A STUDY OF ANAESTHETIC EMERGENCY WORK
BY
T. H. TAYLOR, A. M. C. JENNINGS, D. A. NIGHTINGALE, B. BARBER, D. LEIVERS, M. STYLES AND J. MAGNER

PAPER I:
THE METHOD OF STUDY AND INTRODUCTION TO QUEUING THEORY

Little information has been published on the staffing structures of anaesthetic departments, either for routine or emergency work. The routine work is normally predictable within certain limits, and time and appropriate staff can be allocated for it. Emergency calls on the personnel of anaesthetic departments, however, occur throughout the day and night and are very variable, both in occurrence and degree of urgency. In times of staff shortage routine work may be reduced at the expense of those patients on waiting lists but the emergency workload still remains. In a previous study (Nightingale and Taylor, 1966) the discrepancy between the total working time of anaesthetic junior staff and the time spent in operating theatres was explained as being due to emergency on-call duties. No attempt was made to evaluate the anaesthetic services offered or required. In fact, little has been written to define the whole range of interests and duties of anaesthetists. The authors feel that a good definition is that proposed to the United States Department of Labor (Dripps, 1966), which enumerates the obligations and responsibilities of anaesthetists as follows:

- Anesthesiology is a practice of medicine dealing with (1) the management of procedures for rendering a patient insensible to pain during surgical operations; (2) the support of life functions under the stress of anesthetic and surgical manipulations; (3) the clinical management of the patient unconscious from whatever cause; (4) the management of problems in pain relief; (5) the management of problems in cardiac and respiratory resuscitation; (6) the application of specific methods of inhalational therapy; (7) the clinical management of various fluid, electrolyte and metabolic disturbances.

With the development of intensive therapy units and cardiac arrest call systems, the commitments of anaesthetists, especially to the last three categories, has increased in two ways. First, the work load has increased in absolute terms and, second, a greater proportion of the work is more urgent than hitherto. Thus it has become important that some analysis of emergency anaesthetic work should be undertaken to show (1) the ability of the staff to provide a satisfactory service and (2) the effects of increasing commitments on what is in many hospitals an already undermanned service.

The empirical way in which most departments are organized results in wide variations in the individual member's work and responsibilities and in the quality of service provided. The purpose of the present study is to examine in detail by statistical methods the emergency commitments and staffing of three widely differing hospitals: a London teaching hospital, a district hospital in the Midlands, and a specialist hospital in the North of England. The three hospitals together account for 1.6 per cent of all acute beds in England and Wales and the number of anaesthetists employed is 1.5 per cent of all consultant anaesthetists and 2.4 per cent of junior staff (Hospitals Year Book, 1968; Ministry of Health, 1967).

THE WORK LOAD

Against a background of work which can properly be done when an anaesthetist is next free, there is a varying load of more urgent or very urgent work, such as is required in the treatment of foetal distress or cardiorespiratory arrest, which must take priority if life is to be saved, although not to
the detriment of cases already started. Whether these emergencies are dealt with within “reasonable” time depends on the pattern of the work and the available staff. Obviously it is impractical to provide, say, 10 on-call anaesthetists against the unlikely event of 10 simultaneous very urgent cases. Indeed, whatever number of staff is provided, it will always be possible in theory for there to be a greater number of simultaneous emergencies and no finite number of staff will eliminate this probability completely. It is for this reason, and not through any unwillingness to accept the necessity to provide a complete service, that the writers have chosen to express the chances of a breakdown as one of their measures of the service that might be provided. The study as a whole has been prompted by a desire to express quantitatively the relationship between various arrangements of different numbers of staff and the effectiveness of the resulting service.

It is general experience that the emergency work load varies considerably from day to day and at different times of the day. In order to determine what length of survey would be likely to produce a statistically valid sample, a 7-day pilot study was undertaken in April 1967 at two of the hospitals concerned (The London Hospital and Northampton General Hospital). Analysis of the data obtained indicated that a minimum of 6 weeks was necessary. In the light of experience gained during the pilot survey, modifications were made in the method of data collection and for the definitive survey an 8-week period between September 25 and November 20, 1967, in all three hospitals was decided upon. This time of year was chosen as the period when holidays were at a minimum and seasonal factors were unlikely to have a major distorting effect on the work load.

COLLECTION OF DATA

The individual anaesthetists were asked to fill in a daily form of all emergencies undertaken and routine work performed after 6 p.m. whilst on emergency duty. The following details concerning each case were recorded:

(1) the time of initial request for anaesthetic work; (2) the time of start of anaesthetic work; (3) the time of completion of anaesthetic work; (4) the degree of urgency of the case; (5) the grade of anaesthetist undertaking the work; (6) the type of work; (7) whether an accident or not; (8) the location of the anaesthetist when called.

The “request time” is the time of first learning of a case rather than the time arranged for it to be treated. The latter time is so often the result of an informed guess of when the anaesthetist will be free from work already arranged, and sometimes involves consideration of time for meals and other activities, that it is therefore of little statistical value.

The times of start and completion include the whole time that the anaesthetist is occupied with the case and unable to do any other work.

Each call was allocated to one of five categories of priority:

1) Immediate: requiring service within 3 minutes, e.g. cardiorespiratory arrest (including neonatal resuscitation).

2) Cases not quite so urgent as (1) but requiring service within 15 minutes, e.g. prolapsed umbilical cord, uncontrolled haemorrhage, etc.

3) Emergencies needing the next available anaesthetist, but with no definite time limit, e.g. perforated appendix.

4) Other emergencies which could wait several hours (if unavoidable); many uncomplicated fractures fall in this category.

5) Routine work carried out in emergency time by a duty anaesthetist.

The urgency of the cases was recorded, first, as stated by the person requesting the service and, second, as assessed by the anaesthetist who dealt with the case. Any failure of the anaesthetist to respond to calls in the two more urgent priorities, to keep to pre-arranged times for the treatment of patients with priorities (3) and (4), or if he was forced to leave a patient unattended on the operating table, was specially noted. The completed forms were checked daily by one of the authors and the accuracy of the data was assessed on the next day while events were still fresh in the minds of the individuals concerned. By this means it proved possible to attain a high standard of accuracy and complete coverage of all relevant anaesthetic work. All the information from each case was transcribed on to an “incident form”. This was designed to facilitate the transfer of data on to punched paper tape for input to the computer. The London Hospital’s Elliott 803 computer was
used to store the data in chronological order on magnetic film. The file of information thus obtained was then used for studying various aspects of work load and staffing with the aid of a series of specially written analytical programmes. The results so obtained will be presented in later papers.

**ANALYSIS OF DATA**

To obtain worthwhile information to assist in planning a service it is necessary to use statistical methods for the analysis of the data collected. In particular, use has been made of the concepts and results of a branch of applied mathematics known as "queuing theory". This theory involves the quantitative analysis of particular systems liable to congestion and allows certain predictions, such as the probability of waiting under various circumstances, to be made. By considering examples it is possible to illustrate the method and the results obtainable, without undue stress on the rather complicated mathematics which arise as soon as the system under investigation is examined in detail.

The theory of queues has been developed over the last 60 years, mainly for commercial applications. Pioneer work on congestion in telephone exchanges was done by Erlang in 1907 and since then a large literature has accumulated. Doig (1957) and Saaty (1961) gave references to papers on queuing theory prior to 1960 and Saaty (1966) brought these bibliographies up to date with about 300 more references covering the succeeding 7 years. In medicine the theory has been applied to the problems of waiting in out-patient clinics (Bailey, 1952, 1955; Blanco-White and Pike, 1964; Nuffield Trust, 1955, 1965). The simplest queuing model is sufficient to explain most aspects of the behaviour of out-patient clinics but more complicated models have also been analysed (Pike, 1963; Barber, 1964). At The London Hospital attention has been turned from detailed studies of individual clinics to the recording of studies of all clinics, using the computer, thus enabling them to be continuously monitored (Barber, 1967, 1968). However, there are striking differences between an out-patient clinic and the emergency anaesthetic service. The latter is, of course, a continuing process, whereas the clinics only exist over a limited period, starting afresh at some later date. Also, the anaesthetic service is organized so that the demand is but a small fraction of the actual provision, so as to ensure an adequate service to the patient. In contrast, the appointments for out-patient clinics usually ensure that the demand during the first hour of the clinic is substantially in excess of the service capacity, providing a rather unsatisfactory service to the patient, but making sure the doctors are occupied. Any attempt to arrange a continuing process in this way is clearly doomed to failure but if things are arranged so that the service capacity exceeds demand then the system will not fail. While individual patients may have to wait, on the whole the chances of waiting and the length of wait will approximate to certain equilibrium values. The proof or disproof of their existence and the calculation of their values are some of the tasks of queuing theory.

The extent to which service capacity exceeds demand is, in a certain way, waste. In principle, it is desirable to minimize this and the theoretical optimum which comes to mind is perhaps the point at which there is a minimum number of anaesthetists on call, but at which an adequate standard of patient care is still provided. Queuing theory can assist in selecting this point.

It can be shown that:

1. No equilibrium exists if demand outstrips service facilities.
2. Conditions of statistical equilibrium exist when demand is less than service facilities.
3. The chances of waiting, the mean waiting time of the patients who do wait, and the proportion of patients who wait more than any specific time, can be calculated if certain simplifying assumptions are made about the distribution of operating and inter-arrival times.
4. In the system in which demands arise at random and the service distribution follows a negative exponential, which fits the emergency anaesthetic situation quite well, the mean waiting time of those that wait is given by

\[ T_w/m(1-r) \]

where \( m \) is the number of service channels provided and \( r \) the ratio of demand to service and \( T_w \), the mean service time.
Out-patient clinic studies have, in fact, shown that a ratio of demand to service of about 80 per cent results in patients not waiting unreasonably. That is, if the average consultation takes 8 minutes, patients should be booked at 10-minute intervals when one doctor is present or every 5 minutes for two doctors and so on.

This leads to the other major difference between the planned "cold" service described and the emergency system under consideration. With utilization of 80 per cent and with one server, the waiting time of patients who wait is five times the mean service time and 80 per cent of patients do wait. If four servers are involved in a system with the same utilization (i.e. 80 per cent) then the mean waiting time of the patients who wait would be 1.25 times the mean service time and only 60 per cent of patients would wait. The overall mean service time of cases occupying an anaesthetist in the three hospitals in the present survey was about 45 minutes (see later papers). An 80 per cent occupancy is thus quite unacceptable if much of the work is urgent.

Application of queuing theory to the anaesthetic service.

The occupancy and mean wait which are likely to be attained in practice is best illustrated by examples derived from the present survey.

Figure 1 shows the arrival pattern of emergency work into the anaesthetic department of Northampton General Hospital (582 beds) between 9 a.m. and 9 p.m. on October 10, 1967. The lower line shows what would have happened had there been one person on call who responded by attending to a patient immediately, if free, or in order of arrival as he became free. It can be seen that two patients waited, one for 20 minutes and the other for 3 hours. It so happened that the first emergency (arrowed) was a cardiac arrest which was dealt with immediately because the server was free. If it is desired to know what chance there is that such emergency will be seen immediately, it is the chance of finding the service channel unoccupied. Assuming random arrival of urgent cases, this is the ratio of unmarked parts of the lower line in the total length, or 55 per cent on this day.

One day cannot be taken as representative, hence it is necessary to take a larger sample and average. It was found that over the 8-week period of the survey at Northampton approximately 20 per cent of the whole time was occupied on emergency work. Assuming again that the more urgent cases occur randomly, it can be inferred that, with one person on emergency call, 20 per cent of urgent cases must wait and 80 per cent can be treated immediately. With a mean case length of 33 minutes at Northampton, the mean wait of the 20 per cent who do wait would be 41 minutes. Furthermore, application of queuing theory shows that two-thirds of those waiting (12 per cent of the whole) will need to wait more than 15 minutes. A proportion of these cases would be in the two more urgent priorities and a satisfactory service could not be offered for these with only one on-
call anaesthetist. In arriving at these conclusions the writers have not considered how many were, in fact, on call. This information is not needed to arrive at the result. However, it can be said with reasonable confidence that something like this would have happened. The experiment of having only one on-call anaesthetist has been performed, not by trial of the anaesthetic system with its consequent risk to patients, but by calculation on a model situation.

Figure 2 shows what would have happened with one person on call on September 28, a much busier day. The call arrowed was an urgent case of severe haemorrhage occurring 1 hour after vaginal hysterectomy and requiring immediate re-exploration under general anaesthesia. With one anaesthetist on call this case would have had to wait 30 minutes, even though it would have taken priority over a Colles' fracture notified 2 minutes earlier. In fact the waiting period was only 5 minutes because someone else was available.

Figure 3 shows the service which would have resulted from having two people on call on that particular day.

In this way it is possible (but very time-consuming) to work out the occupancy of any number of...
servers over the 8-week period. Alternatively, with enough information about the pattern of arrivals, the average case length and spread about this average, queuing theory can be used to calculate the probability of the patients having to wait. The formula given earlier can be modified to take account of a priority system, and the chances of waiting more than 3, 15, 60 or any number of minutes for various categories of priority can be calculated. The calculations are more complicated but the advantage of using queuing theory remains. The consequences of any staffing arrangements can be explored and as much information as possible can be extracted from a limited and manageable survey.

In order that the calculations should not be so difficult as to be insoluble, certain simplifying (and therefore possibly distorting) assumptions have been made. Provided that these distortions are not too great we can expect the predictions to be of the right order, if not exact. In particular the writers will be concerned with comparing various possible arrangements to see if some answers are markedly better than others. For the purpose of making decisions concerning the organization of staff this is the important factor.

Basically, only two variations are possible, namely the number of anaesthetists on call and the apportioning of duties between those on call. A lower occupancy of anaesthetists' time is the result of improving the service by increasing the number on call. Here the price of a better service is waste of time. It is possible to measure the waste against the consequent improvement in the service. The answers may be inexact but if at any point there is a steep rise in waste to be set against a small increase in quality of service something helpful in planning will have been obtained.

There may be arguments in favour of reserving one or more anaesthetists for cases of particular types. These may be the very urgent (cardiac arrest), those in a particular situation (such as an intensive therapy unit), midwifery cases, or cases of predictable brevity. By applying queuing theory an attempt can be made to put these in order of effectiveness in terms of quality of service according to various types of cases. Unless the number of persons on call is increased, any rearrangement of duties can only improve the service to one group at the expense of the others. However, it is possible to predict what price is being paid and where. With this information decisions on staffing arrangements can be taken with greater confidence.

Statistical theory cannot make the final decision. Like Oscar Wilde's cynic, it will tell us the price of everything and the value of nothing. The "value" decision is uncomfortable but has to be made, whether consciously or by default; it will be better made if the best information is available to those whose duty it is to decide. The provision of such information is the aim of this investigation.

REFERENCES


*Hospitals Year Book* (1968), Section 8 (Statistics), 401.


