in the management process, and those at the other end who take no part in it whatsoever. In the management of clinical work in hospitals a clinician is free to choose the nature and extent of his participation.

There appear to be four main groups of reasons why clinicians choose not to participate in the management of clinical work:

1. The clinician can not see any problem and therefore feels there is no need for 'management'.
2. Problems can be seen but solutions are not apparent and thus involvement in management is thought pointless.
3. The clinician is unable to manage either because the situation is outside his control or he finds himself otherwise unable to initiate the action to solve the problem.
4. The clinician is unwilling to become involved in management because he has no time to manage or he does not see it as his responsibility.

It has been suggested that one of the main reasons why clinicians fail to take an active part in management is that they are frustrated by the lack of the information they need to play an effective role. The following illustrations will show that in the first two instances this may be so, but in the latter two other factors determine the clinician's unwillingness to participate.

Cannot see the problem. A medical executive committe for a group of hospitals had taken steps to increase the admission load of one hospital. Part of the reason for so doing was to improve the bed occupancy of that particular hospital which was at a lower level than the medical executive committee desired. Whilst their action was being successfully directed at a particular aspect of the low bed occupancy it was not appreciated that an additional and equally important cause of the low bed occupancy was the serious under-use of beds on three particular wards. Data illustrating this was presented to the commit-

tee and subsequently the problem was analysed and discussions initiated with those specialties concerned. As a result of these discussions improvements in the use of two of the wards were achieved. In one case the improvement was achieved by a reallocation of some of the facilities and in another by a change in admission policy. It should be noted that no action resulted from the presentation of data concerning the under-use of the third ward.

The presentation of data to demonstrate a problem which the clinician has not identified usually produces one of two results. Either he grasps the problem and attempts to participate in its solution or he shows an understanding of the problem but displays an inability or unwillingness to participate in its solution. The presentation of data can clarify the reason for not participating.

Cannot see a solution. The clinical and administrative staff of a hospital agreed that an increased number of emergencies should be accepted. Whilst the principle was agreed by all concerned the precise size of the increase was the subject for debate and disagreement. The views expressed included those of the other hospitals in the area and the differing views of the nursing, medical, and administrative staff of the hospital concerned. It was agreed that the hospital should receive the maximum number of emergency admissions possible without interfering with the existing admission and discharge policies for the hospital's planned admissions. At one time no solution could be seen and clinicians began to withdraw their participation from the management process. At a later stage the existing data concerning the admission and occupancy rates of the hospital were analysed in order to demonstrate the effect of accepting differing numbers of cases. On the basis of this information a revised emergency admission policy was agreed and implemented.

The presentation of data can be used to illuminate the situation and suggest solutions to the clinicians. As in the previous example it can aid the clinician in understanding the problem and obtain his participation. Alternatively the clinician may retreat behind some other reason for non-participation.

Unable to manage. Over a five-year period the in-patient and out-patient facilities of a hospital had been extended and a greater number of patients were thus able to attend that hospital. At the beginning of this expansion one of the clinical service departments was working

very near to its level of capacity because of severe restrictions in the amount of space available to it. The general expansion of the hospital's work left that service department considerably over-worked. An extension to the department was planned in the future but the head of the department was faced with a further three years of extreme pressure. Attempts were made by the head of the department with the aid of a considerable amount of data to reschedule and reduce the workload within the department. His attempts were unsuccessful in that rescheduling produced very minor improvements which in no way solved the problem. The only solution was, therefore, to control the work coming to the department. Discussions with the clinicians concerned showed that the only practical way of reducing the workload was to reduce the number of patients attending the hospital, and this they were not prepared to consider. The departmental head was therefore unable to make any effective contribution to the management of the services.

If a clinician is unable to make a contribution to the management process the provision of data is of little practical value. No matter what the cause of this inability no amount of data will make any difference. If the approach of a chairman to his clinical colleagues so antagonizes the group that no agreement can be reached concerning a distribution of resources no amount of data is going to bring about agreement until a change of attitude on the part of those concerned takes effect.

Unwilling to manage. A division was faced with a problem of the under-use of beds on one particular ward unit. Discussions with the consultants concerned produced no change in the use of the ward. The first reaction was to state that there was no problem. The under-use of beds was denied and it was suggested that on the contrary it was difficult to admit some patients to the ward. Data was produced to demonstrate the very low bed occupancy. The second defence followed immediately: the data were wrong. Time and effort was expended to check the collection and analysis of the data. When it was grudgingly accepted that the data were probably correct and that a problem did exist a third reaction was encountered: the data were irrelevant. It was claimed that there would be a considerable difference between the data collected at midnight and the data collected at a time of the day more indicative of the true ward occupancy pattern. When further studies revealed a marked similarity on that particular

ward between the midnight occupancy pattern and the pattern displayed at other times of the day, the clinicians then returned to their first defence and suggested that there was no problem. It was claimed that the occupancy level was quite acceptable for a ward unit of that nature. This example illustrates three lines of defence frequently taken by those clinicians who are unwilling to participate in the management of the resources they use.

Provision of data to the clinician who is unwilling to participate in the management process is of little practical use. It may have a general educative value which may be of benefit when the clinician is ultimately persuaded to participate in the management process, but it may also be used as a reason for avoiding such involvement. Requests for additional and more sophisticated data often only serve to delay the point of decision for as long as possible or even to put off a decision altogether.

If a clinician is not involved in the management process the presentation of data to him concerning a problem that requires solution offers no guarantee that participation will result. It is commonly supposed that information will help bring about the participation of clinicians in management. It has been demonstrated that this is only the case where the clinician is willing and able to participate but is otherwise unable to identify or solve the problem. It is also important to realize that very often the reason given for non-participation hides the real reason. Only too often an apparent inability to see the problem or to find a solution is put to the test by the provision of data and reveals that the real disability is an unwillingness on the part of the clinician to undertake a management role at all.

In fairness it must be recognized that there are occasions when the apparent inability or unwillingness are probably attributable to the lack of information. It may be that the presentation of data illustrated in the example above was not undertaken in a satisfactory manner, or that the data itself did not meet an acceptable standard of relevance, accuracy, completeness, and timeliness. If that was the case then a further danger in data presentation becomes apparent: poor presentation and poor data can force a clinician into a position where he feels unable or unwilling to participate in the management process.

Those who expect databanks and information systems to change the attitude of clinicians should therefore beware, since it is clear from the latter two examples that data is not the key to gaining

participation. This being so, emphasis needs to be placed on discovering what factors motivate a clinician to participate in the management of clinical work. It is likely that research in this area would demonstrate that the creation of the right environment and climate of opinion and the training of the clinician in the needs and techniques of management are as important, if not more important, than the provision of management information.

THE USE OF INFORMATION BY CLINICIANS IN THE MANAGEMENT OF THEIR WORK

Having examined the role of information in relation to the clinician not participating in management it is now proposed to go to the other extreme of the participative spectrum and examine the effect of information on the management of clinical work where the clinician is prepared to participate fully in the management process. It is in this area that the provision of information is expected to improve the decision-making process. Any causal relationship between the provision of information and the obtaining of a satisfactory or improved result is difficult to demonstrate. Evidence can be shown of occasions where the use of information has preceded improved results. The evidence below falls short of scientific proof, for although the improvements shown are more than likely to have been affected by the input of information it is possible that the improvement would have taken place at that point in time regardless of the input of information. Three major difficulties are encountered in trying to provide evidence of a causal relationship:

1. Decisions are never made totally without information and it is difficult to measure exactly what information has been used and what importance was placed on the information used.
2. Identical situations which might be suitable for experimenting with the presentation of different levels of data seldom arise.
3. It is difficult to anticipate situations sufficiently similar to attempt a randomized controlled trial.

Before examining further the use of information in the management of clinical work it might be useful to provide a simple framework upon which to base the illustrations. It is suggested that the management process includes the following stages: anticipating and identifying problems, defining and clarifying problems, analysing problems and searching for solutions, choosing a solution, action, and evaluation.

This very simplified process should be in the form of a loop for the

evaluation stage will become part of the act of searching for and identifying further problems. All but the 'action' stage of the process require an attempt to measure or quantify some aspect of the problem. In the examples that follow the use of the word clinician indicates a clinician who is participating in the management process.

Anticipating and identifying problems. Clinicians and administrators are anxious not only to deal with problems that occur within their sphere of responsibility but also to identify such problems at an early stage in order to provide a warning of a breakdown in services. Work undertaken in one division has indicated that information can help in this respect providing the clinicians are prepared to suggest some expected standards of performance. In one example of this work a dental division agreed upon certain standards which they felt would indicate the numbers of out-patients' attendances required in order to provide enough teaching material for dental students. An estimate was prepared a year in advance of the number of patient attendances required in each department of the hospital on a month-to-month basis. Throughout the period comparisons were made each month of the actual attendance pattern with the standard previously set. Comparisons were also made of the cumulative patterns thereby assisting in determining trends. Both the monthly and the cumulative monthly figures had been predicted in terms of a band in that the upper and lower limits were designed to indicate limits of acceptability. The upper limits were intended to indicate that more patients were attending the hospital than the resources could cope with and the lower level limit that not enough patients were attending to satisfy the teaching requirements of the students.

The experiment showed that early warnings of unacceptable levels regarding patient attendances were given and that sufficient time was available for the dentists to attempt to adjust the workload. In the few cases where unacceptable levels occurred, approximately six months' prior notice of the likely trend was given. This attempt at monitoring was combined with the introduction of an exception reporting system. Not only did the division agree to standards regarding the number of patients' attendances but they also agreed that unless the figures were outside the ranges predicted there was no need to report the actual figure or fluctuations within that range to the division.

This exception reporting system vastly reduced the amount of data

presented to the division and at the same time highlighted areas which required attention. The monitoring process developed in this exercise included the following steps:

1. The identification of measurable characteristics in the current activity.
2. Description of the proposed plans in those terms.
3. Agreement on a standard of performance.
4. Comparison of the observed performance with the standard set.
5. The reporting of discrepancies from that standard.

This monitoring procedure was preceded by the setting of objectives and the preparation of a plan and was followed by the taking of action either to change the performance when the objectives were not being met or alternatively to adjust the plan or objectives.

The exercise indicated that data can become useful and usable information, provided that the clinician is prepared to participate in the setting of standards. The main problems of the exercise lay not with the handling of the data but with the difficulties encountered in the setting of the standards. The exercise indicated that the use of information in setting standards, monitoring performance, and providing exception reports in order to achieve early identification of problems might well prove a fruitful field for research.

Defining and clarifying problems and the search for and choice of solution. The next illustration is used to examine three stages of the management process.

Three surgical wards were each responsible for taking a share of the emergency admissions for a district general hospital. The admission and discharge policies of the three consultants who were each responsible for a ward were such that one ward took a disproportionate number of the emergency admissions. This led to differing workloads on similarly staffed wards. The nursing staff and junior medical staff on the ward taking the highest proportion of admissions were acutely aware of the problems concerned with a high turnover of patients. None of the staff were clear as to the reasons for the apparent high workload. The three consultants and the staff of the other wards at first were not aware of the problem but when it was brought to their attention found it difficult to understand the basis of the problem as each of the wards shared equally seven take days in any twenty-one-day period. The analysis of the data available from Hospital Activity Analysis indicated two problems. Firstly, the way

in which the take day pattern was arranged provided for all three wards a very uneven emergency admission occupancy pattern, which made it difficult to arrange the planned admission cases in an efficient manner. The second problem was that only one of the wards made a deliberate attempt to clear beds for its emergency take days and although it managed to cope with the emergencies on its own take days it had also to take the overflow of other wards' admission days. The information had thus helped to identify more precisely the problems concerned. It was then possible to use the same information to search for a solution to the problem. A simulation of different emergency admission patterns was produced after consultation with the nursing and medical staff and one or two patterns revealed a more equitable distribution of emergency admissions and thus of emergency occupation. Not only were the data able to suggest a number of solutions but also able to help choose the best solution under given circumstances, provided that the clinicians were prepared to set the standards required in the solution of the problem and the order of the importance of the variables. The type of decisions that the clinicians had to make were to determine what were the acceptable levels in terms of the number of days extra beds could be erected upon a ward, the number of times interward transfers could be allowed, and the relative merit of having a more stable and easily understood emergency admission pattern which would, for example, guarantee staff regular off-duty periods.

Management problems can undoubtedly be defined and clarified by the use of information. Having identified the problems concerned with the running of the three surgical wards the consultants could have immediately attempted to solve the problems by changes in the emergency admission pattern which they believed might improve the situation. Without the use of data, however, the wards might well have become overworked during certain periods because of a bad choice of emergency admission patterns, and a trial and error procedure would have been extremely tiresome to the ward and theatre staff who would continually have had to change their patterns of work until a satisfactory method had been established. The use of data in this instance saved a large number of changes and also reduced the risk of an error of judgement. The example also illustrated the method of using information to help arrive at a solution.

It must, however, be remembered that the usefulness of information is always limited by the availability and the accuracy of the data.

It is also limited by the ability of the manager to determine the relative values of various sets of data. The precise order of importance of the variables cannot always be included in a simulation. Ultimately the choice must be made by the clinician and because a perfect decision cannot be supplied by the information alone judgement is required at that point.

Action. Information has no influence on the actual execution of the action which was suggested by the information. Information may be required to assist the plan of action but finally it is the clinician who must carry out the action suggested. This point was clearly illustrated earlier in the text when a group of clinicians were faced with the extremely low occupancy of three wards. The same data were presented to the same clinicians at the same time, but action resulted in only two of the three cases.

Evaluation. Evaluation involves measuring a result and comparing it with a standard. It was demonstrated (see p. 149) that if the clinician can set up a measureable standard then it is possible to analyse the data in such a way as to assist in the evaluation. A medical executive committee made a decision to alter the admission policy of a hospital by accepting additional emergency cases. This was undertaken on the understanding that no other change in the clinical practice of the hospital. These included such items as the number and pattern admission characteristics concerning the existing medical practice of the hospital. These included such items as the number and pattern admission of planned cases, the age distribution of planned and emergency admissions, the length of stay distribution for both planned and emergency cases and the number of inter-ward transfers. Whilst it was appreciated that the type of measures suggested above did not give a complete indication of the hospital's clinical practice, it was considered that the measures used would detect any major changes. It was thus a relatively simple process to evaluate the effects of the change of hospital admission policy.

Conclusions

Although the evidence falls short of scientific proof, the case-studies illustrated above strongly suggest that the use of information can help identify problems previously not seen, give an early indication of

problems and ease and speed up the analysis and solution of problems. The setting up of an information system for clinical management in hospitals should aid those concerned in priority assessment and resource control. Much work remains to be done, but experience suggests that there are considerable rewards to be reaped from further work in the following areas.

Information requirements. Further work is required to help distinguish between real and perceived information requirements.

1. *Information required by management.* Too often the clinician makes requests for data rather than information. Information services should test requests by inquiring what use will be made of the data. If the clinician has difficulty in answering such an inquiry it may indicate that the real information requirements have not been specified. Even when information is used the extent of its relevance to the decision may be questioned. Often the relevance is a matter of subjective judgement and two clinicians or managers may disagree on the weight of the importance that can be attached to a certain piece of information. In this area further research is required to determine what information is needed for particular sets of decisions.

2. *Requirements of information itself.* It should be noted that in these case-studies and examples the data and information provided to the clinicians was not always complete, relevant, accurate, and timely. The standard achieved was, however, high by comparison with the hospital service in general. The completeness of the data was nearly always 100 per cent. A great deal of time and effort was devoted to checking the accuracy of the data and explaining any inaccuracies to the users of the data. Every effort was made to ensure that the data was as relevant as possible by frequent discussions with the clinicians concerned on the information required. Finally the time-scale was in many cases as short as one or three days after the event. It was apparent in many cases that decisions could have been made with data that was less accurate and complete and was not so speedily presented. Whilst the study had questioned whether there is always a need to obtain a 100 per cent accuracy and completeness and achieve a real-time response it does seem that these features significantly increase the confidence of clinicians in the data and the clinicians willingness to use that data. If for no other reason it would seem desirable that an information service should aim at 100 per cent standards.

Experiments in the setting of standards. Experiments involving those responsible for management in the setting of standards of performance may assist in distinguishing information from data and ascertaining information requirements. Monitoring of care will prove an unproductive exercise unless clinicians can become involved in the setting of meaningful and measurable standards of performance. If the difficulties of standard setting can be overcome the use of exception reporting techniques will greatly assist the management process.

Search for usable measures of outcome. Progress in the field of developing information for clinical management in hospitals has been impeded by the lack of suitable measures of outcome of hospital and health care. In the past the fact that the health services have not been integrated has been seen as a hindrance to the development of such measures. The proposed integration of the health services may now test whether that reason given for lack of development is an important factor.

Clarification of the position of information in management. There is a real danger that too much emphasis may be placed on the uncritical collection of data and undue confidence placed in the ability of information itself to solve the managerial problems. The distinction between data that is presented and information that can be used should be clearly understood by all managers before attempts are made to define information requirements, and the status of information in the management process must be recognized by management. It is not a supplement to management but a part of management. The management process will be weaker for its exclusion but its inclusion will not solve all managerial problems. As much if not more emphasis must be placed on the training and education of clinicians in the need for and techniques of management and in creating a favourable environment for its practice.

This paper has illustrated some of the work undertaken in the field of information for clinical management at hospital level. Parallels between clinical management in hospitals and health management in general are not hard to find. The problems encountered in providing information for the management of those two fields also appear similar. The four areas suggested for further study at hospital level might also be studied in the wider field of health service management.

The distinction between information and data must still be made and it must be recognized that information will not prove a panacea for all managerial ills.

ACKNOWLEDGEMENTS

Much of the framework of this paper was based on work undertaken by the author during the period of a research grant to analyse the information requirements of a divisional structure. I am grateful to the Nuffield Provincial Hospitals Trust for the provision of the grant and to the United Birmingham Hospitals for providing the facilities for the research in the past eighteen months. My particular thanks go to Oriole Goldsmith, Dr Gordon Cumming, Jon Evans, and Malcolm Buckler, for their advice during this time. Responsibility for the views expressed and the material presented in this paper remains mine alone.

REFERENCES

1. MINISTRY OF HEALTH (1967). *First Report of the Joint Working Party on the Organisation of Medical Work in Hospitals* (London: HMSO).
2. MCLACHLAN, G. (ed.) (1971). *Challenges for Change. Essays on the next decade in the National Health Service* (Oxford University Press for the Nuffield Provincial Hospitals Trust).
3. ——(1971). *In Low Gear? An examination of 'Cogwheels'*. Occasional Hundreds 2 (Oxford University Press for the Nuffield Provincial Hospitals Trust).
4. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1972). *Management Arrangements for the Reorganised National Health Service* (London: HMSO).
5. BODENHAM, K. E., and WELLMAN, F. (1972). *Foundations for Health Service Management. A Scicon report for the Scottish Home and Health Department on the requirements for a health service information system.* Occasional Hundreds 3 (Oxford University Press for the Nuffield Provincial Hospitals Trust).
6. DEPARTMENT OF HEALTH AND SOCIAL SECURITY (1972). *Second Report of the Joint Working Party on the Organisation of Medical Work in Hospitals* (London: HMSO).
7. STEWART, R. (1963). *The Reality of Management* (London: Heinemann).
8. MCRAE, T. W. (1971). *Management of Information Systems* (Harmondsworth: Penguin).

6

A basis for management decisions on the deployment of nurses

Evaluation of patients' nursing needs: prediction of staffing

Catherine Rhys Hearn
BSc, MSc, PhD
Senior Research Associate,
Health Services Research Centre,
University of Birmingham

Evaluation of patients' nursing needs: prediction of staffing

Summary. *A methodology is identified for relating patients' nursing needs to the nursing workload generated, and hence prediction of staffing. Various successful applications of the method in different clinical environments are presented; future extensions are outlined.*

INTRODUCTION

The aim of this paper is to describe a methodology for exploring the relationship between patients' nursing needs and the resources necessary to meet those needs.

Consider the meaning of the phrase 'nursing needs'. Who decides what nursing care a patient 'needs'? It is well known that the same patient, nursed in different care-facilities, or even in different wards in the same hospital, might seem superficially to have entirely different 'needs'. The reason for this, of course, is that a patient's nursing needs are not defined by any absolute standards, but are interpreted by the medical and nursing staff in charge of the patient's care, and subsequently differ according to the clinical policies and nursing standards of the particular care-facility within which the patient is situated.

The nursing care actually received by the patient, however, is determined not only by his needs, but also by availability of resources to meet those needs, and of course upon the efficient management of the care-facility. Thus if insufficient resources are available, the care received by an individual patient will also depend upon a priority

system worked by the nursing staff which is governed by their interpretation of the simultaneous needs of other patients. In order to decide whether or not existing resources are adequate, it is necessary to devise a means of relating 'needs' to 'resources'. The present studies examine one resource, ie nursing staff, and define 'needs' in terms of patient's requirement for *nursing-staff time* from nurses of different skill levels interpreted according to the policies and standards of each nursing area considered. This means that the proposed scheme can be used in *any* nursing environment, as the common parameter is nursing time which can in each instance be related to the specific needs of the area. It also allows comparison between different facilities, wards, or units in terms of staff needed to provide the nursing care for the patients, and could lead to an evaluation of existing policies current on similar units. The problem, of course, is in deciding what constitute 'similar' units; but the present approach does in fact describe a set of common areas of nursing responsibility and activity within which the differences between units may be identified. A considerable amount of work has already been done on the evaluation of patients' 'nursing dependency' (ie the dependence of patients upon nursing staff) and this is reviewed elsewhere (8). A list of references is appended on pp. 192–3 of papers which describe specific studies made and approaches followed (with the exception of references 6 and 7). Many of these evaluations have proved useful in limited applications, and there is considerable current interest in this field of research. There seems to be a very real need for a generalized philosophy of approach which provides a unified picture of the state of the research at the present time. The present methodology does in fact provide such a framework in that it formalizes the general ideas behind most of the previous studies; the detailed development of the actual methods of measurement, approximations made and applications attempted are the differing factors in each case. In this article both the general formalism, and the detailed method of calculating nursing workload (and hence staffing numbers) developed by the author are described.

General comments

One of the main reasons for the current upsurge of interest in nursing dependency studies is the general feeling that hospital nursing is a scarce and expensive health care resource, and that information

gained from such studies could lead to more efficient use of the nursing resources available. With the planned reorganization of the health services, and changes in policies for providing patient care, traditional patterns of hospital ward staffing are coming under close scrutiny, and reliable means of determining realistic staffing establishments are urgently sought. In order for any method of measurement to prove useful it must be generally applicable to any nursing environment; only in this way can the comparative needs of different types of nursing facility be assessed, and hence viable staffing allocations and policies regarding transfer of patients be evolved.

Any method of measuring nursing workload should meet with the following requirements if it is to be useful for management purposes:

1. It must incorporate *all aspects* of nursing care.
2. It must produce results which correspond realistically with the real nursing workload on any unit.
3. It must be applicable to *all nursing-care* areas (for example, geriatric, ophthalmic, ENT, intensive care unit, specialist areas, psychiatric and community nursing as well as general medical and surgical wards in hospitals).
4. It should measure workload in such a way that the *skills* of the staff required are also taken into account.
5. It should be developed into a simple and routinely usable system, with the cost and effort required for its implementation reduced to a minimum.

With such a system, the information obtained could be used for a variety of purposes, including both long- and short-term planning.

In addition to estimating hospital establishments, long-term objectives include planning staffing for new types of specialist nursing units, new types of patient-care facility with nursing responsibility, and nursing resources in the community. By studying care-plans and patterns of nursing care associated with different groups of patients, it should be possible to predict workload and plan staffing patterns, and transfer of patients between different care-facilities in such a way as to make most effective use of available resources.

Short-term applications of an effective measure of patients' nursing workload would provide information to aid day-to-day ward management in hospitals; detailed predictions of the relative workloads on different wards could help in making decisions concerning redeployment of staff where possible; knowledge of a patient's nursing workload category would assist in deciding where he should be placed

in a system of progressive patient-care based on nursing needs. Also, the optimum size and composition of a hospital nursing pool could be determined, if this was a part of the hospital's over-all nursing policy.

So far, none of the methods developed for measuring nursing workload in association with a consideration of patient's nursing needs conforms with all the requirements listed above. Very often the measures which are quickest and easiest to use are found by nurses to be applicable only in very limited circumstances, and hence are rejected on the basis that the patients are not being properly assessed. Also, many nursing workload measures fail to take account of some extremely important aspects of the nursing function (8). These rely almost entirely upon the assessment of patients' basic-care requirements, with some additional consideration of treatments, procedures, and investigations. Very rarely has sufficient attention been paid to the less tangible aspects of patient-care, such as the reassurance of emotionally dependent patients, or the rehabilitation of geriatric or disabled patients.

The particular method to be described in this article does attempt to take this aspect of nursing into account, and also of special characteristics of individual patients which tend to make the provision of their care more time-consuming. These patient-characteristics, called 'dependency factors', include such things as 'obese, heavy, immobile', or 'physically disabled', or 'incontinent', 'unable to communicate', or even perhaps 'unco-operative'.

The differences between teaching and non-teaching hospitals is absorbed in one of the subdivisions of nursing care (indirect care) which is described in more detail on p. 164. Other differences between hospitals, such as the availability of supporting services, plated meals, ward clerks, and the geographic layout may all be specified within the scheme outlined here. The method has so far been applied in medical, surgical, and specialist units in three teaching hospitals (one American, two British) (1, 2, 4, 5); in a psycho-geriatric unit in a local non-teaching hospital; to neurological patients, both medical and surgical in a specialist hospital, and to ante- and post-natal patients at a maternity hospital as well as in the main delivery suite and intensive care unit. The method was found to be acceptable in all these different environments, and attempts to evaluate the information gained were made. Descriptions of these studies are presented on p. 164, and more details of the method, the general formalism and concepts involved are given below.

Discussion of concepts and details of method of assessing patients' needs

The concept of patients' nursing 'needs' has already been discussed at the beginning of this paper and it has been established that the present approach involves specifying these 'needs' according to the medical and nursing policies of the staff and institution providing the care. The 'needs' can be related to workload generated; this can be measured in terms of nurse-time needed from staff of different skills. The term 'nursing dependency' or 'patient/nurse dependency' which has been in common use for some years now is often taken to mean 'the amount of bedside nursing needed by a patient'. It is used in the context of the present work in exactly that way, where 'amount' means workload (measured as described above) and 'needed' is interpreted in the way already mentioned.

Confusion often arises out of the use of the term 'nursing dependency' as it is taken to mean a variety of different things according to circumstances. For example, 'needed' is often assumed to mean the same as 'received', ie nursing dependency is measured in terms of items of care received by a patient without establishing whether the care received conforms with policy regarding what should be provided. Again confusion often arises out of what is meant by 'amount' of nursing. In many cases, of course, work-study is carried out and nursing dependency categories are related to measured workload. In other circumstances, however, nursing dependency categories are defined according to somewhat arbitrary (intuitive) criteria and a relationship between these categories and workload is assumed without adequate backing. In common parlance nursing dependency is often assumed to be synonymous with 'degree of physical disability', and although there is likely to be a strong relationship between them, it is not always as simple as might be supposed. For example, a highly emotionally dependent patient may well generate a greater nursing workload than a well-adjusted physically disabled patient.

A patient's 'nursing dependency' is also used rather loosely at times to distinguish between *facilities* needed for providing appropriate care.

For example, patients placed in 'high care' units because of their need for special support facilities such as respirators or intravenous therapy are said to have 'high nursing dependency'. This is, of course, in all probability the case, but again a relationship should be established between the nursing dependency categories and the workload

generated. In this particular case, it is probably the different *skills* required from the staff which is the dominant feature.

The third concept which requires some discussion at this stage is that of 'nursing care'. What do we mean by 'nursing care', and 'nursing workload'? There may well be considerable differences of opinion here. In fact a great deal of discussion and study has taken place regarding nursing and non-nursing duties on hospital wards, but it is still true that nurses' responsibilities and duties differ according to environment. Clearly, the amount of work done by the nursing staff is bound to be affected by the availability of support staff, such as housekeeping teams, ward clerks, ward porters, and so on. The hospital's layout and facilities will affect workload; if the hospital has a plated meal service the nursing staff need not spend much time on 'non-nursing' duties connected with serving meals. The fact that a hospital is a teaching hospital will affect workload because of time spent by nurses training pupils, attending teaching rounds, and because of increased time spent in communication with medical staff and students.

'Nursing workload' in the context of this work is taken to include *all* work done by nurses, whether or not some of this work includes items normally considered to be non-nursing duties. The philosophy is that if nurses are doing the work, and this is consistent with the hospital's policy, then it is reasonable to predict nurse staffing numbers needed on the basis of the work they do. If it is suggested that policy be changed and new types of staff introduced to carry out non-nursing duties, then the information is still available as to the relative numbers of nurses and new staff needed to match the workload.

In order to classify the nurses' work into categories that are comparable in different nursing areas, the present approach subdivides nursing work into three categories: 'direct care', 'indirect care', and 'routine care', and these are defined as follows:

Direct care is nursing care 'needed' by individual patients (as interpreted by the medical and nursing staff in charge of their care). This is often called 'bedside care' and differs in workload from patient to patient according to individual need. Hence direct care is *patient-oriented* and includes only those items of nursing care which can be directly attributed to the specific needs of individual patients. A patient's 'nursing dependency' is related to his need for this type of care.

Indirect care includes all the other tasks and duties undertaken by the

nursing staff. In particular this category of nursing work is involved with the care of a group of patients, such as a ward, or a unit, and includes items sometimes known as 'collective nursing'. Hence, we might say that indirect care is *group-oriented*, though clearly this is also the section which is particularly affected by the location, situation, and facilities of the nursing area in which a group of patients is being nursed.

This area of nursing responsibility also includes any teaching commitment of the staff on the wards. Hence it incorporates time spent in supervising and teaching student nurses, and time spent in discussion periods on the ward and in conferences. In addition, indirect care also includes time spent with medical staff, both on rounds, in assisting with procedures, and in conference. In the case of teaching hospitals, there is considerably more time spent by nurses on this aspect of indirect care, including more routine ward rounds, teaching rounds, and in discussions involving medical students. It has been shown (16) that nursing workload increases with numbers of medical staff available, and that hospitals with high numbers of medical staff also need more nurses. Hence it is essential that differences in nursing workload caused by different organization in hospitals should be clearly identified in order to allow realistic comparisons to be made when allocating nursing staff.

Routine care is a category introduced merely for convenience, and includes those items of direct or indirect care which are provided routinely for all patients within a group irrespective of their needs or their nursing dependency. Hence this category of nursing care is policy-oriented and differs from one group (for example, ward or unit) to another according to existing nursing policy. Fig. 1 shows diagrammatically the subdivision of 'nursing care' into the three categories, and presents examples of items of care appearing in each section. (These examples were taken from general medical and surgical wards at a teaching hospital.)

Details of method of assessing workload/staffing. For the purpose of applying the methods of analysis proposed in this work, detailed information for each of the categories of care mentioned above is required for any nursing area to be studied.

Firstly, all the items of care occurring within each of the categories must be identified and listed. Ultimately, it is required to know the work content of the items, and the skill-levels of the staff required to

	NURSING CARE		
	DIRECT CARE		INDIRECT CARE
DEFINITION	*Patient-oriented* care: bedside care provided to individual patients according to their needs; differs from patient to patient.	ROUTINE CARE *Policy-oriented* care: care provided to *all* patients routinely according to policy.	*'Area'-oriented* care (ward, unit, etc.): tasks performed by nursing staff *not* directly attributable to an individual patient.
FACTORS DETERMINING CARE PRESCRIBED	1. Patient's condition. 2. Medical policies. 3. Accepted nursing standards.	Accepted nursing standards of the particular nursing area concerned.	1. Function and layout of nursing area. 2. Administrative policies. 3. Communications. 4. Teaching commitment. 5. Accepted standards of safety and hygiene. 6. Availability of supporting services.
CONSTITUENT PARTS	1. *Support of normal-living functions.* Determined by: Patient's degree of self-care. Accepted standards of basic nursing. 2. *Treatments, procedures, observations.* Determined by: Management of patient's medical problems. Accepted medical/nursing standards. 3. *Professional surveillance.* Determined by: Patient's physical, emotional, and rehabilitative needs as interpreted by the medical and nursing staff in charge of his care.	*Combination* of items of 'direct' and 'indirect' care, determined according to the policies and practices of the particular nursing area considered.	1. Movement and identification of patients. 2. Recording, administration. 3. Communication with and assisting medical staff; reporting. 4. Teaching students and pupils. 5. Maintenance of standards of patient-care, safety, and hygiene. 6. Non-nursing duties caused by inadequate supporting services.
EXAMPLES OF ITEMS OF CARE IN EACH CATEGORY	1. Bed-bathing, turning, feeding, skin care. 2. Dressings, injections, drainage, cardiac monitoring. 3. Emotional support, rehabilitative procedures, explanation of nature of disease, specialling, checking whether patient is comfortable.	Bedmaking, distribution of drinks, pre-meal care, distribution of wash-basins.	1. Admission of patient, discharge, transfer. 2. Records, notes; off-duty rotas. 3. Ward rounds, ward reports, assisting medical staff. 4. Bedside teaching, discussion sessions. 5. Checking drug-cupboards, fire-hazards; correct isolation procedures. 6. Serving meals, accompanying patients to other areas.

Fig. 1. Analysis of nursing care.

provide the items. The former requirement may be met by appropriate work measurement exercises; the latter poses more of a problem.

The main difficulty here is, of course, that the same nursing procedure carried out for different patients may require nurses of different levels of skill because of differences in the patients' personal characteristics. There seems to be no obvious way of classifying patients in advance so as to predict the required degree of skill or experience needed. Another problem arising here is again that of differing policies in different hospitals; for example, one hospital may agree that a nursing auxiliary may measure and record blood-pressures and another hospital may consider this unacceptable. In either case, the proviso 'under adequate supervision, and only on certain patients' would be included. Again, in teaching hospitals, the skill-levels and degree of experience of the students at different stages of their training are considered differently in different hospitals. Hence no over-all rules regarding skills can be made, but the present methodology does allow each individual nursing area (hospital, clinic, or community district) to take account of skills needed according to its own policies and requirements. Three skill levels are identified and are called *basic*, *skilled*, and *technical*, in ascending order of skill level or experience. The decision as to which nurses go into which category is not predefined but is left to the judgement of the nursing administration of the nursing area concerned.

For example, in one teaching hospital, the following grouping was agreed:

Basic = Nursing auxiliaries
Skilled = Nurses in training
Technical = SRN.

No other grades were present. In a maternity hospital which was also part of a teaching group the classification was:

Basic = Nursing auxiliaries, theatre attendant, SEN, nursery nurses
Skilled = Pupil midwives
Technical = SCM.

Whereas in a local non-teaching hospital's psycho-geriatric unit the grouping was

Basic = Nursing auxiliaries
Skilled = SEN
Technical = SRN.

Even with these general groupings, of course, the ward managers still have to match individuals with the tasks, and the general classification of 'skills' in this way is not perfect. However, it does provide a basis for comparison between policies of different nursing areas and it does provide a formalism for expressing 'skills' and including this factor into nursing dependency studies.

Workload derived from *indirect* and *routine* nursing needs may generally be assessed for the unit (ward, or care-area) as a whole. The variability in nursing workload is caused mainly by the variable *direct* care needs of the individual patients. Hence, the next stage in the programme is to design a method of capturing the necessary information regarding individual patients' nursing needs. For convenience, *direct* care is subdivided into three sections (see Fig. 1) as follows (these are described in detail elsewhere [2]):

Support of normal living functions. This includes 'nursing' care which the patient needs because of his inability to perform for himself the normal functions of daily living. Hence this is basic care which the patient needs whether or not he is in hospital; in many instances this care is provided by relatives in the community.

It includes basic nursing items such as bedbathing, feeding, oral hygiene, turning, ambulating, and skin care (2). The provision of this care is determined by *nursing policy* and differs quite considerably from one hospital to another. It is necessary to specify standards of basic care deemed acceptable in the particular care area being studied, and to relate these to the workload generated for each category of patient. Considerable work has been done on this already (13) but work-study results would prove very useful here in the particular ward being studied.

Treatments, procedures, and observations. This category of nursing care is fairly well defined and is directly related to the patients' medical-care requirements.

These items include dressings, injections, suction, oxygen administration, and so on (2). Hence the patients' needs for these items are directly related to the *medical policies* of the medical staff in charge of the patients' care, as well as the nursing policies of the ward nursing staff. Again, these differ considerably from one consultant to another, even within the same hospital, and it is necessary to record the patients' needs in some detail in order to assess the workload

generated. (Work measurement exercises provide background data for workload content of individual items in this category.)

Professional surveillance, patient education, rehabilitation. This area of nursing care is frequently neglected both in workload calculations and in practice. Examples of items occurring in this category are: providing support to emotionally dependent patients, teaching patients and relatives to cope with the patients' after-care, checking that pills are taken, oxygen is properly used, monitors and drips are functioning properly, and specialling very ill patients. (A detailed description of this aspect of nursing care is given elsewhere [2].) The provision of care of this type is related both to medical and nursing policy, and clearly depends upon the emotional and psychological as well as physical needs of the patients. In order to obtain the information necessary to assess the workload generated by care given in this area of nursing commitment, the ward nursing staff must be asked to record their assessment of the surveillance needed by the individual patients. At the present time, this is only done occasionally, for some patients, with comments such as 'observe closely', 'check dressings frequently', or 'patient is emotionally disturbed'.

Patients' 'dependency factors', as mentioned previously, are also recorded where relevant; this allows the workload assessment to take account of personal characteristics of patients which extend the nursing staff time necessary to provide them with the care they need (examples of dependency factors observed in general medical and surgical wards are given elsewhere [2]). In addition, movement of patients in and out of wards is also recorded, as it is observed that fluctuations in the numbers of admissions, discharges, transfers, etc., occurring per day can cause the workload to vary quite considerably. Clearly, in order to obtain the required information, data collection forms of some kind have to be designed. A *generalized format* has been developed, and consists of four sections, where the appropriate details for the particular ward or area being studied can be recorded. The four sections are related to the three areas of nursing commitment already described, and the fourth section is for recording dependency factors. The total ward workload consists of: indirect care workload for the ward (basic, skilled, and technical); routine care workload for the ward (basic, skilled, and technical); direct care workload summed for each individual patient on the ward (basic, skilled, and technical); and the results over time may easily be con-

verted into numbers of nurses of different skills required to meet this workload.[1]

It should be stressed once more, in conclusion, that this method of assessment is generally applicable in any nursing environment, and measures workload (and hence staff needs) according to the standards and policies of the particular area being assessed. It also provides a formalism for exhibiting the distribution of this workload between the various areas of nursing responsibility defined; this provides information for comparative studies with other nursing areas and hence allows decisions to be made regarding optimum staffing patterns, and provides a means of evaluating the effects of proposed changes in nursing policies.

Description of applications

As explained previously, this paper describes a generalized formalism for expressing patients' nursing needs in terms of the amount of nurse-time required by the patients from staff of different skill levels. The meaning of the word 'needs' in this context has already been indicated and the concept discussed at some length on p. 163. Since the method employed is generally applicable, and can be adapted to all kinds of nursing situations, there is no reason to restrict its use to hospital nursing only; however, all of the applications to be described in this section have in fact been in hospital wards.

The present scheme was first developed in 1970 at the Beth Israel Hospital, Boston, Mass., USA, to help with a problem that existed in the hospital at that time. The Director of Nursing felt that there were not sufficient nurses to meet in full the nursing needs of the patients. She felt that since the time that the staffing levels had been agreed, the pattern of nursing care within the hospital had changed, and that the changed workload caused by changes in medical practice and admission policy was no longer being met by the available nursing staff. However, without some kind of quantitative workload-measure related to the nursing needs of different types of patient, it is very difficult to justify demands for greater resources, and to express these demands in quantitative terms. Merely quoting the very high bed-occupancy occurring in the hospital at the time was unhelpful, since in fact, the bed-occupancy had not changed. It was the nursing 'needs' of the patients in the beds which had changed, since the average

1. List of stages required in setting up studies, data-collection forms with instructions, and details of staffing calculations may be obtained from the author.

length of stay had been reduced, the apparent number of emergency admissions had increased, and the medical staff were introducing new policies and methods of patient-management which involved increased nursing commitment.

In an attempt to obtain more quantitative information, the hospital had arranged to take part in a nursing work-study exercise, which involved many of the hospitals in Massachusetts, and was carried out by the Massachusetts Hospitals Association.

It was subsequently suggested that the results of the work-study combined with a relevant (and realistic from the nursing point of view) classification of patients might lead to the development of a useful tool, to provide information both for long-term budgeting purposes and to assist with daily deployment of staff currently available.

The present method was developed from these beginnings. As already indicated, the essence of this method is that it requires an analysis of the activities undertaken by the nurses plus an assessment of the workload generated by these items. This then involves the acquisition of information regarding *individual* patients' needs within each of the areas of nursing activity identified. Hence it removes the necessity for 'classifying' patients in the traditional way. This means, of course, that the initial studies require the collection of detailed information for each patient on a daily basis. Initial studies in the Beth Israel Hospital, Boston, were made over a period of about two months, on various acute medical and surgical units. These were retrospective studies (ie data were recorded for the past twenty-four-hour period) and the results of the studies were used to determine optimum establishments for the wards. From the daily fluctuations in calculated workload it was possible to estimate the optimum size and composition (in terms of trained staff and nursing aides) of a hospital nursing 'float'.

An attempt at evaluating the results was made. The nursing administration approved the staffing levels calculated, as the numbers agreed very well with their own, intuitive, establishments. In addition, an attempt was made to determine whether care was affected when staffing numbers were low compared with the calculated values. There was already in current use at the hospital a 'quality control' measure designed by the Massachusetts Hospitals Association. This provided an indication of whether or not hospital nursing policy was being adhered to in a fairly general way. It was certainly shown, even though with this somewhat crude indicator, that 'quality' defined in

this way was reduced when the calculated workload exceeded staff availability. It was also shown that adequate 'quality' (by this measure only) was achieved when the staffing agreed with the calculated staff numbers and grades. This 'quality control' measure was in fact a checklist of items which were deemed to be necessary in providing nursing care to a ward of patients. It was really more in the nature of a monitoring sheet; many of the items recorded whether routine hygiene and fire hazards were being avoided and whether ward tidiness and general patient comfort were being supervised adequately. However, although this evaluation of the nursing dependency method was somewhat superficial, it provided sufficient information and insight into the nursing workload generated by different patients' needs to convince the Director of Nursing and the administration that it would be useful and worth while to develop the method for use as an information tool to assist with deployment of staff on a daily basis.

Clearly, in this case the objectives were somewhat different. The important factor now was to predict the extent of fluctuation in workload likely to occur during the next time period, in this instance over the next twenty-four hours, ie for the next three nursing shifts. It was decided considerably to reduce the nursing effort required in recording the data. For a research study it was reasonable to ask the charge nurses to record detailed information about each patient, but for a daily routine operation, development of a much simplified version became a very high priority. Also, it should be noted that since the recorded information was to be used to *predict* staffing over the next three shifts, the data forms had now to be filled in *predictively*. This meant that the nurses were being asked to consider the following facts:

1. The patient's likely 'degree of self care' over the next twenty-four hours.
2. The amount of professional surveillance he was likely to require during the next twenty-four hours.
3. The treatments, procedures, observations, etc., ordered for the patient.
4. Any dependency factors relevant to the patient.

Clearly, the nurses' recordings for 1 and 2 usually referred to their experience over the past twenty-four hours, unless they knew, for example, that the patient was due to undergo surgery next morning; items under 3 were well-defined as they were generated by written orders signed by the medical staff; dependency factors, 4, as already

mentioned (p. 169), were personal characteristics of the patient deemed by the nurses to make the patient's care more time consuming.

An acceptable and simplified routine daily system for capturing the required data and producing the staffing predictions was implemented, and is described in detail elsewhere (1).

The assessments of patients' needs recorded on each of the eight medical and surgical units were entered into a computer via an on-line typewriter terminal before the start of the night shift. A conversational programme had been written for the computer which guided a clerk to enter the data in a very simple-minded manner. Built into the programmes, of course, were allowances for indirect and routine nursing, and allowances for probable admissions and discharges expected to occur during the next twenty-four hours. The computer produced a list, for each ward, of the patients' care groupings, predicted admissions and the predicted staffing needs for the coming night shift, the following day shift, and the evening shift the next day (1).

These lists were then delivered to the night supervisor when she came on duty at 21.00 hours. A list of all high-care patients and their location was also produced, and this was found to be very useful to the night supervisors. An additional feature of the system was that of updating the lists if the situation should change radically during the night. If a high-care patient was admitted to a ward, or transferred out of a ward, or died, or if a patient's condition improved or deteriorated to such an extent as to move him in or out of the high-care category, the supervisor could enter this information into the computer terminal and new lists were produced. A new high-care patient list was made available and new staffing lists appropriate to the areas concerned. The staffing lists were used by the day supervisor to deploy staff as appropriate.

The system was used routinely for some months and an attempt was made, though in a somewhat qualitative way, to evaluate its predictions. A modified and extended version of the existing 'quality control' questionnaire was answered daily on every unit by the supervisors. The intention was to monitor the nursing care given, in order to see whether it matched with the prescribed care (which was the basis for the staffing numbers predicted by the system). The results of this monitoring study showed, in a qualitative way, that the predicted staffing numbers (when assigned to the units) provided appropriate care as judged by the nursing policy of the hospital and monitored by the study. It was also shown that when the staffing available was

lower than the predicted requirements, some items of care were not provided as prescribed, and that general ward standards also deteriorated. Although the evaluation was by no means rigorous, it provided some measure of support to the system which was already approved by the Nursing Division.

As a general result of the nursing studies, further resources were made available and more nursing staff of all grades have been recruited. The system was found to be a useful predictor of nursing resource needs; hence it was decided to extend it to routine use throughout the hospital. Since then a new member of staff has been appointed to the Nursing Division, one of whose primary tasks is to undertake the reorganization of this system for daily nursing workload assessment on a much larger scale, for routine use as part of the hospital's information system.

STUDIES IN THE UNITED BIRMINGHAM HOSPITALS GROUP

The general approach, as indicated in the introduction, is designed to be applicable in any nursing environment: the details of the data-collection and grouping of items of care may differ according to speciality, geography, policy, and type of care-facility, and differs as between teaching and non-teaching hospitals.

It was thought by the Chief Nursing Officer to the United Birmingham Hospitals Group that this approach could well be used within the group to identify differences between specialist areas. It was decided that the method developed in Boston could be modified and adapted to suit the British environment, and a system could be developed to assess nursing workload in different specialist wards in hospitals within the group, with the specific objective of determining whether present staffing patterns matched the workload generated by these patients under current medical policies, both in terms of numbers and skill-levels of staff. Unfortunately, we had not the benefit of work-study results within the UBH to incorporate into the scheme. Hence, it became necessary to use the results of other British work-study exercises; in particular, results of an extensive Scottish study (3) were used. Where times for items were not available, these were measured by the nurses taking part in the studies, or MHA results were accepted.

Detailed initial studies were made on eight wards in the Queen Elizabeth Hospital and the General Hospital, Birmingham, in order to develop the method for future use within the group, to compare

workload calculated by this method with staff availability, and to attempt to use the method to make predictions as to staffing needs.

These initial objectives were met; it was shown that the method of approach was acceptable to nurses and gave results compatible with the intuitive assessments of the nursing administration. In fact the method did seem to provide a measuring tool whereby nursing workload could be compared with staff time available.

Again, evaluation of this method was attempted. A 'monitoring' study was carried out simultaneously with the dependency study on two wards at the Queen Elizabeth Hospital. A monitoring form was designed in collaboration with the Chief Nursing Officer, and the Nursing Officer for some of the medical wards, which included detailed questions regarding whether or not various items of care were in fact being provided for patients. The questionnaire was subdivided into five sections:

1. Basic nursing care.
2. Skilled and technical care.
3. General condition of ward areas.
4. Professional surveillance.
5. Ward administration and teaching.

The section on basic nursing care included some detailed observations regarding totally dependent patients which involved the Nursing Officer in considerable effort. For example, she checked whether paients' skin care, oral hygiene procedures, turning, and so on were being done as frequently as had been prescribed by the ward sister when recording the details for that patient in the nursing dependency study. She also made some general observations regarding the basic nursing of other patients on the ward; she checked the distribution of fluids and the fluid balance charts; she checked the comfort of incontinent patients and frequency of bathing and personal hygiene procedures. Various treatments and procedures were also monitored, for example, timing of drug rounds, frequency of suction and chest care procedures, adequate care of dressings, correct isolation nursing procedures strictly observed. All these items were monitored by making frequent visits to the two wards concerned.

The third section of monitoring concerning the general condition of ward areas involved recording facts related to general hygiene policies, fire hazards, patient safety, and the maintenance of facilities and equipment.

Recording factors associated with the adequacy of the surveillance

provided to the patients was considerably more complex. A list of items to be monitored was drawn up and this included checking whether patients had adequate explanations of their treatments and procedures, and whether they were given reassurance and information regarding the nature of their disease. The Nursing Officer checked whether rehabilitative procedures were adequately carried out, whether dressings, drugs, monitors, feeds, and drains were checked as frequently as necessary; she checked for promptness in answering patient calls and attempted to determine whether emotionally dependent patients were receiving sufficient attention.

Clearly the Nursing Officer's judgement enters into this assessment, but since she was herself responsible for *all* the monitoring, the recordings though subjective were none the less consistent.

The final section of items to be monitored, associated with ward administration, teaching of students, and staff supervision also involved the subjective judgement of the Nursing Officer. Items monitored included the availability of time for bedside care training of learners, attendance at ward rounds and conferences, discussions with medical and other staff, and patients' relatives; checking for adequate ward reporting and paperwork, and whether patients' notes and X-rays were readily available, and whether drug cupboards and refrigerators were checked as dictated by hospital policy. The ward sister's views regarding adequacy of staffing levels was also recorded, and whether overtime or meal-break working was necessary.

The monitoring study was made on a general medical ward for a four-week period, and simultaneously on a ward containing neurological, psychiatric, and dermatology patients for the same period.[1]

The results of the study did provide support for the nursing workload measure in that whenever the staffing levels dropped below the calculated values, considerably more of the monitored items were found to be left undone. There is obviously some difficulty in relating these omissions to decreased 'quality of care' in that some omissions are clearly more serious than others. The present study makes no attempt to distinguish between omissions, or to weight some omissions more heavily than others. However, the following observations lead to some insight into the way in which the standards are lowered when workload is not matched by staff availability.

It was found that section 5 items (regarding ward administration and teaching) were the first to suffer when pressure was felt on the

1. Copies of monitoring forms can be obtained from the author.

wards. Almost immediately, also, were the items of section 3 (regarding acceptable standards of safety, hygiene, and general tidiness of ward areas). Here, a great many items were found to be omitted when the wards were under pressure. It is arguable, of course, whether these omissions were of any great significance in decreasing the quality of patient-care; however, the items were only included in the survey because they were considered part of the hospital's nursing policy. As has been pointed out already, the present studies and measurements are not designed to produce some absolute standards of staffing needs for all hospitals; they merely *measure* the needs of a particular environment subject to its own policies, and restrictions, both geographic and with regard supporting services. If the calculated staffing is not available, the detailed results of the studies should enable the managers to identify those policies and restrictions which might be changed in order to provide a better match between workload and staff availability.

The next set of omissions monitored were found to occur in section 4 (professional surveillance). These are unfortunate omissions, as they refer to those aspects of patient-care which are rather difficult to define, but which are regarded by most nurses to be the essence of their responsibilities. However, when the wards are under pressure, some of this surveillance is omitted, because it is easier to leave out than a specific item of care such as a bed-bath. If workload is not matched by staff availability, and policies are not changed regarding specific but possibly less important aspects of nursing care, then omissions will be made by default and not by policy, and the surveillance of patients may well be seriously affected.

Basic nursing items were generally provided adequately. However, in one or two instances it was found that some of the more general aspects of basic care were neglected when the wards came under extreme pressure. For example, fluid balance charts were not always kept up as well as expected by the Nursing Officer, and occasionally skin-care and oral hygiene was not provided as frequently as prescribed. However, the wards were never really understaffed to the extent that basic nursing care suffered. Omissions from the second section, skilled and technical nursing procedures, were very infrequent indeed. In fact the only recording in this section was regarding the timing of drug rounds. These were late occasionally when the ward was under pressure. High-care patients were observed always to receive the care prescribed for them (though during the course of the

study there was never a situation in which the wards were grossly understaffed).

In conclusion, therefore, the original studies made at the Queen Elizabeth Hospital and the General Hospital, Birmingham, showed that:

1. The method developed for assessing patients' nursing needs is acceptable and comprehensible to nursing staff, and is applicable to patients from a variety of wards and specialities in the United Birmingham Hospitals group. (These studies included patients from general medicine and surgery, cardio-thoracic surgery, neurology, psychiatry, dermatology, and dentistry.)

2. The assessment of the individual patients' needs in conjunction with other analyses of nursing work on these wards leads to realistic predictions of nurse-staffing requirements, both in terms of numbers and grades of staff.

3. Confirmatory studies, involving monitoring actual care received by patients, support the validity of the staffing patterns predicted.

STAFFING A NEW UNIT

A new nursing unit consisting of two open wards joined by several smaller bedrooms was to be opened and staffed at the Queen Elizabeth Hospital, Birmingham. One ward was to nurse neurological patients, and the other general medical and some dermatology patients. It was thought that the small bedrooms could be shared by both wards, and could be used for 'high-care' patients. In order to determine the likely number of high-care patients who would need this accommodation, and also in order to assess in advance the probable staffing needs of this unit, it was suggested that a study be made of the type of patients who were to be nursed on the new unit, using the method of measurement already mentioned.

The general medical patients of the consultants concerned were already being nursed on a single ward; the neurological patients, however, were being nursed in beds on four different wards, and some also at the Midland Nerve Hospital, Birmingham. The dermatology patients were at that time situated on a ward which also held neurological and psychiatric patients.

A four-week study of the patients concerned was initiated in these hospitals; this study has been described in greater detail elsewhere (4, 5). In this case a simplified version of the data-form was used, with categories determined as a result of the earlier studies carried out in the Queen Elizabeth Hospital.

The results of the study showed that the number of beds available for sharing between the two proposed new wards would seem to be quite adequate to accommodate the expected number of high-care patients. Staffing needs were also predicted by the study and the new unit was, in fact, staffed according to the results of the exercise. A follow-up study on the new unit is now planned, in order to evaluate the predictions made previously. This will be in the form of a nursing dependency study using the same measure as before, and a concurrent 'monitoring' study similar to that mentioned in the previous section.

With these studies we hope to show that:

1. The predictions made were correct according to the method used (ie that the number of high-care patients on the new unit is consistent with the predictions).
2. That the staffing numbers (as predicted) seem adequate, as tested by 'monitoring' the care provided.

In conclusion, therefore, it may be stated that this method of assessing nursing needs has been used, and used successfully, for a specific purpose, ie planning and staffing a new nursing unit. Since the composition of one of the wards concerned was to be different from any other at that time in the hospital, it would have been difficult to assess the needs of that speciality in comparison with any other ward without some sort of generally applicable measuring tool such as this one. The proposed follow-up studies should yield useful information as to the detailed reliability of the measure used as a predictive device.

STUDIES AT A MATERNITY HOSPITAL

It was suggested that the present method of measuring nursing workload might be applied equally successfully to a maternity hospital. In Birmingham Maternity Hospital there were some problems regarding the staffing of an antenatal unit during the night. It was thought that a study of the type already mentioned could throw some light on to the matching of workload by staff availability as existed at that time.

Naturally, the data-forms used in the general medical and surgical areas discussed previously were no longer appropriate in detail. However, since the methodology remains the same in any nursing area, it was necessary only to redefine the actual items of care to be recorded on the data form, and to re-acquire the necessary back-up data which forms the quantitative basis for the calculations.

This was done in collaboration with the Principal Nursing Officer

and the nursing officers whose units were scheduled to be studied. These consisted of:

1. A professorial unit, comprising two wards, one antenatal, one post-natal. (The professorial delivery suite was not included in the study.)
2. A consultant ward unit, comprising two wards, both of which held antenatal and post-natal patients.
3. The main delivery suite and intensive care unit.

Since the particular problem in the maternity hospital had been stated to occur at night, the detailed data forms were completed twice a day in these studies, ie during the day and during the night. Clearly, due to the special circumstances of obstetric care, the workload during the day and night, particularly on the main delivery suite, were often indistinguishable. The studies were continued for a period of twenty-eight days (after initial trial periods of about a week in each ward). The data-recording on the main delivery suite was very extensive, particularly regarding the frequency of certain observations and treatments.

In general the results of the study showed:

The professorial unit. Although the workload at night was generally balanced by staff availability, the fact that all the antenatal patients were situated in one ward, and all the babies in a nursery at the end of the other ward made the comparative workloads of the two wards rather uneven. Hence staff from the antenatal ward were helping with baby-feeds and other baby-care and leaving the antenatal patients at times unsupervised. This is against hospital policy. As a result of the study, it was proposed that the babies be situated in an area near the antenatal patients, in order to even out the workload at night.

There was little evidence of imbalance between workload and staff availability on this unit during the daytime, though the calculations did show that the ratio of skill-levels of staff might possibly be altered in favour of introducing more nursing auxiliaries to provide some of the basic care.

The consultant ward unit. The results of the study showed a consistent imbalance between workload and staff availability during this time. There seemed to be a shortage of staff of each skill-level on these units, both during the day and at night.

Since not all the wards in the hospital were included in the study it was difficult to say whether there was any spare capacity anywhere else in the hospital. However, if it were so, it was suggested that a 'hospital pool' of nurses if formed might be able to provide extra staff for this unit when the workload was exceptionally heavy.

Since no *additional* staff could be made available in general, it was necessary to consider possible changes in policy which might improve the situation on this unit. It was noted that the amount of time spent by qualified midwives on medical ward rounds was very high considering the imbalance of workload and staff availability. Clearly, some of the nursing policies of the hospital were not being upheld, due to this imbalance. It was questioned whether time on ward rounds could not perhaps be reduced in order to free the nurses for other tasks.

Partly as a result of these discussions, and partly because of the nurses' feelings that 'patient care would be improved' it was decided to split this unit so as to accommodate all the antenatal patients on one ward and all the post-natal patients on the other, but sharing the baby-care at night (in order to avoid the problems of the professorial unit). It was considered that this strategem would considerably reduce the time spent on ward rounds, as each ward would then have only half as many consultants visiting. Follow-up studies on this unit and on the professorial unit are planned for the immediate future, as soon as the units are considered to have 'settled down' after the changes. These follow-up studies should indicate whether the changes in management instituted as a result of the nursing studies have had the desired effect, ie improved use of available nursing resources.

The main delivery suite and intensive care unit. As would be expected, the care requirements of the patients in this unit were far in excess of any other ward studied. The most striking observation was the enormous fluctuation in workload, both during the day-time and at night. In fact, it was shown that the workload fluctuated on the main delivery suite and intensive care unit to such an extent that in the worst possible case, there was a shortage of 6 nurse-shifts (mainly skilled) and in the other extreme an excess of 9 nurse-shifts available (again, mainly skilled nurses) between 08.00 and 21.15 hours. The same sort of fluctuation occurred at night, the worst cases being a shortage of 8 nurse shifts (3 basic, 4 skilled, 1 technical) during a night (21.15 to 08.00), and in the other extreme an excess of 5 nurse-shifts (2 skilled,

3 technical) on another night. However, it was shown that the average values for staff availability and workload agreed very well; but there was no evidence that the peaks in workload could be predicted, and met by redeployment of staff from elsewhere. Perhaps with certain types of more predictable delivery, and studies to establish the nursing workload of these deliveries, it may be possible in the future to schedule extra staff on to the unit when peaks in workload are most likely to occur, though, of course, it is unlikely that a perfect match can ever be achieved. Again a hospital pool of nurses could be used with advantage here.

It was noted during the survey that a great deal of nursing time was being spent with patients receiving epidural analgesia. Closer analysis of the data collected on the intensive care unit showed that a patient being delivered after epidural analgesia generated twice as heavy a nursing workload as other patients. The introduction of new patient-care policies and new medical techniques are bound to have some effect upon nursing workload. In some cases workload will be reduced but very often the introduction of new procedures and techniques generates a heavier technical nursing workload. Moreover, the obvious commitment of the nurses to caring for patients undergoing new or specialized procedures is inclined to take precedence over the normal, more commonplace nursing tasks, particularly those in the more nebulous areas of commitment described here as 'surveillance'. This may be the best policy in fact, bearing in mind the limitations which exist regarding staffing numbers available, but it ought to be considered in a positive way. No-one would deny the benefit to patients of some of these new procedures, but their effect upon nursing workload ought to be evaluated, and nursing policy formulated with these new commitments in mind. It ought to be a conscious policy change that some aspect of care previously provided now be allocated a much lower priority, in order to release sufficient nursing resources to cope with the new commitments. Similarly, in many cases, patients' length of stay has been reduced by changes in medical policy regarding the management of obstetric patients. This too has changed the pattern of the nurses' workload, without this being taken into account when formulating nursing policy. The danger is, therefore, that if workload is not matched by staff available, the decisions regarding the 'lowering of nursing standards' are left to the most junior nurses who simply skimp or omit those items of care which are least likely to be noticed. Far better, therefore, whenever changes in clini-

cal policy, administrative procedures or supporting services are proposed, that their effect upon nursing workload be evaluated, and new nursing policy formulated accordingly. The particular example of the increase in the number of patients undergoing epidural analgesia serves to demonstrate how some changes in nursing policy must be made in order to cope with the increased workload. If it is impossible to increase the allocation of staff to the unit, it will be essential to reassess the priorities existing at present, and try to provide best possible care with existing resources. The present method of measuring patients' nursing needs is a very useful tool in making evaluations of this kind. Follow-up studies after changes in policy have been effected can also show to what extent the changes have had the required effect. Until some sort of generally accepted 'standard of care' exists for all hospitals, comparison between different hospitals is very difficult. At least the present method enables a hospital to judge itself by its *own* standards. The general nature of the approach allows considerable extension of the method into other nursing care areas.

Further applications and developments

The ultimate objective of the present programme of research is the development of a 'package' which allows any nursing area to assess its patients' nursing needs and relate them to staffing requirements. Such a package will comprise a well-validated formalism for analysing the nursing workload with appropriate numeric data and work measurement results where applicable to particular types of hospital; it will include data-collection forms for assessing the patients' nursing needs and computer program (or set of instructions) for calculating the staffing numbers. It will also provide details of evaluation procedures, if relevant. Before such a package can be put together however and recommended with confidence for general use, it is necessary to test the validity of the method of assessment further, in as many different nursing environments as possible. It should be tested in non-teaching hospitals, on different medical specialities, and in hospitals with various different layouts with differing levels of supporting services. Hence the present method of assessment which has been applied successfully at hospitals within the United Birmingham Hospitals group, is currently being used in a number of hospitals outside the group for a variety of purposes. Studies are in progress in two non-teaching hospitals, one hospital at a postgraduate centre and in

another case on a large scale at selected hospitals throughout a region. The objectives of the studies in each case are slightly different; the flexibility of the method and hence its potential value as an information tool are thus amply demonstrated.

To determine staffing numbers of optimum skill levels. This study is being made on a psycho-geriatric unit at All Saints' Hospital, Birmingham, which is a non-teaching mental hospital. The objective is to determine whether the ratio of staff of different skill levels is optimal at the present time. The same methodology is used as in the general hospitals, with the details of the data form modified to include appropriate treatments and procedures for this unit. Again, since there are no work-study results available for this hospital, the Nursing Officer for the unit and the staff there have made some measurements of their own and estimated appropriate indirect and routine care workload factors. No results of this study are yet available, but the initial trial showed that the method itself was applicable in this speciality.

Prediction of required nursing resources. The objectives in this case are somewhat different. The National Hospitals for Nervous Diseases, London, are undertaking a hospital simulation project in order to provide information to assist management decision-making. The simulation model is to be used to determine the service areas within the hospital which are under pressure and hence represent 'bottlenecks' in the management of patients. Also the model is used to predict the effects upon the hospital as a whole of proposed changes in policy designed to improve effective use of resources. The objective is to optimize resource utilization whilst still retaining the high standard of patient-care. The project, which is being implemented by the Management Services Unit for the London Teaching Hospitals, uses a simulation model developed in Birmingham by the author (6, 7) which simulates the way in which appropriate medical and nursing care is provided to individual patients being managed within the hospital and takes account of the limitations of the resources of the hospital.

Since one of the sources of the hospital is, quite clearly, nursing, it was decided that an assessment of patients' nursing needs should be included as a parameter to the model. The present method of measurement has again proved to be sufficiently versatile to be applied in these specialist neurological and neurosurgical areas also. The hospi-

tal simulation model will predict the changes in the numbers of nurses of different skill levels required for each new hospital management policy tested. It was felt by the hospital administration that any change in hospital policy or resources made to improve effective use of the resources as a whole should certainly include an evaluation of nursing needs, as this resource, though not at present under undue pressure, might well become so. For example, a reduced length of patient-stay resulting from the elimination of delays for investigation results could increase the nursing workload to a considerable degree. This factor must of course be included in the evaluation of the proposed policy change. Hence in this case, the assessment of patients' nursing needs is used to *predict* staffing requirements under changed conditions in the hospital. The nursing workload measure was previously used in this way as a parameter to the hospital simulation model at the Beth Israel Hospital, Boston. In this case, nursing was a very scarce resource indeed and also expensive. Hence, since the objective of the exercise in this case was to optimize the distribution of planned new beds between the medical and surgical services, making best use of current resources and minimizing expenditure on additional resources, the inclusion of the nursing assessment was of prime importance. The simulation would then *predict* the increased nursing resources needed for the planned new nursing units, as well as the changes needed in the capacities of the hospital service departments.

Nurse allocation within a region

The Regional Chief Nursing Officer to the North-West Metropolitan Regional Hospital Board decided to test out the method over a large number of hospitals to see whether it could be used to throw some light on the differing needs of different hospitals. By setting up a Working Party to agree in a general way on 'standards' and by specifying those areas of obvious difference between hospitals according to agreed categories (for example, those hospitals who have ward clerks, plated meals, teaching commitments, and so on) it is hoped to be able to produce a realistic comparison by which it will be possible to determine how much any particular hospital falls short of its own standards, and how it differs from others within the area. This, it is hoped, may lead to improved information upon which to base staff allocation as between the groups within the Region.

Evaluating the nursing needs of a hospital with 'race-track' design

The layout of the nursing units at Walsgrave Hospital, Coventry, is of the 'race-track' design, comprising about 50 beds, situated within small (2- or 4-bed) well sound-proofed bedrooms. The units are subdivided into three sections originally intended for 'low-', 'medium-', and 'high'-care patients in a system of progressive patient care. However, due to general changes in medical and nursing policy, the predicted low-care or convalescent patients have not materialized, since they are now normally nursed in the community. Problems have arisen in that the original staffing establishments have proved insufficient to meet the workload now generated by the patients nursed in the hospital; at times it has been necessary to close down beds due to shortage of staff. It is also felt that the higher average dependency observed for the patients than was originally expected is partly due to the fact that the additional surveillance needed for patients nursed in an area having this particular layout was grossly underestimated.

Hence the PNO at Walsgrave Hospital, Coventry, decided to set up a study in order to determine the extent to which the staffing numbers should be increased to provide an acceptable nursing service to all the units in the hospital, or alternatively, to identify those areas of nursing workload which might be reduced or eliminated by changes in management policy. This study is to be made, using the author's method of assessment on all the units in the hospital, with a view to assessing the patients' nursing needs and relating these to staffing requirements. As a result of the studies, it is proposed to rationalize the present distribution of staff and to determine the best policy for optimizing use of resources currently available. The analysis of the nursing workload necessary for the application of this method of calculating staffing requirements provides a framework for evaluating various alternative management policies in terms of use of nursing resources, and a viable alternative to closing down beds may well be identified.

Future developments

It is felt that this method of assessing nursing needs has already shown itself to be useful in a variety of circumstances. It has been shown to possess four out of the five characteristics of a 'good meas-

ure' mentioned in the introduction, when used in a number of specialities and different types of nursing units. The most important future development now is to achieve the fifth characteristic, ie to provide a simple and routinely usable 'package' for assessing patients' nursing needs and hence the staffing requirements for any nursing area.

In order that this package may be recommended for general use in all areas, however, some work measurement is needed to establish a firm quantitative background for the workload calculations. Research studies in this field are planned in Birmingham; the effects of dependency factors are to be studied and assessed in some detail, and the extent of professional surveillance and its overlap with other items of care considered.

Also, the method of evaluating the results obtained for staffing requirements from the original assessment of the patients' nursing needs will be developed in greater detail; weighting factors should be associated with omissions of various items of care monitored so that the effects of understaffing can more realistically be evaluated in terms of reduction in standards of patient-care. For example, omitting to turn a patient or provide skin care appropriately on one occasion might well be considerably more important in terms of failure to meet required standards of patient care than forgetting to remove the television plug from its socket, even though the latter duty is part of the hospital's policy regarding fire hazards. Work is continuing in Birmingham on the refinement of this evaluating method and it is intended that it be included in the total package proposed for general use.

It is planned that the nursing-needs assessment package shall contain all the necessary information to enable any nursing manager to set up an assessment study in any nursing area, determine the required staffing and perform an evaluation of the results. The package will include a simple step-by-step guide to the analysis of the nurses' tasks within the area being studied; appropriate data-collection forms will be made available and relevant work-measurement data wherever possible. A computer programme for calculating the workload and staffing values (or instructions for manual computation) will also be supplied, along with a general scheme for implementing the study. In addition, an evaluation mechanism will also be included; checklists and instructions for implementing the evaluation study will be supplied as a part of the total package. Eventually, it is envisaged that simplified versions of the system for common specialities will be

produced, where areas of difference between the nursing policies in different hospitals are identified and accounted for. These simplified schemes can be used for day-to-day deployment of staff or admission of patients in areas where fluctuation in workload from day-to-day has been shown to be great.

When the package is completed and is made generally available, the information to be gained by its use will be immediately valuable. Not only does this method of assessing patients' nursing needs allow individual nursing areas to determine their own staffing requirements according to their own policies and standards, but it also provides a formal framework for comparison between apparently similar nursing areas with different calculated staffing needs. Hence, the effects of different nursing standards and different medical policies upon the nursing workload can be evaluated, and more effective use of available nursing resources made. Eventually, it would be desirable to achieve at least a minimum standard of nursing care in all nursing areas; however such minimum standards are presently ill-defined, and each separate nursing area has its own accepted levels. The framework provided by the present methodology gives a means of predicting the consequences (in terms of workload) of changing the existing standards, and also of experimenting with possible changes in management policy.

Prediction of nursing workload associated with patients of different types has always been an objective urgently sought. If patients' nursing workload could be predicted accurately on a day-to-day basis before admission, this would provide information enabling nurse managers to allocate staff and schedule admissions to optimize efficiency in the use of nursing resources. Unfortunately, such predictions are only possible at present with a few 'types' of patients. One of the main problems is that of classifying patients into 'types'. Diagnostic categories are of very little value in predicting patients' use of resources, including nursing. However, certain minor surgical cases, patients in coronary care units with a defined policy regarding their care, and patients admitted to special short-stay investigative units do in fact have well-defined management plans which are predictable. Further study of the nursing commitment generated by these plans, assessed according to the scheme described in this article, would allow prediction of nursing workload in advance at least for these patients on any unit. Work currently in progress in Birmingham is directed towards the development of formalized problem-

oriented patient-management schemes; extension of this work to a large number of medical problems would lead to more reliable predictions regarding the use of many patient-care resources, not the least of which is nursing. The scheme presently being developed allows medical staff to formalize predetermined management plans for patients with particular medical problems. Hence, the required resources to back up these plans are known in advance. Similarly, the scheme can allow acceptable standards of nursing care to be agreed for each predefined management plan; the resulting nursing workload can then be predicted for each patient being managed according to that plan.

Of course, a fairly general prediction regarding nursing workload can be made, as explained earlier (p. 178) in the description of the study made of neurological patients, predicting staffing numbers for a new unit. In fact, day-by-day prediction only becomes of a high priority if the workload fluctuates violently, and there is flexibility in the number of staff available to meet these fluctuations. However, further studies in appropriate nursing areas will provide sufficient information to set up a simplified version of the system which allows prediction of staffing needs at least over the next twenty-four hours (1). It is hoped to produce a 'set' of simplified versions for different specialities which may be used in similar nursing areas with equal validity.

Conclusion

The methodology described in this article has been shown to be applicable in a variety of circumstances to the assessment of patients' nursing needs, and the prediction of nurse staffing requirements both in terms of numbers and grades of staff. As may be seen from Fig. 2 the method may be used for a number of purposes, with only the actual details of the studies to be carried out differing in each case (2). The method of measurement is a tool which could prove extremely useful as a means of assessing nursing needs at a time when almost all hospitals claim a shortage of nurses. Furthermore, with the approaching reorganization of the health services, some means of measuring 'needs' of all kinds, and of measuring them in a way which is consistent both in the hospitals and in the community, will become a very high priority. In order to evaluate the alternatives for treating patients using different health care resources, it is necessary not only to

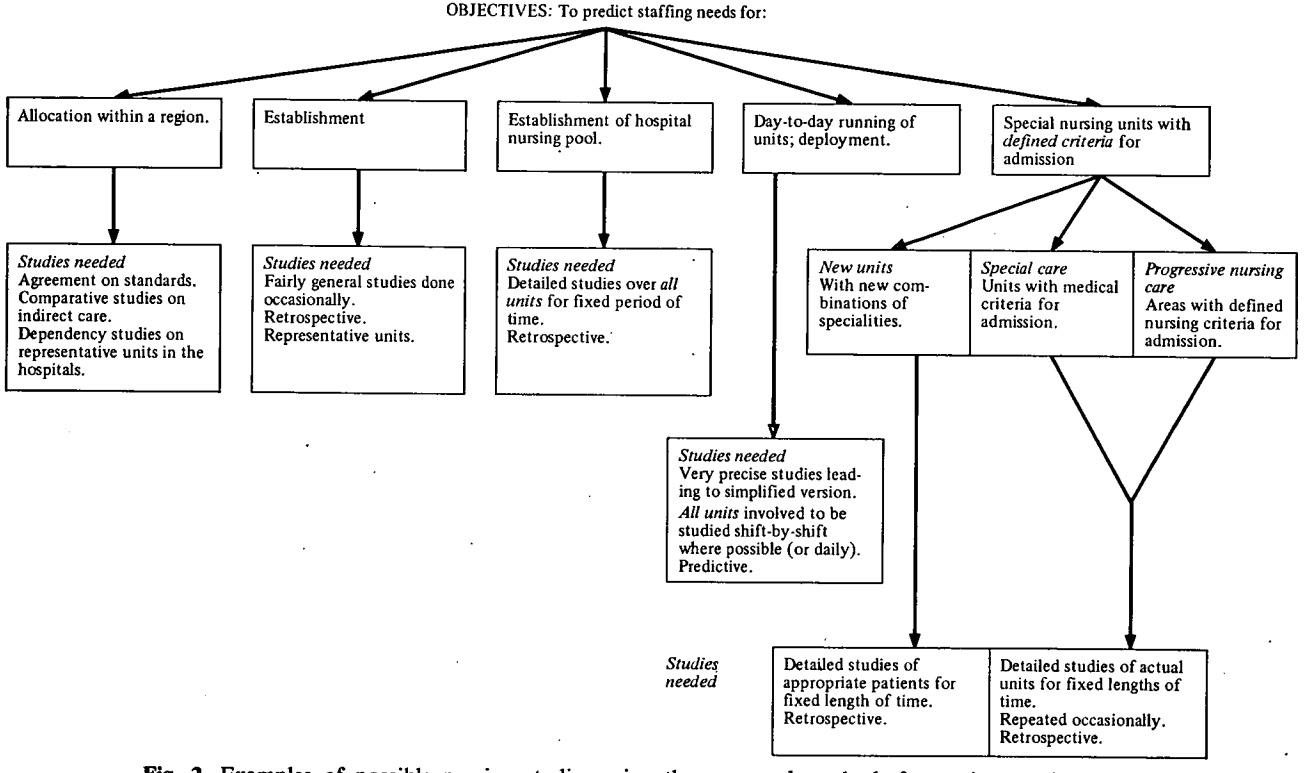


Fig. 2. Examples of possible nursing studies using the proposed method of assessing nursing needs.

examine the costs and outcomes of the alternative management plans, but also to consider whether in fact the alternatives are viable. For example, if patients are sent home from hospital sooner than in the past, not only do we need to consider the changed nursing workload in the hospitals, due to faster turnround and a higher average dependency, but also we should consider the increased workload in the community, and to measure whether in fact the resources are available to meet this increased workload. The present method of measurement is sufficiently flexible to be used in both areas. However, the planned extensions to the present studies, and provision of a 'nursing needs assessment package' should provide an even more robust tool which could be used with confidence by managers at any level to plan the deployment of their scarce nursing resource. The method of evaluation developed as part of the nursing-needs assessment package provides a means of determining whether prescribed care is in fact being given. This indicates in a general way whether staff availability matches nursing workload. Results of studies of this kind have shown that the nurses themselves clearly employ a priority system regarding care given, such that omissions are made in these areas which they consider least important. Unfortunately the surveillance of patients is often neglected when the wards are under pressure. The methodology proposed here offers the opportunity for the effects of other priorities to be tested out in terms of workload generated. For example, it may be questioned whether or not bedfast patients need a bed-bath every day when other patients are not being given adequate teaching and emotional support. Again, the presence of qualified nurses on medical ward rounds may carry a very high priority, but may mean that there is insufficient time to carry out rehabilitative procedures with some of the more physically dependent patients. Formalizing the nursing function into particular areas in this way allows both medical and nursing staff to take a new look at their policies and practices, and reassess their priorities if their standards of care are not being met.

Comparison with policies and practices in other places, and realistic assessments of the standards proposed for each area can produce an entirely more objective viewpoint from which to judge comparative needs of different health service facilities. With further research, integrated with that taking place in different parts of the country, perhaps eventually an acceptable set of nursing standards may be evolved which is generally applicable in all areas, and which could be

maintained by monitoring the patients' 'needs', using the present methodology and allocating staff to different areas according to this measure.

ACKNOWLEDGEMENTS

I would like to thank the Nuffield Provincial Hospitals Trust for their continuing encouragement and financial support. I am also very grateful to Miss B. I. R. Dodwell, CNO, and the nursing staff and administrators of the United Birmingham Hospitals Group for their interest, enthusiasm, and extensive commitment to this programme of research. I am particularly indebted to Miss C. Myerscough, Nursing Officer at the Queen Elizabeth Hospital, without whose tireless efforts the monitoring and evaluation studies could not have been completed.

REFERENCES

1. RHYS HEARN, C. (1971). 'The more fruitful use of hospital resources', in Abrams, M. E. (ed.), *Spectrum 1971—A Conference on Medical Computing*, 57–74 (London: Butterworth).
2. —— (1972). 'Evaluating patients' nursing needs', *Nursing Times*, **63**, no. 17, Occasional Paper.
3. SCOTTISH HOME AND HEALTH DEPARTMENT (1969). *Nursing Workload per Patient as a Basis for Staffing*, Scottish Health Services Studies no. 9. (SHHD)
4. RHYS HEARN, C. (1972). 'How many high-care patients?—1. How Birmingham tackles the problem', *Nursing Times*, **68**, no. 14, 472–4.
5. —— (1972). 'How many high care patients?—2. Deployment of nursing staff', ibid. no. 17, 504–5.
6. —— and BISHOP, J. M. (1970). 'Computer model simulating medical care in hospital', *Br. med. J.* **3**, 396–3.
7. —— (1969). 'Application of computer simulation techniques to patient care studies in a hospital', Doctoral thesis, University of Birmingham.
8. BARR, A., MOORES, B., and RHYS HEARN, C. (1973). 'A review of the various methods of measuring the dependency of patients on nursing staff', *Internat. Nursing Stud.* (in press).
9. BERNSTEIN, A. *et al.* (1953). *A Study of Direct Nursing Care Consumed by Patients with Varying Degrees of Illness* (New York: New York University).
10. LEEDS REGIONAL HOSPITAL BOARD (1963). *Work Measurement as a Basis for Calculating Nursing Establishments—An Analytical Study* (Leeds RHB).
11. CONNOR, R. J. (1961). 'Effective use of hospital resources', *Hospitals*, **35**, 5 30–39.
12. RESEARCH UNIT, OXFORD REGIONAL HOSPITAL BOARD (1965). *Measurement of Nursing Care*, Publication no. 9 (Oxford RHB).

13. NORWICH, H. S., and SENIOR, O. E. (1971). 'Determining nursing establishments', *Nursing Times*, **67**, Supp 17–20.

14. HOSPITAL SYSTEMS STUDY GROUP (1967). *Nursing Study—Phase 1* (Saskatchewan: University Hospital, Saskatoon).

15. MCDONNELL, T. A. K., *et al.* (1968). *Timing Studies of Nursing Care in Relation to Categories of Hospital Patients* (Manitoba: Winnipeg Hospitals, Winnipeg).

16. OPIT, L. J., and CROSS, K. W. (1973). 'Comparing hospital costs', *Hosp. and Hlth Serv. Rev.* **69**, 2, 44–49.

THE FUTURE—AND PRESENT INDICATIVES

This collection of essays the ninth in the *Problems and Progress* series has a common theme—some vital issues which depend on the adequacy of information for the review and development of medical care in the future.

Altogether the various issues are of particular importance at this time on the eve of the establishment of the new authorities in the reformed NHS, for while these are not to come into legal existence until 1 April 1974, certain policies will doubtless be formulated in advance of that date.

The particular accent on the future but with present indicatives, is struck by Sir George Godber's brilliant essay on the future of epidemiology which is the leading article in the volume.

£2·50 *net*

ISBN 0 19 721380 4