

WORKING PAPER 391

WHY DO REFERRAL RATES VARY? SOME THEORETICAL
CONSIDERATIONS ON THE STUDY OF MEDICAL PRACTICE

FRANS BERKHOUT

WORKING PAPER
School of Geography
University of Leeds

WORKING PAPER 391

WHY DO REFERRAL RATES VARY? SOME THEORETICAL
CONSIDERATIONS ON THE STUDY OF MEDICAL PRACTICE

FRANS BERKHOUT

School of Geography
University of Leeds
Leeds
LS2 9JT

June 1984

Why do Referral Rates Vary? Some theoretical considerations on the study
of medical practice

Contents:

1. Introduction
- 1.1 Scope of the research
- 1.2 Historical development and relationships
- 1.3 The problem and aims of this research
2. The referral system in general practice
- 2.1 Reasons for referral
- 2.2 The decision to refer
3. Factors related to the decision to refer
- 3.1 Analysis
- 3.2 Community and resource availability
- 3.3 Factors about primary care
- 3.4 Factors about the patient
- 3.5 Factors about the hospital
4. The interpretation of clinical freedom and some speculations
on a theory of referral rates
5. Review of on-going empirical work
6. Conclusions
- Bibliography

Why do Referral Rates Vary? Some theoretical considerations on the study of medical practice

1. Introduction

1.1 Scope of the research

The research reported in this paper constitutes part of a wider project to construct a strategic planning model for a District Health Authority (DHA). The set of models being constructed include models of demand for health care, the utilization of facilities and the consumption of resources. In this paper we specifically focus on an analysis of the general practitioner (GP) referral system.

This paper is the first of two that examines this component of the health care system. In the second paper we present some detailed results of our empirical work. Here we review the general problem and provide a context for the study.

It is estimated that specialized medicine, centred in hospital in-patient departments, presently consumes up to 80% of the National Health Service budget at a district level (Forte, 1983). The demand for these resources is for the greater part stimulated by GPs through the referral system - 80 to 90% of patients arriving at a hospital do so via a consultation with their GP. The GP is therefore an important regulator of patient flows into the hospital system, and referral decisions, through the mediating influence of the hospital-based consultant, can be seen to have a bearing on

- (a) the utilization of hospitals (inputs of patients)
 - (b) the expenditure on hospital facilities (costs)
- and (c) the quality of care afforded to acutely ill patients.

Each one of the above must be an important component of any analysis of health services. Referral patterns are therefore of key importance in any attempt at comprehensive, systematic planning in the NHS. This paper sets out to provide a more complete examination within the referral process by placing the decision to refer into the context of health service structures at the DHA level and by attempting to understand more deeply the medical and

non-medical factors which affect that decision. This new understanding, together with the results of a survey of GP referral behaviour will then in the second paper be applied to assess the spatial ramifications on referral patterns of changes in the location of inpatient facilities in the Dewsbury DHA.

In the present climate of public service expenditure cuts it is especially important that the link between general and specialized medicine is more accurately understood. It is clear that any optimization of resource use in the NHS will, to a significant extent, be dependent on the well-informed, judicious use of the referral system. Little progress can be made towards better planning of health service resource allocation without a detailed and systematic reappraisal of GP behaviour and their interaction with the hospital system.

1.2 Historical development and relationships

The definition of referral may be given as:

"A permanent or temporary transfer, including sharing, of responsibility for the care of diagnosis of a patient from one primary physician to hospital medical or paramedical personnel" (Loudon, 1977).

The requirement for these transfers of patients, which occur in space, stems from the particular organization of health services in Britain into the tripartite system of community health care, general practice and hospitals. Indeed, the referral system, or the need for the establishment of such a system, has long been seen as central to the ethics of British medicine. Every citizen, it has been argued, should have easy access to a general medical practitioner. The equity-efficiency problem of universal medical coverage and an efficient division of labour could only be resolved by:

"... a primary division into general practitioners and specialists... The role and ambition of the general practitioner should be to know something of everything, the role and ambition of the specialist to know everything about something. Each must recognise the advantages and limitations of his position. The work of each is of vital necessity to the community. Neither must encroach upon the other's province." (L. Shaw, A BMA leader, 1918, cited in Honingsbaum, 1979.)

A rudimentary division of this sort and the development of a referral system had been encouraged by the Health Insurance Act of 1911. Under its

'panel' system GPs were paid a fixed sum per annum for each registered insured person. This stabilized the income of general practitioners and lead them to refer on difficult (and hence costly) patients to out-patients departments in voluntary and local authority hospitals (Gill, 1975).

The advent of the National Health Service in 1948 produced a bitterly contested but more decisive separation between general practitioners and consultants. The final organizational structure of the NHS was determined by the outcome of negotiations between the state and various establishment groups - principally leaders of the medical profession. But the profession itself was composed of three main sectional interests with different bargaining positions and different financial and political goals. Overall, as Doyal (1979, p. 180) observes:

"... the GPs gained little more than a modestly increased and more reliable income [while retaining their status as independent contractors]. It was the specialists who dominated the negotiations and who gained the most."

Consultants were left in control of the administration and running of the hospitals to the virtual exclusion of GPs. Particularly important was their complete control over the teaching hospitals. From this position they played a major role in determining the nature and priorities of medical research and medical education. As a group they were now able to maximize their control over the practice of medicine.

So developed the current structure of health care provision in which nominally everyone is registered with a GP who serves as the initial provider¹ and arbiter of the management of illness. The GP has also assumed a wide range of other apparently less clinical duties such as advice giving and counselling. Above the GP stands the superstructure of specialized medicine based in the general hospital. These services may be called upon through the referral process in cases of acute or complex illness.

¹ The Standing Medical Advisory Committee (1963) estimated that 90% of all medical episodes were handled from start to finish by general practice, this matching with Loudon and Stevens' (1980) estimate of 70-90%.

Between the two branches of the profession, there remains a tension which Honingsbaum (1967) named the 'antiseptic barrier' and we must see that the interaction of the two groups, in referral, works through its mediating influence, an influence based on the authority of the consultant. The scars of the elaborate demarcation dispute of 1948 are still present in the ambiguous linkage between general and specialized medicine. Cooperation and liason over, for instance, resource allocation remains limited, with several identifiable effects. Power in the planning of hospital services resides firmly with the specialists who have a view of medical need which is particular to their situation as hospital-based medical practitioners. This view is reinforced by the need to retain the consultant's status and career structures in the profession as a whole. The hospital serves the consultant's need at least as well as the patient's. A pertinent example of an extension of the consultant's power comes with the support they have given to the closure of small voluntary and 'cottage' hospitals, so restricting the direct access of GPs to beds and also jeopardizing the standards in the continuity of care which were attained through the use of these facilities.

On a wider level, Navarro (1978) argues that increasingly technological, hospital-based medicine has lead to a proportionately decreasing expenditure in general practice. This apparent diminution of the role of general practice has strained the links with specialized medicine as they were originally conceived.

This then is the institutional context through which the referral system is constituted and which is underwritten by a whole range of political and economic questions which will be briefly discussed later.

1.3 The problem and aims of this research

Although at first glance it would appear that the decision to refer was a relatively simple one, related directly to the severity of the illness whether of an individual or mass of individuals, it is found that the numerous data of the referral process reflect wide variations at all levels of investigation. The second National Morbidity survey (1979) showed marked regional variations in referral rates from 14.5 to 38.9 for inpatients, and 67.2 to 160.7 for out patients per annum per 1000 population. These figures

cannot be rationalized by appealing to complementary regional variations in morbidity patterns. There is no evidence to suggest that variations of this order in the occurrence of illness requiring hospital treatment exist. At a lower level of resolution, the DHA level, it is expected that the variations would be even more pronounced and so on down to reported variations between individual GPs from 8 to 205 referrals per annum (Butler, 1980).

At each of these levels, regional, district and the group or solo practice, we are interested in a different, though overlapping, set of variables. At the regional level variations in referral rates are seen at least partly to derive from variations in the provision of services. Regional Health Authorities (RHA), of which there are 14 in England, are structured so as to provide a comprehensive range of health care; only for very highly specialized services such as heart transplants and certain types of neurosurgery are patients referred out of a RHA. The Resource Allocation Working Party (RAWP, 1976) employing fairly coarse measures of resource provision, focussing on capital stock provision, showed that there were still serious disparities in the provision of services between relatively well provided for regions such as the four Thames authorities, and relatively deprived regions such as Trent and East Anglia. These broad variations are reflected in referral rates since it is clear that patients cannot be referred to hospital specialists and beds which do not exist. The finding by Cooper and Culyer (1976) that there is a relation between the level of provision of GPs and the provision of hospitals is more difficult to interpret in this light. As Butler (1980) showed there is no direct relationship between list-size and referral rates, so that relative deprivation in terms of GP numbers cannot be expected to have a systematic effect on hospital utilization, or the demand for hospital services produced by the referral process. We will argue in this paper that the relation between the supply of and demand for health services is a high order problem which cannot be properly understood with reference to aggregate, national or regional data. The level of resource allocation and the organization of these resources is of interest when comparing medical practice at a broad scale, but at the level of referral decision-making, which would normally utilize information from a DHA level (ie. the GPs direct knowledge of facilities), there are a wide range of self-regulating devices and practices which in most cases, psychiatric and geriatric care being notable exceptions, equalize the pressure on services

and also, apparently, the quality of care. The connection between general and specialized medicine is one which is manipulated by medical practitioners in a wide variety of ways to reach outcomes in health status or quality of care which seem to be indistinguishable. That at least appears to be the position at a stage where no easily applicable measurement of the outcome of health care is available.

We will attempt here to investigate the range of ways in which those patients deemed to be acutely ill might be coped with by the NHS, and our main object of interest will be the referral decision. At this the finest level of analysis, the individual GP, less tangible variables such as training, expertise, age, personality and assiduity will prove to be major determinants of action.

Most of these aspects of the referral process have been discussed at some point in the extensive literature, although never together, in a broad-based synthesis, and the referral system, despite its importance is still very poorly understood. Because there is still a great uncertainty about how the system works, especially at a meso- and micro-scale, many studies have concentrated on trying to offer suggestions for one-dimensional causalities in the system, usually by testing very limited data. A typical conclusion to this type of work is found at the end of a paper by Morrell et al (1971)

"A hypothesis that the observed differences [in referral rates] were due to the different age, sex, social class and diagnostic characteristics of the patients seen by [3] doctors was not substantiated. The possibility that a doctor's age and the duration of this experience in a particular practice is inversely related to his referral rate to hospital is examined [but equally not substantiated].

The attempt at empirical research, mostly by non-social scientists without the necessary analytical tools to tackle the data in any case, would therefore appear to have failed to produce very much more than some normative models. Green's (1973) typification of a high pathology department user is one such example.

Studies which take the referral process as their subject (cf. Berkeley, 1976; Evans and McBride, 1968; Forsyth and Logan, 1960; Fraser, Patterson and Peacock, 1974; Fry, 1971, 1977; Hopkins, 1956; The Journal of the Royal College of General Practitioners, 1978; Loudon, 1977; Martini et al

1973; Morrell et al, 1971; Price, 1973 and Stracey, 1961) are normally based on highly conditional, partial evidence, and produce generalizations which posit the GP as being ultimately the source of all aspects of the referral process. We suggest that this is an unhelpful perspective which moreover produces information that can be of little use in planning.

What little work has been done on the theory of the referral process has also been drawn to making generalized statements from which little re-usable meaning can be extracted. Re: Spencer, 1971,

"the process of referrals hinges on the formal organization of the service modified by an informal medical and community network".

A psychological perspective is introduced by Shortell (1971) in his application of Homans' (1961) social exchange theory to referral relationships between physicians in the United States. This application, aside of its theoretical limitations, is only of marginal interest to the British context due to the different balance of clinical interests in a state system of health care provision. Nevertheless, this type of work, which goes beyond the actual transferral of patients in referral, and which tries to understand the broader structure within which illness is perceived and health care is provided, has produced the basis onto which the present work is set.

No exposition of the referral system in the NHS has considered the implications of a changing structure of health care facilities for rates and patterns of referral. It is therefore unsurprising that the referral process is ignored in health service planning, notwithstanding its theoretically important ramifications on health delivery systems. This must partly be due to the lack of procedures in the planning system which make that inclusion possible. In addition, the paucity of data from the interface of general practice and the hospital service makes the construction of such procedures extremely difficult. Unclear, generalized statements of intuitive causal links in the system have been the inevitable result of seeing referral patterns as a mainly clinical problem, which can be resolved with reference to an objective qualitative description of (a) the patient and (b) the referring GP, with slight deference to the backcloth of available health care resources. We believe that a major contribution to this field

can be made by the placing of those clinical decisions within the context of the spatial, organizational and political constraints of the NHS infrastructure. Further, following a more general argument set out by Giddens (1979), this infrastructure of hospitals and clinics, and the structures of medical knowledge and care which operate through it are essentially involved in their own production. There is a circular, self-reinforcing relation between a culturally produced conception of health, and the institutions which exist to cure or ameliorate disruptions in health status. Medical knowledge and the need for medical care are, we find, extremely slippery concepts.

We begin with the premise that as most health care in the British health service is delivered at the DHA level, this spatial entity appears to be the most interesting and productive to investigate.

2. The Referral System in General Practice

2.1 Reasons for referral

Figure 1 shows the numerous possible routes taken by patients in receiving medical care. The most important flows in terms of this study are those between the GP and the consultant, and between the GP and the hospital directly. These are labelled a, b, c and d.

Referral to an out-patients department (a) is the largest component. It arises out of a need:

(i) for more specialized treatment after diagnosis has been made by the GP. This is particularly required of general surgery, ear, nose and throat (E.N.T.), orthopaedic, and physical medicine departments (Morrell et al, 1971). An appointment is made by the GP for the patient to see a consultant who takes over temporarily or permanently the management of a patient's illness but does not necessarily admit him or her to a bed. This may include cases of gallstones or hernias for surgical out-patient treatment or continuing care which lies outside the scope of general practice including serious conditions such as the leukemias, hypertension and diabetes. In terms of costs incurred by the hospital, once a patient, including day-cases, has occupied a bed he or she is an inpatient.

(ii) for a second, expert opinion from a consultant in cases where the GP is uncertain of a diagnosis. Enquiries of this type are most frequently made of medical, paediatric, and ophthalmology department (Morrell et al, 1971). The outcomes of these consultations in terms of resource utilization are equally uncertain.

(iii) for clinical or radiological investigations unavailable to the GP. Access to these facilities is variable, depending mainly on the size of the practice. Large health centres might have x-ray equipment while solo practices would be unlikely to justify such installations.

(iv) to satisfy a patient's demand for referral or dispel anxiety.

According to Hull (1972) this accounts for about 14% of all referrals, and is related to highly variable lay knowledge of medicine and health services.

In the year 1981/82, depending on the RHA, only 8.7 to 12.9% of hospital services expenditure was devoted to out-patient services (DHSS, 1983). Out-patient referrals assume a larger importance than this because an out-patient consultation may lead to an admission, either immediately or after a delay; (c) and (d) in Figure 1 above. Out-patient referral may therefore be turned, by the consultant, into in-patient referral. A certain amount of confusion may be avoided by employing this formulation. Although the final decision of admission is made by the consultant, the location of the admission will be highly constrained by the initial referral decision of the GP.

In the hospital, especially in surgical specialties, the distribution of medical care is, in most cases, regulated through a system of waiting lists (d)¹. The variation in the size of these waiting lists between different specialties, between different consultants in the same specialty, between different areas and regions, and through time is so wide however that no generalised statement is possible (Loudon and Stevens, 1980). Clearly though, while some conditions are fairly stable over long periods of time, medical intervention therefore not being urgent, others are more volatile in their severity and carry higher risks to health and of mortality. The waiting periods associated with waiting lists must therefore be in some direct relation to the anticipated time-span or time constraints of the group of illnesses which a given specialist attends to. This idea is further developed in Section 3.2. For our immediate purposes the individual GP's local knowledge and assessment of waiting lists is an important consideration since this, theoretically, will have a bearing on referral decisions; a desire on the part of the GP for good quality care as quickly as

1 95.6% of waiting patients in 1969 were for surgical specialties since medical cases are likely to show marked deleterious effects if admission is delayed (Culyer, 1976, p. 109).

possible being assumed.

Referral directly to an in-patients department (b) occurs when the GP decides that the patient has a condition which can no longer be safely treated at home. The actual distinction between out-patient and in-patient referral lies in the perceived urgency required in the provision of a hospital bed for the patient in the latter, but this too is contingent on the availability of resources (beds, equipment, staff) at any particular time. The GP, in deciding on an emergency admission, opts to bypass that other regulator of patient flows, the consultant, and lays claim, on the strength of a personal diagnosis, to more intensive, expensive hospital care. After this, responsibility for the patient's care is assumed by the consultant until such time as the patient is discharged and the problem of continuity of care is resolved. Action on this perceived requirement for a bed will usually be a function of the gravity of an illness, but this distinction merges with, and is complementary to, other possible categorizations of the motivations of referral. For instance, Berkeley (1976) broadly classifies the reasons for referral of both direct and indirect (via the consultant) admissions into medical, nursing and social reasons. The medical reasons were: gravity of illness, the need for investigation, or for special treatment; the nursing reasons were the need for intensive nursing, inadequate home conditions, or the lack of necessary equipment on the part of the GP; the social reasons were living alone, overcrowding, or pressure from relatives. All these reasons were in different instances, and in varying combinations, given by GPs as determinants of both in-patient and out-patient referrals. The way in which each of these factors is considered and measured against each other is central to the problematic of referral decisions. Armstrong (1983) suggests that there are basically two ways in which a medical problem may be considered. In the first, only the patient's organic (pain or other physical symptoms) presenting symptoms are considered, and other matters, particularly psycho-social factors are ignored. In the second, the problem is seen as multi-faceted and some details about the patient's biography and environment are incorporated into the doctor's assessment. Again these two

distinct methods may be seen as just two points along a continuum, and in the last instance they are probably related to the nature of the information which the patient gives the doctor, either spontaneously or under questioning. Any analysis of the content of a consultation would need to apply rigorous attention to the communicating skills of both the doctor and the patient

Given that we can tell apart the more stringently clinical view from a more social orientation Armstrong goes on to note that:

"A large proportion of patients in medical and surgical wards have some emotional problem. This might be a consequence of their organic problems; it may be because emotional and organic problems tend to 'cluster' in people's lives; or it might be because a patient with psycho-social problems is being unnecessarily investigated and treated, with all that entails in terms of costs, inconvenience and side-effects." (p. 95)

The possibility of such unnecessary treatment then requires the inclusion of psycho-social factors into a diagnosis. This too is fraught with difficulties. Can a psycho-social factor/symptom be distinguished from organic symptoms? What are the limits to relevant evidence? Can medical solutions be found for them? And lastly, is medicine appropriating, possibly harmfully, yet more of everyday life into its area of control, into the province of 'doctor's orders'? Illich (1975) argues that medicine's extending grasp on our lives leads to an increasing, and detrimental, dependency on medicine for solutions to human problems. Not only are those solutions not always available, but ill-health, iatrogenic disease, may actually be caused by treatment.

There are no stipulated rules by which referrals are made, although as we have seen the referral system is by no means simple. It depends wholly on the discretion of, at first, the GP and later the specialist. In the brief analysis of diagnosis we find that the assessment of symptoms may be a highly complex problematic, suggesting that there are only very broad limits to the way discretion may be exercised by GPs. This stems from the different preconceptions of a doctor's function and of the nature of ill-health which are held within the profession. While it is worth stressing that many diagnoses do turn out to have been correct, a large number do not, and it is our purpose here to do an analysis of the, as it were, marginal demand

for hospital services where these vague and underdefined categories come into play. Some mistakes in diagnosis are never discovered because the illness is self-limiting, that is, the body's own immune system can cope with it, and the ones that are discovered can easily be referred on to a specialist for a second opinion.

Having specified some of the problems inherent in making clinical judgements and decisions we must now assess some of the constraints placed on these decisions by the structure of health care resources in an area.

2.2 The decision to refer

There are, as Shortell and Anderson (1971) pointed out, two components to the referral decision.

(i) whether to refer (producing referral rates)
and (ii) where to refer to (producing referral patterns).
The first has been discussed above. The second may be further subdivided into two areas of interest; (a) the pattern of resource availability and (b) professional and personal relationships and interactions within the local medical fraternity.

(a) The spatial allocation of resources in the NHS is notoriously variable. Pre-1948 historical distortions, the crude planning procedures used since then, and the way in which planning decisions are made and implemented have all led to an incoherent and inequitable distribution of facilities. It is therefore difficult to generalize about spatial components in health care delivery at a national level, let alone within such a comparatively small, and comparatively new entity as the DHA. However, since the level of allocation is so important to our general argument it is necessary to give a general impression of the types of facility available in a DHA. This viewpoint will provide contingent limits to the crucial question of choice in the referral process. The surrounding theory will, in this way, be grounded in a suitable empirical bed.

Most GPs have a personal experience of one to three general hospitals (which may include a teaching hospital) and a small number of special hospitals (Loudon and Stevens, 1980). Only in London and a few major urban centres is this experience likely to be wider. It does not follow, however, that GPs would have a choice of up to three consultants or departments at each referral event. Although specialty services have been introduced to all parts of the UK since 1948, their distribution and planning remains distorted by localized anomalies in historical development. The variable size of specialty departments gives further complexity to the problem of choice in referral. In one DHA a large department may dominate a particular specialty, in another this specialty may be divided among two or three smaller departments. Specialty-mixes of hospitals may also have an effect on waiting lists. The allocation of GP beds¹ is exclusively based on local historical conditions as no national policy exists on their provision.

Bearing in mind these qualifications, DHAs may be characterized as having the following specialties commonly available:

general surgery, general medicine, gynaecology, geriatrics,
paediatrics, orthopaedics, ear nose and throat (ENT),
dermatology, special care babies, dental surgery, and
psychiatry.

Other specialties are provided at a regional or sub-regional level and tend to be concentrated in large urban centres. The problem of referral to these latter specialties will be addressed later in a review of referral interactions between the Dewsbury DHA and surrounding DHAs.

Our interest in spatial aspects does not come primarily from a concern with the disutility experienced by the patient in getting to hospital, although we appreciate that this might be relevant to (for instance) an assessment of ambulance use and the convenience of visitors, but rather that changes in the location of particular specialties, or changes in the size and staffing of particular specialties may have large-scale effects

1 Beds mainly for adult medical admissions who need hospital care, but not the special facilities available in large hospitals.

on patient flows, and the cost and quality of care.

For example the contraction of an ENT department will have an effect on the waiting lists of that department. A concomitant feedback to GP perceptions of that impedance will complicate or reorientate relevant referral decisions. Either the referral will remain towards the original location, or the referral will be to another (probably more distant) location, or the decision to refer might be suspended altogether especially if referral is not critical to the management of an illness.

Alternatively, the installation of a new department might have very complex and far-reaching effects on waiting lists and bed occupancy rates in the whole region by providing care for patients inside the DHA and also by attracting patients from other districts.

(b) Equally important in terms of our investigation must be the more elusive relations of respect and deference, like and dislike, and that medical competitiveness, often declared as desirable, mutual professional criticism. There are, quite obviously, wide and very localized variations in these qualities and processes because they are associated with individual actors, and although attempts have been made to systemize them (cf. Spencer, 1971; Hummell et al, 1970; and Shortell, 1979) we can make little practical use of these schemas.

Having seen that our problem lies in the interaction between a highly variable resource base and actors (or agents) with highly variable characteristics it is now necessary to organize our knowledge of these individual actors.

3. Factors Related to the Decision to Refer

3.1 Analysis

3.1.1 Introduction

We have emphasized so far that the understanding of referral patterns is only possible at a fine level of scrutiny and it is now necessary to turn to the DHA to identify salient, and, as far as possible measureable, interpretable variables which might define a model of the referral process. As is implied above, the bulk of specialty services are provided at this level and we will assume at this point that the unit is small enough to allow distance not to be seen as a significant resistance to referral trips¹. Each specialty department is potentially as 'attractive' to the referring GP as the next. This is a problematical assumption since it is unlikely that administrative regions are in all cases (especially at the margin) coterminous with the geography of care conceived of by the GP, even for those 'common' specialties listed above. Nevertheless, in terms of available data and in an attempt to keep referral choice as wide as possible it would appear to be a reasonable first approximation.

We may, following Martini et al (1973), usefully identify four groups of pertinent variables which should be addressed.

1. Factors about the community and resource availability
2. Factors about primary care
3. Factors about the patient
4. Factors about the hospital(s).

3.2 The community and resource availability

These are details which serve to characterize the region under investigation and may be further subdivided into:

¹ This may not be the case in larger mainly rural DHAs.

(a) Population characteristics such as age and sex structure and class or occupation structure both at the level of the district and the individual practice. For example, an elderly, working class population is deemed to require quite a different range of hospital care than one which is young and middle class. Similarly, demographic trends must be a primary consideration of any assessment of health needs (cf. Clarke and Spowage, 1983 and 1983; Clarke and Wilson, 1983 and Anderson, 1973).

(b) Background morbidity rates which will provide a sketch of the underlying trends in morbidity. This, as we shall argue later, is strongly related to the level of provision of services in a district.

(c) The location, type and size of medical facilities and services in the DHA will detail the spatial dimension of the referral process. Some assessment of community medical, nursing, social and welfare services is also necessary since these are the institutional measures which exist partly to limit referral to hospital. The Standing Medical Advisory Committee (1971) suggested that adequate and properly organized community services might accept up to 28% of patients with a questionable need of hospital care.

and (d) The bus and other transport services which might later be used to interpret constraints on movement.

3.3 Factors about primary care

These may be discussed under two headings: (a) the GPs social and professional background, and (b) practice characteristics.

(a) (i) The GP as individual

As we have seen the individuality of the GP has been stressed by previous work on the referral process. The failure of this particular perspective to provide a theoretical underpinning to the process leads us to a branch of social theory which has attempted to recombine social structure and human agency. This theory of structuration (Berger and Luckman, 1966; Giddens, 1976, 1979, 1981; Bhaskar, 1979; Bourdieu, 1977 and Thrift, 1983) claims that neither the structuralist-determinist nor the voluntarist

approaches are satisfactory to understanding social action. Rather, capable and knowing agents (actors in society) are able to reconstitute or transform the social structures which have produced them, and which they, through social practice, reproduce. The purpose of this section is to understand medical practice more fully at the level of human intentionality and motivation, that is, at the level where difference is introduced to social action. We are therefore sketching out a theory of practice or a theory of practical reason in referral. Such a project may be seen as an attempt to introduce Giddens' (1979) notion of 'practical consciousness'¹ to our analysis. While a GP is making an examination, or a diagnosis, or while thinking over possible referral, knowledge is being applied, as it were, instinctively; there is no consideration of, for instance, the origin or reliability of the knowledge which is being accessed. Reflexive consciousness, that is the consideration, rearrangement and transformation of medical knowledge and practice requires a different order of thought and is achieved over longer time periods. We are concerned then with deeper structures of motivation and intention, how they arise, and how they may mutate or be modified. What, in other words, is the basic structure of motive, knowledge and ethics in medicine, how does it become manifest in everyday medical practice, and lastly how can we investigate it?

Our problem is that medical practice is a very special and complex social practice which stands at the conjunction and occasionally conflict, of a large number of structures, each of them highly complex in themselves. Professional, specialized and lay medical knowledges and perceptions, career and professional structures within Medicine, medical technology and its development, the training and specialization of medics, the provision of a diverse range of medical resources according to shifting government policies,

¹ The "tacit knowledge that is skilfully applied in the enactment of courses of conduct, but which the actor is not able to formulate discursively". (Giddens, 1979, p. 57). And as such the set of "reason-givings" (Thrift, 1983) which determine an individual's behaviour.

and at a more local scale, the intensely complex interaction between a doctor and a patient, and the systems of power deference and dependence which may be assigned to that situation, must all in some way be incorporated into a comprehensive theory of medical practice. GPs in their long training no doubt come to resemble each other quite closely and might, following Bourdieu (1977, cited in Thrift, 1983), indeed be inculcated with a common 'habitus', that is, a common cognitive and motivating (reason-giving) structure. Once in general practice however the doctor commands a wide range of freedom of action which is borne out in the measured variations in referral rates. In Section 4 we will argue that this freedom is mobilized in referrals dominantly to produce ways of coping with a given range of resource allocation in a DHA. The range of freedom of action available to all GPs is apparent in the wide substitutability of treatments which exist for many illnesses. That this level of arbitrariness is not viewed with unease is a consequence partly of commercial interests in the drug industry, but also, and more importantly, the uncertain effectiveness of many treatments. These unknowns have been assumed into medical practice, although they are not uniformly distributed across medicine. Some areas have more uncertainties than others. Lastly, gaps in a GP's knowledge and expertise play a role in the variability of general practice.

In the end, critical interest must be focussed on the way in which the GP assesses the relative needs of patients, since it is the criteria by which "any conflict of interest between patients competing for scarce resources" (Cooper, 1975) is resolved which delineate medical practice. All medical practice carries with it an implicit rationing, but this rationing is only partly informed by medical knowledge.

Having attempted to deal with agency at a broad theoretical level we are left without a methodology with which to bring the perspective to bear on the real world. We must return to the basic, empirically measureable categories commonly employed in social theory and begin to relate what is above to reality.

An important time component in medical practice, the age of the GP or, perhaps more pertinently the number of years in general practice can be seen to have several effects. The problem it presents is centred on the time period in which significant medical innovation, whether technical, pharmacological or otherwise, can be predicted to take place. What might a twenty-six year old doctor know that a colleague thirty years a GP would not? (and indeed, vice-versa) and how have information flows in the medical community distributed what we presume to be progress in knowledge and practice? Information networks in medicine take many forms, varying from directives given by government to stop prescribing a drug to articles in an enormous periodical literature. But how thorough is the medical literature which is made easily available to GPs? How widely is this literature read? What resistance is there to the assimilation of new drugs and techniques? Direct contact with technical changes through, for example, hospital appointments would also serve to keep older GPs in touch with clinical and pharmacological innovations. But how many GPs hold such positions and in which departments? Expertise in particular specialties may also be gained by doing 'refresher' or postgraduate courses.

The age of the doctor, as Phillipson (1982) noted, is also related to the age of his or her patients. Doctors tend to age with their patients so if there are significant resistances to change in medical practice we might assume that in general innovations in geriatric care are less rapidly taken up than in, say, paediatrics. But again this is a multi-layered problem relating to local demographic changes and government policies of resource allocation once a new type of care - residential care for the mentally handicapped, for example, has gained professional acceptance.

An even more difficult problem relating to time is the retention of practical knowledge, and the ability of a doctor to exercise it within the routine of general practice. Uncommon medical conditions are difficult to recognise and act upon because their symptoms are infrequently presented and may be mistakenly subsumed under ordinary or commonly used categories. Early detection and treatment is in most acute illness beneficial (although we are here talking mainly about presymptomatic stages in disease), so delays caused by initial unsound treatment might prove important later in terms of hospital utilization and the quality of care. This whole problem seems to

be in the very nature of professional practice (Friedson, 1975).

Accepting that the application of practical knowledge is a complex and ambiguous area of general practice we must now investigate the substance of the consultation and in particular the doctor/patient relationship.

It is apparent that patients in the British NHS cannot exert the same control over their treatment as patients in private health care systems where the physician is directly dependent on the patient for his or her standing and prosperity. But despite an evident dominance in the consultation, the doctor must exercise extensive communicative and social skills to both aid the patient in describing symptoms and to give confidence in the therapy which will be prescribed. Like any face-to-face interaction the consultation is fraught with instability, and experience in communicating and dealing with patients is clearly a valuable asset (cf. Treadway, 1983). It should be remembered that even this element of medical practice is not uniform or static. From experience it is plain that GPs manage their relationships with patients in many different ways, although whether the patient returns for future consultations seems to depend more on the efficacy of previous treatments than on the way consultations were directed. Lay knowledge of medicine can incorporate many different styles of practice. Over time and over a number of consultations a firmer relationship may be established with the patient. Progressively, consultations may become more useful to both parties, as they, in the words of Bloor and Horobin (1975)'... socialize each other into a set of expectations of the other's role' (p. 279).

Associated with the communication of presenting symptoms by the patient is the crucial factor of a GPs concern over the certainty or uncertainty of a diagnosis. Uncertainty is characteristically caused by overlapping distributions of a symptom for healthy and sick persons (Cochrane, 1972) or by overlapping distributions of symptoms between different diseases. There is a level of uncertainty in all medical practice, but what is felt to be an unacceptable probability that other more serious pathologies are present, is, we presume very variable among doctors. As Armstrong (1983) notes:

"The patient's underlying question when presenting a symptom, whether it is organic or psychological, is: Am I normal doctor?" (p. 78).

The determination of what is normal is as problematic in medicine as in

other areas of social activity and it is not possible, statistically or otherwise, to state the bounds of normality. Doctors will assess deviance from a norm of well being in a wide range of ways. Since these norms are socially created, and not set by some isolated body known as the medical establishment, the whole wealth of reason-givings apparent in ordinary life are also in some way present in medical assessments. A response to illness then, is tied in with economic, political and historical as well as psychological forms and processes, and a significant medical decision, like the decision to refer a patient to hospital, is a manipulation by the doctor of all these extremely complex structures.

Where the GP does feel that there is an unacceptable uncertainty in diagnosis a referral for the second opinion of a specialist or for diagnostic tests may be made. Such referrals require an extra effort on the part of both the doctor and the patient; the doctor in making a telephone call or writing a referral letter and so subjecting his or her own professional authority to that of a specialist, and the patient in having to make the journey to hospital. Again we find resistances in medical practice and again complex problems arise in its evaluation. That these referrals are worthwhile is borne out by the findings of the Morbidity Statistics from General Practice 1971/72 (DHSS, 1979) that diagnoses were changed in 30% of cases following referral to hospital (see also, Elwood and Pitman, 1966 and de Dombal et al, 1972).

The actual content of an examination will inevitably have a bearing on eventual diagnosis. A more general, thorough examination may, for example, uncover a wider range of symptoms and allow a better diagnosis. But an examination includes both physical manipulations or investigations, and some sort of conversational interaction. The patient must lead the doctor to a conclusion. In this way it is not sufficient to produce, as Hull (1968) has done, a schematic categorization of procedures used in an examination and attempt to make broad deductions about the quality of medical practice. Each examination presents a different set of problems which may or may not be resolved.

A further problem arises out of the specialization of medical

practitioners. This narrowing of interests is inherent in the requirements of hospital-based medical schools, and in the way in which promotion is generally gained within the profession. The whole process leaves doctors with varying knowledge and skills and this may be expected to have a significant impact on the content of care they provide. Postgraduate experience or, more generally, a special interest in a given specialty has been found to produce higher rates of referral in that specialty (Evans and McBride, 1966). This apparent paradox is attributed to a greater awareness of alternative pathologies and treatments.

Although western medicine is overwhelmingly hospital-oriented, there are serious uncertainties about the values of hospital care in many of the instances in which it is prescribed and provided keeping a patient at home may be as efficacious for recovery as treating him or her in hospital. The general shortening of hospital stay-lengths in the NHS since 1948 is partly as a consequence of this realization. It is to be expected that attitudes among GPs as to the value of hospital care - both in general and in particular cases will mirror these various orientations. Spencer (1971) has labelled the extremes of attitude the 'whole person' orientation and the 'technical' orientation. The latter being more likely to depend on hospital care to provide solutions to ill-health. Mechanic (1970) had additionally the categories 'withdrawers' and 'moderns'. The latter here had both a 'whole person' and 'technical' orientation, the former neither. Behind these individual distinctions in the approach to medical practice lies a more general development of professional attitudes as a whole. There has been a steady increase in referral rates over the last 30 years (Fry, 1970; and Loudon, 1977) indicating an increasingly 'technical' orientation to care, although Mechanic's 'modern' category (i.e. currently taught practice) includes a significant 'whole person' component, thereby anticipating a move towards a deeper assessment of needs by physicians, and a broadening perception of their own function.

As a final comment we must acknowledge that it is extremely difficult for a social scientist to penetrate very deeply into the bewildering

complexity of medical practice. We cannot for instance elaborate on the substitutability of treatments or the current debate around generic drugs in any detail. The medical profession presents in its public face a very strict ethos of concerned, detached care of the sick under the aegis of a formidable technical and scientific knowledge. To stand outside the institution and theorize about the different styles of care of GPs as individuals is necessarily a limited approach. What is presented above is essentially a patient's view of medical practice.

(ii) The professional integration of the GP into the local medical community

Part-time attachment to a local hospital, or active membership of professional forums such as the local Medical (Advisory) Committee as well as informal socializing in the medical community may widen personal contacts and underscore professional relationships. This network of relationships may serve to provide information for the GP, but may also affect medical practice in more obvious ways. In the very different organizational context of the West German health service, Hummell et al (1970) found that private contacts among physicians can intensify referral relationships but only if they are combined with a recognition of the recipient physician's competence.

(b) Practice characteristics

(i) The GP as actor in the practice

The first major group of constraints on the GP that we are able to identify stems from his or her immediate working environment. Since the GP is dependent on a wide range of ancillary services (laboratories, nurses, examining equipment, an hygienic clinic and so on) which are not uniformly allocated, variations in medical practice must be at least partially related to the structure and level of provision of these services. Also there are constraints produced by the demand for the GPs services and for a large part this is dependent on the number of people who have registered themselves with the practice. Surveys by several workers show however that the rather coarse measure, list size, does not appear to have a systematic

effect on referral rates, although finer measures such as daily workload might offer a more direct relation (Butler, 1980). If we once more consider the mutability of medical practice this turns out to be as expected. Styles of practice are various and may be adapted to particular clinical and workload characteristics; the degree of freedom of action must always outreach variations in structural characteristics. Appointment systems, for instance, have been widely interpreted as being structural features imposed to lower workloads by encouraging a greater degree of prior-diagnosis by the patient. It is felt that the number of trivial conditions presented in general practice will as a consequence diminish.

A proper evaluation of workloads would seek to define a threshold at which the required thoroughness in medical practice was forfeited. The long debate over the best average list size, reported by Butler (1980) has tried to deal with this problem, albeit unsuccessfully. The amount of work which a doctor is expected or expects to do is connected with a far wider set of relations in society as a whole. And it is from these same relations that patient's expectations of health care are derived. Either we can assume that a GP's training is better than is required in most practice settings - even a rushed GP makes the right observations and diagnoses, or the uncertainty of medical practice and the monopoly the profession has over its administration serve as a buffer to deep investigation.

Butler implies that the 'structural aspects of the quality of care' can be shown to be more immediately to the content of care. By this he means the level of ancillary services. A larger practice with a concomitant, a greater resource of clinical skills, more nursing and other ancillary staff, and, in some cases, direct access on the premises to pathology and/or X-ray facilities will, by virtue of its size, have characteristics which tend to limit referrals to certain hospital departments. There appears to be no work on the relation of practice size and referral to for example pathology, general medicine and accident and emergency departments. Spencer (1971) calculated that ancillary nursing staff could save up to 28% of the GP's workload. Whether this time could be used in more thorough examination and consultation and hereby alter prescription and referral rates, or whether care, in a larger practice, is generally of a different type in crucial ways is a much more difficult problem.

(ii) Direct access to GP beds

Access to GP beds relates both to the historical development of health resource allocation and the integration of the GP into the local medical community. Certain regions such as Oxford and S.E. Thames are relatively well provided for with GP beds while others, like the Dewsbury DHA have no such facilities. Even where beds are available they are scarce compared with the number available in general hospitals and there are no guidelines for their utilization or additional training for those GPs who do have access to them. The capability to assume a wider control of the patient's treatment by having recourse to a GP bed will usually lead to a decrease in referral to some specialties. Clarke and Mulholland (1973) found that diagnoses of pneumonia, cerebrovascular accident, influenza and bronchitis were most often attended to in GP beds in the Farnham area. General medicine and general surgery departments would therefore be most affected in terms of utilization.

3.4 Factors about the patient

While we have concentrated up to now on the delivery of health care, the patient, as primary object of these structures and contingencies, must also enter into our analysis. And again we must identify features which might prove to have implications to the decision to refer.

(a) The illness type and its severity, or following Parsons (1964), the extent to which spontaneous recovery can be expected on the one hand, and the patient's socio-economic background and circumstances on the other, provide key inputs to a GP's decision on treatment. As we have argued in Section 2.1 the categorization and assessment of an illness is a highly complex problem. While the doctor is usually seen in the role as a 'functionally specific' technician of disease, a patient's psychological and social needs are sometimes explicitly and probably always implicitly taken into account. Culyer (1976) in detailing a set of inquiries which may be considered in the course of a diagnosis, or the recommendation of treatment, demonstrates this point. The patient will be assessed by the application of some of the following queries: How fast will the condition deteriorate? What other complications could set in? How serious is the pain and mental suffering

being endured by the patient? What are the patient's familial responsibilities? What degree of hardship would be caused in the patient's household by failure to admit? Could the care the patient requires be provided equally well in the home, rather than in the hospital?

For each of these to be answered requires an enormous knowledge of the patient quite peripheral to clinical concerns. It may be built up over time as we have suggested, or through a thorough enquiry into a patient's medical history, but the main assessment of deviance or degree of deviance is probably carried out tacitly, by the playing of roles by the actors and by reference to more subtle ideological constructs.

These are extremely detailed speculations which appear to be difficult to follow up empirically. Perhaps the main conclusion to be drawn is that in analysing health service systems we should be more aware of the characteristics of the patients being treated, and consider them together with the other indicators we normally use to describe those systems, particularly models of utilization.

(b) Age, of the patient as of the GP, presents a multi-dimensional problem. First, and as a general statement for the whole of the UK, admission for chronically ill over-65s become progressively more difficult with age; psycho-geriatric admissions being what Loudon and Stevens (1980) called a 'recurrent nightmare', although the need, both for clinical and other reasons, is in general greater among the elderly and geriatrics, and community care for the elderly is often badly stretched. In instances of exceptional pressure on resources in poorly provided for authorities, or as a result of an unexpected pattern of emergency admissions, beds may become interchangeable between specialties. A general medicine bed may be occupied by a geriatric case and so on. There appears to be no documentation of this, perhaps widespread, adaptability of hospital facilities.

This highlights another problem in health care delivery, which is the matching of an ill person to the rules and expectation of medical knowledge and the strictures of medical resources (especially hospital resources). Several themes arise from this question, each suggesting an uncertainty attendant in this 'fitting' process.

First, although western medicine in general has an elaborate, seemingly exhaustive system of categorizing disease, there are significant differences in the emphasis and content of this system in different countries. While each of these systems rely on the others for new knowledge and technical innovation organizational differences persist. There is no unified canon of medical knowledge. Second, these systems change through time either due to the emergence of new branches, or the shift or coalescence of established branches of the profession. The monopoly which the profession holds over what is regarded as legitimate in medicine means that the assimilation and alteration of specialties is often a long drawn-out and complicated process. Third, planners may react to changes only relatively slowly, it may be therefore that medical facilities always lag behind the state of knowledge, in a profession which, as we described earlier, requires a wide range of ancillary services. We may qualify these observations by adding that they do not apply to those sections of medicine which have proved to be effective. The case in point here, geriatric care, is not on the whole one of these.

Access to hospital then, is variable with the age of the patient. On a practical level, more aged patients are less likely to own a means of transport, and may be unable to afford bus or taxi fares to distant hospital facilities. Young mothers, with their children, may find similar problems, especially if the bus routes necessitate several changes in the journey. In emergency cases ambulance services may be called upon to transport patients who depend on public transport but this incurs large costs to the health service. The trauma of a long bus journey for a sick patient may also be a consideration to referral.

These problems of accessibility to hospital services, which at a coarse level is related to car ownership, will, it is felt, be considered by most GPs and serve for certain age and income groups as effective deterrents to the prescription of hospital care by the doctor.

The factor age is also mediated by the sex of the patient, middle-aged women, and male over-65s showing the highest referral rates (cf. Morrell et al, 1971; and Hopkins, 1956), although why exactly this should be has not been further analysed.

3.5 Factors about the hospital

Having made a decision to refer the GP must make a choice (assuming still that one exists) as to which consultant/hospital to send the patient.

(a) Referral patterns probably exist to a large extent through habit which even major additions or alterations to the structure of health care provision may not dislodge. Initially then the medical reputation of a consultant and subsequently the GPs experience of this and other consultants will influence the continuance of habitual referral or modification to a more complex pattern. Waiting times for appointments, the promptness and the informativeness of communications, and feedback from patients are the main processes of this experience.

(b) The ease of admission to a bed in the case of an emergency depends mainly on the stock of empty beds, but also on the relationship between the hospital secretary or registrar (who allocates beds to emergency admissions) and the referring GP. This mediator effectively controls the GP's access to beds. Through experience, ease of admission to hospitals and their departments will be known to the GP.

Ease of admission to hospital is often stated in terms of waiting times or waiting lists in a specialty or to see a consultant. Such measures usually indicate the waiting periods for a consultation at the hospital as an out-patient. Waiting lists have, it would seem, more complex effects on referral patterns than is often indicated. Broadly, they appear to be self-regulating (Feldstein, 1967) since longer lists are likely to encourage GPs to treat more cases on an out-patient or domiciliary basis¹. It is the consultant who specifies the priorities by which lists are set, and by his (most usually) style of practice, determines the length of the waiting list,

¹ The Institute of Hospital Administrators (IHA)(1963) reported that many GPs did not send their geriatric or surgical patients to hospital if they knew there was no chance of admission for a long time (Culyer, 1976).

so controlling referral rates and patterns, especially in the surgical specialties. There is an increasingly strong connection between the private practice of consultant surgeons and the waiting lists they maintain in the state system. Uncomplicated surgery (patients who will normally be day-cases) is relatively easy to get in the private sector, where there are shorter waiting lists. Complex, and therefore expensive, surgery becomes the priority of the NHS. The benefits of paying BUPA subscriptions, and being able to 'queue jump' therefore appear to be increasing as the NHS is forced to cut back vital hospital services. Clearly these developments will affect referral rates and patterns, and the advice which GPs are able to give their patients as to how to get the treatment they have partly been responsible for prescribing.

GPs are regularly notified of waiting list lengths and times by circulars from their local DHA. In practice it is found that the priority systems of consultants vary in ways not deducible from the information in circulars. Trivial complaints may be dealt with very rapidly on occasion by consultants with long published lists. This type of experience will obviously strongly influence referral in the future. In general however consultants with shorter lists are likely to receive more referred patients. And the priorities by which lists are ordered will be some criterion of maximum benefit from treatment, or, more formally, decision rules for admission (Luckman et al, 1969; and Phoenix, 1977).

(c) The GPs assessment of the relative quality of nursing staff and facilities may override either of the considerations above. A modern hospital with a happy staff is expected to be more attractive than an older hospital.

(d) Some conception of the relative accessibility of all the specialized medical facilities in an area will, as we have seen, to a greater or lesser degree affect the referral decision. The doctor's awareness of bus routes and car-ownership patterns in the practice will probably be the main components of this consideration.

4. The Interpretation of Clinical Freedom and some Suggestions for a Theory of Referral Rates

The immense and necessarily incomplete effort of interpreting and modelling behaviour, detailed in the last section, is required due to the insistence of the medical profession for guaranteed clinical autonomy or freedom; the power to control the treatment of an individual patient in a way that they judge best (Thunhurst, 1982). As independent contractors GPs have a wide range of discretion over the care of patients, generating highly variable referral rates, and inconsistencies in all other quantifiable aspects of their work (Lees and Cooper, 1963). But, if there are no regulations which demarcate the division of health and sickness, this division is, as we will argue below, fundamentally described by the availability of health care resources. That is, the cultural norms of health (accepted by both doctors and patients) are to a chronic extent predicted on a society's ability to provide care. The limits of 'need' for health care therefore exist within a particular frame of reference produced by the level of knowledge and technology as well as the organization of health service provision. For example, surgical repair work is carried out twice as regularly in the US as in the UK, within the context of a very similar level and type of medical knowledge and against a background of falling life expectancy in the US (Illich, 1977). This difference in practice, intimately related to 'need', is clearly the result of organizational differences between the health services in the two countries. Indeed, interestingly enough, the different rates of surgery closely resemble the relative numbers of surgeons in the two countries (Armstrong, 1983). Further emphasizing the importance of some evaluation of professional career structures in the assessment of defined medical need.

The continuing growth of the health service and the welfare state at large has been, as Gough (1979) explains, partly the result of growing social needs. These needs are due both to increasingly dysfunctional social conditions for example, the middle aged redundant, the victims of urban development; and to an intensification of the utilization of welfare services, what has been called the 'medicalization of life'¹ (Illich, 1977).

1 Illich argues that peoples ability to tolerate pain has lessened,

At the level of medical practice it is also found that growths and contractions in utilization are possible over short time periods. For instance, where the number of beds in hospital is increased, less work is done on an out-patient basis and the hospitalization rates and lengths of stay increase. The process has been popularized in the observation that "the number of beds used are the number of beds available" (Forsyth and Logan, 1960). Nonetheless, in certain specialties (those for which hospital care is essential and effective) it is theoretically possible to saturate the 'real' (not just realized) need for services². As such, need, in a developing and changing health delivery system, is always in flux, there are with each event of resource allocation new needs being awakened and (partly)³ satisfied.

"Collectively the medical profession appears to reassess its conception of need in line with actual levels of provision" (Cooper, 1975, p. 22).

These expanding needs (measured in terms of the elasticity of demand for health care), which are, of course, to a certain extent site-specific, are expressed both by the patient and the GP and do not occur suddenly but

their capability to cope with life's trials has weakened and like an addict they gradually become wholly dependent on the doctor for more and more minor complaints (Armstrong, 1983).

- 2 Considering estimations by Torrence et al (1972) that 1/4 of admissions to a teaching hospital did not require the skill and facilities available there, and the even larger wastage of costly resources (cf. Forte, 1984) predicted by Forsyth and Logan (1960) it appears that this saturation has already occurred in certain specialties in some places.
- 3 The continuing growth in the output of welfare services may not result in a similar growth in the need-satisfactions of clients. Social services, as Gough (1979, p. 93) points out, may have to keep running in order simply to stand still

rather as a gradual, 'cultural' shift in expectations of and demand for health status and the treatment of sickness. At the centre of this complex field of expectations is located the GP who "is responsible for fashioning NHS care both on the basis of the patient's personal circumstances ... and within the resource limits determined for the NHS as a whole and locally" (Tolliday, 1978). Referral patterns 'grow into' the space of health care infrastructure. It is not possible to provide normative figures for referral rates because they are a function of the way groups of GPs cope with the facilities, given patient morbidity, to which they have access. Therefore, the mix and size of specialties (bed availability), the staffing levels and characteristics, and the waiting lists which spring from this organization, together produce an endogenous (to the DHA) expectation and process of health care, while being controlled and altered through both exogenous ("prevailing practice" and DHSS planning policies) and endogenous (needs expressed by the medical profession in the DHA affecting planning) factors.

Feldstein's (1967) investigation of the elasticities of demand for hospital beds by age, sex and diagnostic category provides empirical evidence for this discussion. The elasticities of demand for each type of patient show how the number of beds used for each diagnostic category responds to bed scarcity (and the provision of new beds) and how much of this elasticity is concentrated in each of the two components of bed use. Of these we are concerned mainly with the elasticity of admission. This is shown to be on average almost twice as high as the elasticity of the average duration of stay (see Figure 2). Each result is derived from a regression analysis of eleven figures of the demand for hospital care, for a given disease, by Regional Hospital Board. The range of the data has in this manner been collapsed and in a sense standardized. Nevertheless Feldstein believes that the straight-line used in his regression analyses would be pertinent to the whole range of possible levels of supply of hospital services.

"Unfortunately there is no evidence that as availability increases bed-day demand tends toward ... a limit; on the contrary, the data suggests that, *at least within the observed range of supply*, bed-day demand increases proportionately with availability" (p. 198).

He gives little credence to a logistic curve, which intuitively would provide a more elegant interpretation of the development of GP and patient response to health resource allocation. To quote him again

"there seems no reason to accept this exponential in preference to the simpler straight line".

We would suggest however that whether by directive, "prevailing practice", or simply by the eventual disutilities of health care (the immediate hospitalization of a person suffering a mild headache would in most cases be thought of as unnecessary) there would be a limit to demand for hospital care for most illnesses. In many specialties it would be much higher than the present levels of demand.

The demand for admission to hospital could then be graphically represented. The general form of this graph (Fig. 2h) pertains to all diseases or the specialties they are subsumed by. Different conditions will display graphs with different slopes, marking the different characteristics of response to a growing provision of services. The lower segment up to the point of inflexion relates to the build-up of hospital resources. Both manifest demand for the specialty is low and then what treatment is required may be provided in other medical departments, or referred out to departments in hospitals outside the region. There would appear to be a qualitative leap made at some point in the evolution of health service allocation, probably relating to the appointment of a consultant to that specialty. For the referring GP this department now becomes a real alternative to previous modes of referral. It is likely that most of Feldstein's data corresponds to this relatively straight segment of the graph, although the usefulness of his figures in localized (DHA) contexts is questionable. At the level of one new department in a DHA the responses to that change will relate to a whole range of factors "averaged out" of Feldstein's figures. After this, as we have suggested above, a limit will gradually be reached if the extension of services is continued. We might therefore imagine a whole series of curves for each disease which interconnect at nodes relating to the actual provision of beds to a specialty in a region (in this case a DHA). That is, the level of services in a given specialty will determine the demand for those services from a whole range of diseases and further that the level of resource allocation (and hence demand) may be seen as being at a different point in the development of services than another specialty, i.e. the two

specialties lie at different points on Graph 2a. The elasticity of demand is determined by the level of services, which is a response to the nature of the illness. Clearly the level of services is very roughly equalized by national policy on hospital building etc., but it must be the role of the planner to find inadequacies or holes in provision at a local level. For example, diseases which in terms of resource allocation remain in the first segment of the Graph (2b) above. We have attempted to place the variations of elasticity found by Feldstein into a theoretical framework. Table shows that demand for admission to hospital for acute appendicitis, nationally, is relatively inelastic. We might interpret this as being due to a high level of provision of services to provide this simple surgical repair work; diagnosis is routine and treatment has been proven to be highly successful¹. Hence it has been possible to place some value on the need for these services, and this need has been saturated (i.e. the provision of services for acute appendicitis would be at a point on Graph 2a near the limit). Acute upper respiratory infections, possibly because they are less serious, and because there are substitutable non-hospital treatments available, are less well provided for with hospital resources and hence display a far higher elasticity of demand. An inelasticity of demand corresponds with a limit in the growth of demand; a level of provision having been reached which in terms of the contemporary consensus in medical knowledge (including the contemporary nosology or scheme of classification of illness), can effectively cure all those patients suffering solely from a particular illness. Similarly high elasticities may be taken to correspond to lower levels of provision in which there are still large volumes of recognised illness unprovided for with hospital care. This formulation provides enormous problems for interpretation.

First, if we take the received view of medical knowledge, illness is a condition describable in scientific terms. Further, all illnesses are, or

¹ Although Lichtner and Pflanz (1971) found, in a study carried out in West Germany, that in one in four cases where appendectomies were carried out the appendix was found to be healthy.

will be, once suitable drugs and technologies have been developed, treatable within medicine. From this position we might then ascribe high elasticities of demand, as measured by Friedman, as the result of historically-induced variabilities in, say, bed-to-population ratios, and also to policy decisions made by planners about such ideas as 'social efficiency'. This is some complex relation of the cost of providing medical facilities to the value (in health status) which those facilities are anticipated to realize. It is entirely unmeasurable. It is clear that 'technological' hospital-based medicine is extremely expensive, and increasingly so. Implicit in its development must be the conception of rationing the service it might provide. High elasticities, here, become dissolved into the problematic of medical ethics.

These certitudes we find, however, are challengeable. Gabbay (1982) states that:

"... subjective cultural elements ... are the very stuff of which medical knowledge is made".

If we balance this with Camaroff's (1982) observation that "the record of bio-medicine in relation to the burden of modern morbidity is poor" and that this is due, at least in part, to its stripping of disease aetiologies from their social and environmental contexts, then we begin to see that there may be a conflict between current nosology (and medical knowledge in general) and the services arranged against it. The suitability of hospital care for certain cardiovascular disease, cancer, musculo-skeletal disease and others is then called into question. Indeed, Culyer (1976) found that "some costly treatments in hospital may actually harm patients" (p. 54)¹. From this perspective it would appear that the evolution of demand as we have mapped it in Graphs 2a and 2b would not be resolved in some major areas of medicine. Parallel to the dichotomies of cost and availability then, run the problem of the effectiveness of modern hospital-based medicine. However, "difficult questions surround both the choice of appropriate

1 Illich's (1977) argument is that much high technology medicine, especially in terminal illness, is orientated towards consumption like most other products in capitalist society, rather than towards health.

measures of the outcome of care and the links between structure, process and outcome of care" (Butler, 1980, p. 120) and no sensitive methodologies appear to exist in which these questions are satisfactorily answered.

Within the overlapping grey areas of medical knowledge and the outcome of care lies the fact of substitutability of treatments, and this has particular bearing on choice in referral decisions. The problem of variation in medical practice now becomes the rationalization of these unknowns by the individual GP.

5. Review of On-going Empirical Work

Having acknowledged that there are major theoretical and methodological problems attached to understanding how the referral process works in different contexts with different actors, it remains to discover how far each of these segments of analysis might relate to the real world. To this end a survey was carried out of the referral behaviour of 26 GPs in the Dewsbury DHA.

Our interest is especially in the provision for the Ear, Nose and Throat (ENT) specialty in Dewsbury. Eight beds have been allocated as in patient facilities for ENT but these are occupied by patients from other specialties. All ENT inpatient referrals, and a number of out-patient referrals must therefore be made outside the Authority to Leeds (West or East), Bradford, Huddersfield or Wakefield. The problem of interpreting referral rates is for this reason extremely complex and interesting. It is felt that GPs referring in this way are highly self-conscious about their decision so that data collected is more likely to yield to interpretation.

It is envisaged that the data collected in the survey (see sample) will be used to produce a map of referral patterns and then, using information both from the questionnaire and from other data sources, to suggest how these patterns might change given various alterations in the system, particularly given the appointment of a full-time ENT consultant at the major hospital in Dewsbury.

6. Conclusion

This paper is based on the premise that a more comprehensive view of the referral process, incorporating the structure of health services in an area, is necessary to its understanding.

Any study of health services must recognize that medical practice varies widely. There is no validity in the claim to normative medicine, or to the portrayal of the physician as a disinterested, objective technician. The problem of medical knowledge is extremely complex and nonstatic; the perceived need for hospital care, given its effectiveness is intimately related to actual levels of provision, although changes in 'prevailing practice', technology and pharmacology also have major effects.

The doctor may be seen as a manager of illness, acting within wide and under-defined professional limits, and the referral process as an important mechanism of adjustment of care within the structural constraints in which the doctor is set. It is therefore unsurprising that referral rates vary enormously.

We have suggested, following Feldstein (1967), that one useful measure of the state of medical care provision and the referral system which functions through it in any place is the elasticity of demand for admission through time. The major problem with this approach is that very little data exists with which to construct such measures. Further, it would need to be based on data collected over a long time period to measure the influence of infrastructural constraints on the referral process in a given area. Health service infrastructure changes form only very slowly and incrementally. There is we believe no way to overcome these difficulties by surveying techniques. A question like "Would you refer more patients if ... ?" is unlikely to produce valuable answers.

We have also seen that the whole problem of medical knowledge is called into question in this research area, especially the effectiveness of hospital care. There is only a limited, unsystematic literature on the outcomes of courses of treatment. If information like this was made available, for

example the effectiveness of hospital care as compared with out-patient or domiciliary care given the illness and its severity, then some of the ambiguities present in referral decisions would be eliminated. With information about elasticities of demand, differences in the quality of health care as related to levels of provision could then be estimated. Presently a range of levels of provision can, apparently, through the mediating control of the GP, provide equally satisfactory levels of care. Waiting list lengths, which have on occasion been used as a proxy for quality of care measures, are only partial indicators of demand/'need' with differential affects on different illness and age groups.

There remain also, a host of problems concerning the role of Community Health Services, preventative health and environmental health measures in diverting patients away from hospitals. These have not been covered in this paper.

In our second paper we present results from our survey work and describe how these can be incorporated into our broader, systematic modelling efforts.

BIBLIOGRAPHY

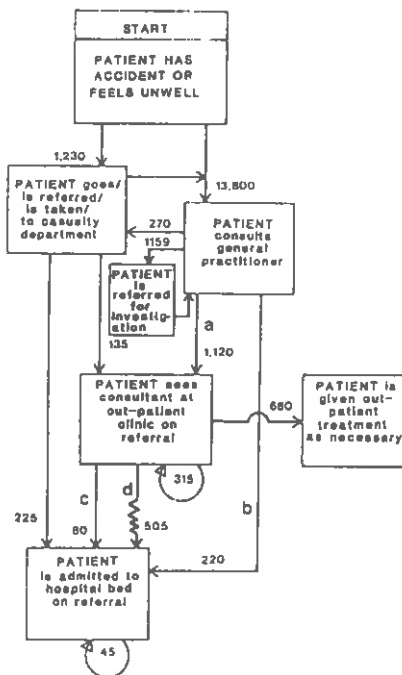
- Anderson, J.G. (1973) 'Demographic factors affecting health service utilization: a causal model', Med. Care 81(2), 104-20.
- Armstrong, D. (1983) An outline of sociology as applied to medicine, Wright PSG, Bristol.
- Berger, P. and Luckmann, T. (1966) The social construction of reality: a treatise on the sociology of knowledge, Penguin, Harmondsworth.
- Berkeley, J.S. (1976) 'Reasons for referral to hospital', JRCGP 26, 293-96.
- Bhaskar, R. (1979) The possibility of naturalism: a philosophical critique of the contemporary human sciences, Harvester Press, Sussex.
- Bice, T.W. and White, K.L. (1969) 'Factors related to the use of health services: an international comparative study', Med. Care VII, 2, 124-33.
- Bloor, M.J. and Horobin, G.W. (1975) 'Conflict and conflict resolution in doctor/patient relationships' in Cox, C. and Mead, A. (eds.) A sociology of medical practice, Collier-Macmillan, London.
- Bourdieu, P. (1977) Outline of a theory of practice, CUP, Cambridge.
- Butler, J.R. (1980) How many patients? Occasional Papers on Social Administration 64, Bedford Sq. Press.
- Clarke, M. and Mulholland, A. (1973) 'The use of GP beds', JRCGP 23, 273-79.
- Clarke, M. and Spowage, M. (1983) 'A District Health Authority simulation model: the in-patient sub-model, specification and first results', Working Paper, School of Geography, University of Leeds.
- Clarke, M. and Spowage, M. (1982) 'The specification of a micro-simulation model for a District Health Authority', Working Paper 338, School of Geography, University of Leeds.
- Clarke, M. and Wilson, A.G. (1983) Modelling for health service planning: an outline and an example, Working Paper 360, School of Geography, University of Leeds.
- Cochrane, A.L. (1972) Effectiveness and efficiency: random reflections on the health service, Nuffield Provincial Hospitals Trust, London.
- Comaroff, J. (1982) 'Medicine: symbol and ideology' in Wright, P. and Treacher, A. (eds.) The problem of medical knowledge, University of Edinburgh.
- Cooper, M.H. (1975) Rationing health care, Croom-Helm, London.

- Cooper, M.H. and Culyer, A.J. (1973) (eds.) Health economics: selected readings, Penguin modern econ. readings, Harmondsworth.
- Cullis, J.G. and West, P.A. (19) The economics of health, Martin Robertson, Oxford.
- Culyer, A.J. (1976) Need and the National Health Service, Martin Robertson, London.
- de Dombal, F.T. et al (1972) 'Computer-aided diagnosis of acute abdominal pain', BMJ Vol. 2, 1 April.
- D.H.S.S. (1983) Health service costing returns, HMSO, London.
- Drury, P. (1983) 'Some spatial aspects of health service developments: the British experience', Progress in Human Geography, 7, 60-77.
- Elwood, P.C. and Pitman, R.G. (1966) 'Observer error in radiological diagnosis of Paterson-Kelly webs', Brit. J. of Radiology, Vol. 39,
- Evans, E.D. and McBride, K. (1968) 'Hospital usage by a group practice', J. Royal Coll. Gen. Prac., 16, 294-306.
- Feldstein, M.S. (1967) Economic analysis for health service efficiency, North-Holland, Amsterdam.
- Fordyce, A.J.W. and Phillips, R. (1970) 'Waiting list management by computer', The Hospital, Vol. 66.
- Forsyth, G. and Logan, R.F.L. (1960) 'The demand for medical care: a study of the case-load in the Barrow', OUP for NPHT, London. Furness Group of Hospitals.
- Forte, P. and Wilson, A.G. (1983) Models of hospital costs, Working Paper, School of Geography, University of Leeds.
- Fraser, Patterson and Peacock (1979) 'Referrals to hospitals in an East Midlands city - a medical audit', J. Royal Coll. Gen. Prac., 24, 304-19.
- Friedson, E. (1975) 'Dilemmas in the doctor/patient relationship' in Cox, C. and Mead, A. (eds.) A sociology of medical practice (op.cit.)
- Fry, J. (ed.) (1977) Trends in general practice 1977, BMS for RCGP, London.
- Gabbay, J. (1982) 'Asthma attacked? Tactics for the reconstruction of a disease concept' in Wright, P. and Treacher, A. (ed.) op.cit.
- Giddens, A. (1976) New rules of sociological method, Hutchinson, London.
- Giddens, A. (1977) Studies in social and political theory, Macmillan, London.
- Giddens, A. (1979) Central problems in social theory: action, structure and contradiction in social analysis, Macmillan, London.

- Gill, D.G. (1975) 'The British National Health Service: professional determinants of administrative structure' in Cox, C. and Mead, A. (eds.) A sociology of medical practice (op. cit.)
- Gough, I. (1979) The political economy of the welfare state, Macmillan.
- Green, R.H. (1973) 'General practitioners and open-access pathology services', JRCGP 23, 316-25.
- Homas, G.C. (1957-8) Social behaviour as exchange, Am. J. of Soc., 63, 597-606.
- Honingsbaum, F. (1967) General practitioner work in British hospitals, Soc. & Econ. Admin., April 1967, p. 35.
- Honingsbaum, F. (1979) The division in British medicine, Logan Page, London.
- Hopkins, F. (1956) 'Referrals to general practice', BMS(ii), 873-77.
- Hull (1972) 'Diagnostic pathways in general practice', J. Royal Coll. Gen. Prac., 22, 241-58.
- Hummell, H.J., Kaupen-Maas, H., Kaupen, N. (1970) 'The referring of patients as a component of the medical interaction system', Soc. Sci. Med., 3, 597-607.
- Illich, I. (1977) The limits to medicine, Kogan Page, London.
- Jaques, E. (1978) Health services, their nature and organization, and the role of patients, doctors, nurses and the complementary professions, Heineman, London.
- J. Royal Coll. Gen. Prac. (1978) 'Practice activity analysis 5, referrals to specialists', J. Roy. Coll. Gen. Prac., 28, 251-3.
- Knox, P.L. (1978) 'The intra-urban ecology of primary medical care: patterns of accessibility and their policy implications', Env. & Plan., A, 10, 415-37.
- Last, J.M. (1963) 'The iceberg: completing the picture in general practice', Lancet, 2, 28-31.
- Last, J.M. (1971) 'Measurement of quality of care and medical audits' in Sackett, D.L. and Baskind, M.S. (eds.) Methods of health care evaluation, Hamilton, Ontario, MacMaster University (mimeo).
- Lees, D.S. and Cooper, M.H. (1963) 'The work of the general practitioner', JCGP, 6, 4, 408.
- Lichtner, S. and Pflanz, M. (1971) 'Appendectomy in the Federal Republic of Germany: epidemiology and medical care patterns', Med. Care, ix, 311.
- London, I.S.L. (1977) 'The general practitioner and the hospital' in Fry, J. (ed.) Trends in general practice, BMA for Roy. Coll. Gen. Prac., London.

- Lo don, I.S.L. and Stevens, R. (1980) 'Primary care and the hospital' in Fry, J. (ed.) Primary care, Heinemann, London.
- Luckman, J., McKenzie, M. and Stringer, J.C. (1969) Management policies for large ward units, Instit. of Operational Research, Health Report No. 1.
- Martini, C.J.M. et al (1973) 'Interaction between GPs and hospitals: work in progress', Health Services Research Group, Dept. of Community Health, University of Nottingham.
- Mechanic, D. (1970) 'Practice orientations among general medical practitioners in England and Wales', Med. Care, Jan-Feb Vol. VIII, 1, 15.
- Morrell, R.L., Gage, U.G., Robinson, N.A. (1971) 'Referral to hospital by General Practitioners', JRCGP, 21, 77-85.
- Navarro, V. (1978) Class struggle, the state and medicine, Martin Robertson, London.
- Office of Population Censuses and Surveys (1979) Morbidity statistics from general practice 1971/72, HMSO, London.
- Parsons, T. (1964) Social structure and personality, Free Press of Glencoe, N.Y.
- Phillipson, L. (1982) Capitalism and the construction of old age, Macmillan, London.
- Phoenix, C.I. (1972) 'Waiting list management and admission scheduling', in Abrams, M. (ed.) Spectrum 1971: a conference on medical computing, Butterworth, London.
- Price, D.T. (1973) 'A survey of the use of hospitals by a provincial, suburban practice', Update, 7, 617-25.
- Shortell, S.M. and Anderson, O.W. (1971) 'The physician referral process: a theoretical perspective', Hlth. Ser. Res. 6., 39-48.
- Spencer, P. (1971) 'General practice and models of the referral process', Health Report 6, Tavistock Inst. of Human Relations.
- Stracey, C.J.H. (1961) 'A hospital outpatient referral survey: a study of the referral habits of a group of general practitioners', JCGP, 4, 214-22.
- Thrift, N.J. (1983) 'On the determination of socialization in space and time', Env. & Plan. D: Society & Space, Vol. 1, 23-57.
- Thunhurst, C. (1982) It makes you sick: the politics of the NHS, Pluto Press, London.
- Tolliday, H. (1978) 'Clinical autonomy' in Jaques, E. (ed.), The Health Services, Heinemann, London.
- Torrance, N. et. al. (1972) 'Acute admissions to medical beds', JRCGP 22, 211-19.
- Treadway, J. (1983) 'Patient satisfaction & the content of general practice consultations', JRCGP, Vol. 37, 327, 769-71.

Figure 4



THE REFERRAL PROCESS IN THE NATIONAL HEALTH SERVICE
(source: Spencer, 1971)

Numbers refer to estimated referrals in one year for typical
population of 10,000 in about 1960

--- patient put on waiting list
--- patient transferred to another hospital department

NOTE: In 1971/72 flow a = 920
flow b = 182 (source: Morbidity Statistics
from General Practice (1971/72),
DHSS, 1979)

Figure 2b.

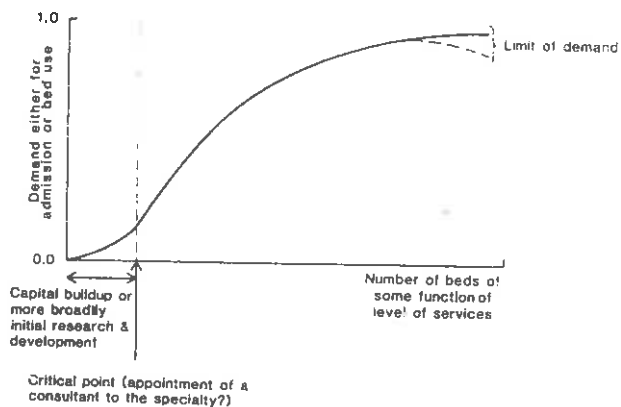
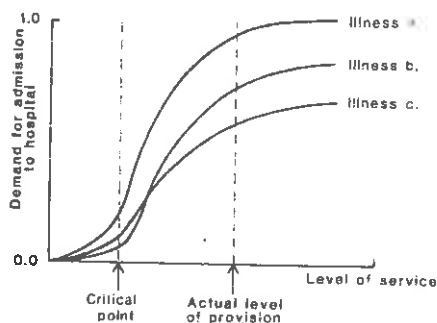


Figure 2a.



*Elasticities by diagnostic category**

Disease (1)	Beds used** (2)	Admissions** (3)	Mean stay (4)
Acute appendicitis	0.15(0.36)	-0.16(0.33)	0.31(0.17)
Acute upper respiratory infections	2.57(1.00)	1.53(0.52)	1.04(0.74)
Peptic ulcer	0.85(0.52)	0.29(0.40)	0.56(0.51)
Abdominal hernia (female)	1.39(0.44)	0.52(0.22)	0.87(0.44)
Haemorrhoids	1.14(0.62)	0.70(0.48)	0.44(0.24)
Tonsils and adenoids	0.55(0.46)	0.23(0.38)	0.33(0.38)
Arteriosclerotic heart disease	2.22(0.70)	1.14(0.51)	1.08(0.99)
Malignant neoplasms	0.58(0.30)	0.68(0.29)	-0.10(0.20)
Varicose veins (female)	1.40(0.70)	0.78(0.41)	0.62(0.67)
Males	1.03(0.14)	0.66(0.13)	0.37(0.15)
Females	0.97(0.11)	0.63(0.17)	0.34(0.21)
All Persons	-	0.65(0.15)	0.35(0.15)

* Elasticities calculated with respect to total number of beds used per 1 000 population.

** Per 1 000 population, 1960.

Figures in brackets are standard errors of responsiveness index values.

FIGURE 2

From M.S. Feldstein (1967) *Economic Analysis for Health Service Efficiency*, North-Holland, Amsterdam (pg. 219).

**Produced by
School of Geography
University of Leeds
Leeds LS2 9JT
From whom copies may be obtained**