Discussion Topic: Waiting Lists and the Supply and Demand for Inpatient Medical Care*

by

J.G. Cullis

* Acknowledgement is made to the Department of Health and Social Security for a grant to the Institute of Social and Economic Research at the University of York for research in the economics of hospital waiting lists.

Introduction

Despite an 87% increase in the number of cases treated since the inception of the NHS in 1948 there was a total of a half a million people on inpatient waiting lists in 1971. This is the basic problem with which this discussion paper is concerned: that the number of people waiting for inpatient treatment has appeared to be largely insensitive to the volume of medical services provided.

One of the motives behind the formation of the NHS was that individuals should receive medical advice, care and treatment "... and that their getting of these shall not depend on whether they can pay for them." As a result for the last twenty five years the population has had access to health services at almost zero money cost by signing on the list of a general practitioner - about 97% of the population in England and Wales have done this. However it is a basic economic principle that the lower the price of a good or service the greater will be demand. So that in this case the demand will increase to the rate at which the marginal valuation of additional units of medical care is zero. If the supply of medical care is less than this 'saturation' level there will be excess demand which will not be eradicated by the market response of rising money prices but rather take the form of a queue of people who are paying the non-money price (partly a time price) that equate supply and demand. Let us consider this argument in more detail under the economists favourite heads of supply and demand.

The Demand and Supply of Medical Care

Considering an individual's demand it seems reasonable to follow Berki (1972) by distinguishing between expected and perceived health states and defining a health state disequilibrium as occurring when the former exceeds the latter. The individuals expected health state at any given time H_t^* is a function of several variables.

$$H_t^* = f(A, E, R, S, N, H_{t-1})$$

where A is age, E is education, R is role expectation, S is societal and family status, N is cultural norms and H_{t-1} the individuals previous health state. At any point in time the individual also has a perceived health state H_t and when $H_t^* - H_t > 0$ a health state disequilibrium exists with an incentive to adopt one or more of the following courses of action - wait for the disequilibrium to correct itself, self mediciation, seek information from non medical sources, seek help from the NHS normally through a G.P. in non emergency cases. Let us assume for the community that the extent of health state disequilibrium that become known to G.P.s and are considered serious enough for treatment

$$\alpha \sum_{i=1}^{n} H_t^* - H_t$$

where n is the population size

The supply of hospital medical care will depend on human agents, (the number of consultants, registrars, other hospital doctors, nurses and other staff), physical capital (the number of hospitals, available beds, operating theatres, X-ray and other machines) and the state of technique. So that

$$S = f(H.A., PC, T)$$

where the letters refer to each of the categories above.

Combining the demand and supply sides we have the extent of health state disequilibrium to be treated in period t as

$$\alpha \sum_{i=1}^{n} H_t^* - H_t$$

of which some proportion β will be treated by the G.P. so that the demand for inpatient care becomes

$$(1 - \beta) \quad \alpha \sum_{i=1}^{n} H_t^* - H_t$$

Of this demand some proportion γ will be referred back to the G.P. for treatment or treated on an outpatient basis so that the final demand for hospital inpatient care becomes

$$(1 - \gamma)(1 - \beta) \; \alpha \sum_{i=1}^{n} H_t^* - H_t$$

which we can assume requires S^* units of medical care to eradicate, whilst from the supply side we have S units of inpatient medical care available which can be taken to be less than S^* so that

$$S^* - S = D^*$$

where D^* is the excess demand for units of inpatient care. By dividing though by the average number of inpatient units consumed per case, A, we have derived the size of the inpatient waiting list in period t, $\frac{D^*}{A} = W_t$, as a direct measure of excess demand for inpatient care. If any lists inherited from previous time periods are added to W_t then the data becomes a measure of the backlog of excess demand. Hence waiting list data could be compared with supply experience in some appropriate way to see if excess demand increases/decreases more or less than proportionally to supply changes.

This picture is false to the extent that no account has been taken of the influence of G.P.s and consultants on waiting list construction. It is implicitly assumed above that α, β and γ are determined objectively by medico-technical factors. However it is well known that G.P.s and consultants act as arbiters rationing the available supply of medical care. G.P. referral decisions are likely to be different depending on individual decision thresholds and the length of the inpatient waiting list. Similarly different consultants are likely to have different decision thresholds both at any point in time and over time. These considerations considerably complicate the simple approach presented above since the supply and demand sides can no longer be assumed to be independent of one another. However as long as the recorded waiting lists bear a fairly constant relation to the 'true' waiting list it is justifiable to use them as a proxy for excess demand for inpatient care. This was done using several equations with recorded waiting lists or changes in recorded waiting lists as the dependent variable. Several sets of data were used: both time series and cross section. The variables were adjusted for population size to eliminate any scale effect.

Feldstein's two equations

$$W = \alpha + \alpha_1 S$$

$$W = \alpha + \alpha_1 S + \alpha_2 S^2$$

where S the supply term is the number of beds used per 1000 population provided a starting point. For time series data the equations were modified to include time as a variable. The supply term was considered to be weak in that it takes no account of differences/changes in the average duration of stay so that a new supply variable S_c, defined as the number of available beds (daily averages) divided by the average duration of stay, was substituted into the equations. None of these equations yielded significant coefficients or acceptable adjusted R^2.

A second broad approach centred on staffing levels as having a major influence on the length of waiting lists. A basic equation run against several sets of data was

$$W = \alpha + \alpha_1 C + \alpha_2 R + \alpha_3 J + \alpha_4 S_c$$

where C was the number of doctors in the consultant categories, R was the number of doctors in registrar categories and J the number of doctors in the junior categories. The coefficients were insignificant and the adjusted R^2 very low or negative even when several other variables were added (separately). The other variables were new outpatients, existing outpatients, total outpatients and the number of surgical operations. The latter being used as a proxy for theatre capacity, the bulk of waiting lists being concentrated in surgical specialties. Another equation just contained a constant, the number of G.P.s and the number of consultants.

A third approach tried to test if waiting list decision takers were influenced by past changes in the waiting list and changes in supply capacity so that the coefficients of the equation

$$\Delta W_t = \alpha + \alpha_1 \Delta W_{t-1} + \alpha_2 \Delta S_c$$

were estimated against five sets of time series data. The coefficients proved to be insignificant and the adjusted R^2 generally negative. Some of the earlier equations were also run in the form of first differences or with the square of the variables introduced but with very little or no success.

The Models and the data

None of the models yielded econometrically successful results, for which there could be a number of explanations. Although in theory the waiting list should be an expression of excess demand the influence of the health service arbiters - G.P.s and consultants - distorts waiting lists in such a way as to make them a largely meaningless statistic. On the other hand it is possible that waiting lists are a good measure of excess demand but none of the models specified the supply side correctly or sufficiently well to expose a relationship. Thirdly it could be argued that the models are far too general and cannot take account of the myriad of influences affecting the waiting list as recorded for an individual consultant. This, perhaps, could be done with a far more micro study of only several individual consultants but the weakness here would be the extent to which the results could be generalised.

As regards the data, several points can be made. From the patient's point of view the time spent on the waiting list is of greater importance than the number of people waiting alongside him so that waiting times may have been the appropriate data to use. However it is worth noting that the waiting list and the mean waiting time are not completely independent of one another as one would expect ($r = 0.65$ approx). The collection of evidence on the severity of waiting list cases would enable the 'weighting' of waiting lists e.g. using a three or five point scale. If this sort of information was available it would be possible to judge if the NHS was overtaking the demand for care by seeing if the waiting list was composed of less serious cases over time. Weights according to severity may have considerably transformed the waiting list data yielding better results. A prerequisite of collecting such data however would be uniform criteria for categorising cases more or less universally applied so that the measuring rod would not bend between individuals and overtime.

It seems clear that increased efforts should be made to make recorded lists more accurate as there is much debate as to whether they are inflated or deflated estimates of those actually waiting. Arguments for the former include the movement of patients to another area where they join a second list or are admitted without their names being deleted. Against this must be set those whose names are removed from the list because of the failure to reply to a letter of enquiry after admission default but remain untreated.

Conclusion

The general conclusion is essentially a negative one that waiting lists are not, on this evidence, systematically related to any of the variables considered. This is, perhaps, not surprising given flexible decision thresholds and the nature of demand. As regards the latter "no country or delivery method within it has dared to engage in an experiment to determine what the saturation point of demand for health services might be. Certainly there is such a point it is inconceivable that all people would be in physicians offices or in hospital beds all of the time if supply were ample. What we do know is that by arbitrary decisions based on quite arbitrary criteria each country (USA, England, Sweden) sets up controls on demands from the professional gatekeepers, the physicians, well before the theoretical saturation point is reached". (Anderson, 1972).

It may well be that the only way to tackle the waiting list problem is to take up the dare and conduct this experiment and 'decide' about what proportion of saturation demand society is willing to meet and then attempt to reduce the waiting list to an "optimum" having placed a finite limit on demand. Implicit in such an approach would be more rigidly laid down and adopted criteria to guide medical practitioners in their actions regarding the waiting list.