THE UNIVERSITY of York

CENTRE FOR HEALTH ECONOMICS
YORK HEALTH ECONOMICS CONSORTIUM
NHS CENTRE FOR REVIEWS & DISSEMINATION

Modelling the Behaviour of General Practitioners

A Theoretical Foundation for Studies of Fundholding

Carlos Lerner
Karl Claxton

DISCUSSION PAPER 116

MODELLING THE BEHAVIOUR OF GENERAL PRACTITIONERS:

A Theoretical Foundation for Studies of Fundholding

Carlos Lerner

Karl Claxton

MARCH 1994

The Authors

Carlos Lerner is an MPhil student in health economics at the University of York. Karl Claxton is a Lecturer of Economics at the University of York.

Further Copies

Further copies of this document are available (at price £5.00, to cover the cost of publication, postage and packing) from:

The Publications Secretary
Centre for Health Economics
University of York
York YO1 5DD

Please make cheques payable to the University of York. Details of other papers can be obtained from the same address, or telephone York (0904) 433648 or 433666.

ABSTRACT

Despite its potentially profound repercussions, the general practitioner fundholding scheme has received relatively little attention from researchers. We provide here a theoretical foundation for empirical studies of fundholding. We begin by reviewing the incentives of the fundholding scheme. The responses of GPs to these incentives depend on the objectives of GPs, raising the questions: what are GPs trying to maximize? what GP objectives are useful in explaining their behaviour? After addressing these questions, we formulate several models of behaviour, focussing in turn on referrals, prescribing, workload, and the agency relationship between the GP and the patient. These models illustrate how we can provide a conceptual basis for researching various aspects of fundholding. Modelling GP behaviour yields several benefits: it induces us to explicitly identify our assumptions; it illuminates the interactions between incentives and GP objectives; and it can suggest interesting empirical questions. Studies based on such explicit analytical frameworks can elucidate how GPs respond to the incentives of the fundholding scheme.

1 INTRODUCTION

General practitioner fundholding constitutes one of the most important elements of the recent National Health Service reforms. The fundholding scheme, which allows eligible practices to take financial responsibility for providing a range of hospital services to their patients, fundamentally changes the incentives shaping GP behaviour. Because GPs perform a "gatekeeping" role in the NHS, controlling the entrance of patients into the health care system, these changes have repercussions beyond primary care. Yet, despite the potentially profound impact of fundholding, it has received scant attention from researchers. Knowledge about fundholding remains largely anecdotal or speculative. The few published studies lack any theoretical framework, limiting their value.

In this paper, we explore the development of models of GP behaviour, erecting a conceptual foundation for empirical studies of fundholding. By specifying a model of GP behaviour to guide our empirical analysis, we make explicit our assumptions about the incentives of fundholding and about the objectives of GPs, clarifying as well the interaction between these elements. Towards creating such models, we first review the structure of GP fundholding, examining the incentives that it engenders. Next, we explore which GP objectives may be relevant in explaining the responses of GPs to these incentives, drawing on the American experience with Health Maintenance Organizations and on the literature of supplier-induced demand. We then formulate several models of behaviour incorporating diverse assumptions about incentives and GP objectives. We conclude with a detailed example of how we can apply the models to design a study of fundholding. This example highlights the benefits of having an explicit theoretical framework. Figure 1 presents an outline of the paper.

None of the models presented here purports to explain all aspects of GP behaviour. Building on the simplest model, each subsequent model focuses on a particular behavioural dimension. We seek to illustrate how one can construct simple models of behaviour that can provide theoretical foundations for investigating a variety of empirical questions.

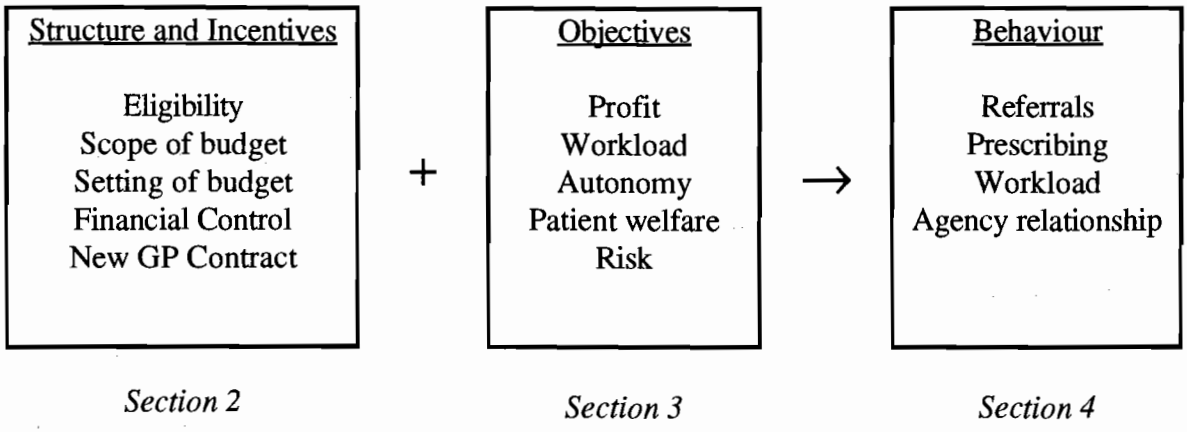

Figure 1: Outline of the paper

2 STRUCTURE AND INCENTIVES OF FUNDHOLDING

The Government instituted the GP fundholding scheme with ambitious goals. In the working paper delineating the fundholding proposal, it asserted that practice budgets offer GPs opportunities

- to improve the quality of services on offer to patients,
- to stimulate hospitals to be more responsive to the needs of GPs and their patients,
- to develop their own practices for the benefit of their patients,
- to enable practices that take part to play a more important role in the way in which NHS money is used to provide services for their patients. (DoH, 1989b)

Although the aims of fundholding were clear at the outset, the Government provided merely the broad outlines of its operation. The regions had to develop many of the working details, such as the eligibility criteria and the budget-setting process, during the scheme's implementation. The incentives of the fundholding scheme depend crucially on these working details, as well as on other recent changes in GP practice. After more than two years of operation, some aspects of fundholding still remain ambiguous, jeopardizing its effectiveness.

2.1 Eligibility

To be eligible for fundholding, the Government stipulated that practices should have at least 11,000 patients. The minimum size was subsequently lowered twice, first to 9,000 and finally to 7,000 for the 1993 cohort (the "third wave"). Besides a minimum list size, the Government requires that practices show the ability to manage budgets. Practices receive some financial assistance in developing these managerial capabilities. Fundholders receive 75 percent of the costs of improving their computing capacity to accommodate the information requirements of the scheme. During the preparatory year, moreover, practices receive an allowance covering the administrative costs of becoming fundholders.

Faced with this vague guidance regarding minimum practice size and management ability, the regions set their own standards for selecting practices. In one of the first studies of the fundholding scheme, Glennerster, *et al*, (1992) analysed the screening procedures in three regions for the first wave of fundholders. The regions applied judgmental criteria focussing on management arrangements, computing facilities, and support of all the partners in the practice. For the first wave, each of the regions accepted fewer than half the eligible practices expressing interest in the scheme.

Given this selection process, any comparison of fundholding and non-fundholding practices must recognize that--even after controlling for practice size and other measurable dimensions--the two groups are likely to vary in many relevant respects besides fundholding status. Those practices that choose to apply for fundholding, for instance, may be less risk averse and more prone to innovate. Bosanquet and Leese (1988) explore the links between practices' propensity to innovate and characteristics such as number of partners and location. Their results suggest that practices that apply for fundholding status may respond differently to financial incentives than other practices. Among the applicants, furthermore, regions select only the ones with greatest motivation, managerial ability and computing capacity. Without a longitudinal design, then, in some instances it may be impossible to determine whether differences between fundholders and non-fundholders are due to the scheme itself or to selection bias.

2.2 Scope of the Budget

Working Paper 3 (DoH, 1989b) delineated the services encompassed by the budget. The scope has been gradually expanded. In 1993/4, the budget includes:

- Hospital services. The budget covers in-patient and day-case surgery for most elective procedures, outpatient services, and diagnostic tests. It excludes medical and emergency admissions.
- Pharmaceuticals prescribed.
- Community care. As of April 1993, the budget has been extended to cover NHS community care services, including health visiting, district nursing services, mental health services and learning disability services.
- Practice costs. Fundholders receive as part of their budget the proportion of staff costs (presently 70 percent) and accommodation costs directly reimbursed under the current GP contract.

To limit the fundholder's financial risk, the District Health Authority becomes responsible for expenditures of more than £5,000 per year by any individual patient. After fundholding practices receive a global budget, they can choose freely how much to spend in each of the various categories, retaining any savings to improve their practices. Fundholders can contract with any hospitals or laboratories--whether public or private--for the provision of services within the scope of their budget.

2.3 Setting of the Budget

The mechanisms for setting fundholders' budgets play a large part in determining the incentives of the scheme. In the White Paper *Working for Patients* (DoH, 1989a), the Government declared that the hospital services element of the budget would be based on a weighed capitation basis, adjusting for "social and other local features that affect the use of hospital services." The Department of Health may have recognized the difficulties in developing a capitation formula, the lack of necessary information to implement it, or its potential for causing disruptive changes in service patterns. In any case, the subsequent working paper on practice budgets explained that while "it is the Government's intention to

move towards a weighed capitation approach to setting budgets...," initially "budget setting will need to have regard to the different expenditure components contributing to the total budget." (DoH, 1989b)

The present method of setting budgets--relying heavily on past expenditures--has led to wide variations in the amounts per patient that each practice receives for hospital services. The allocations for hospital services have varied from £30 to £104 per patient; it is not clear "how much these variations reflect real differences in need between practice populations, in clinical practice between GPs, in hospital costs, or simply in the difficulties of performing the calculations." (Audit Commission, 1993)

This approach to setting budgets generates some perverse incentives. It rewards high spending practices with higher budgets while it punishes low spenders. In the preparatory year, it may motivate practices to increase their referrals. To reduce the variations in practice budgets and to eliminate the undesirable incentives created by the current method of budget setting would require a shift to a weighed capitation formula. The allocation formula, however, needs to be carefully constructed so that it itself does not create incentives towards "cream skimming"--enrolling those groups of patients with expected expenditures below those predicted by the formula.

The drugs element of the budget is set in a manner similar to the hospital services component. For GP fundholders, the indicative prescribing budgets described in Working Paper 4 (DoH, 1989c) become the actual budgets. These budgets reflect the Net Ingredient Cost (basic list price) of dispensed prescriptions. They take into account the characteristics of the patients, the historical prescribing costs of the practice and the average costs of comparable practices in the same Family Health Services Authority. In 1993, regions have begun using the ASTRO (age, sex, temporary-resident orientated) prescribing unit to provide capitation-based benchmarks. (see Roberts and Harris, 1993) Regions will gradually increase their reliance on these benchmarks, giving less weight to historical costs.

2.4 Financial Control

Practices can exceed their yearly budget by 5 percent, leading to a corresponding reduction in their budget the following year. If a practice overspends by more than 5 percent, or persistently overspends by a lower percentage, it is subject to a "thorough audit, including a review by other doctors of any medical judgments which seem to be causing budgetary problems." (DoH, 1989b) Persistent problems may lead to revocation of fundholding status. Practices may choose to leave the scheme at any time, after giving sufficient notice to their region, a factor that ultimately reduces the risk assumed by fundholders.

2.5 The New GP Contract

A year before GP fundholding, the Government introduced changes in GP remuneration that can impinge on the incentives of the fundholding scheme. Under the new GP contract--which applies to all GPs, whether fundholding or nonfundholding--the proportion of income that GPs receive through capitation payments increased from 46 to 60 percent. Concurrently, the Government introduced measures to improve the ability of patients to choose GPs. These measures, outlined in the White Paper "Promoting Better Health" (DHSS, 1987), require Family Health Services Authorities to provide more comprehensive information about local GPs, encourage GPs to distribute leaflets describing their practice to prospective patients (a form of advertisement), and simplify the procedures for changing GPs. These changes strengthen the incentive for GP fundholders to attract more patients to their practice and purportedly increase the competition for patients.

2.6 Confusing Signals

Confusion regarding several aspects of the operation of the fundholding scheme weakens its financial incentives and potentially jeopardizes the achievement of the scheme's objectives. The scheme lacks clear, stable rules governing budget setting and budget savings. First-wave fundholders, for instance, did not know how their budgets for the following year would be set. They found themselves "in the strange position of being

unable to respond fully to the incentives of fundholding, feeling that an efficient running of their budget which resulted in savings, for example, would reduce their allocation their next time round." (Glennerster, *et al*, 1992)

Subsequent developments have fomented this uncertainty. In 1993, the Department of Health announced that the 1993/4 prescribing budgets would be set using 1992/3 as the historic baseline, rather than the practice's preparatory year as originally planned. ("New drug budgets...", *Fundholding*, 21 Jan 1993) Fundholders who reduced their expenditures on prescriptions in 1992/3 thus stood to receive lower budgets.

Fundholders have also received mixed signals about their ability to keep any savings they generate. Several regions pressured fundholders to return "fortuitous" savings. After complaints from fundholders, the Department of Health intervened, indicating that fundholders should be allowed to keep any savings to invest in patient care. Nevertheless, regions still attempt to regulate how fundholders use their savings.

The confusion regarding the "rules of the game" suggests that the financial incentives of the fundholding scheme may have a smaller effect than expected. For the first wave of fundholders, pressure to maintain a "steady state" may have further blunted the incentives. Fearful that fundholders would cause sudden shifts in service provision, disrupting the introduction of the NHS reforms, regions encouraged fundholders not to change their practice patterns drastically. (Glennerster, *et al*, 1992) The Oxford region, for example, introduced a rule under which "fundholders agreed to contract for 80 percent of their hospital service budget in the first year to go to the same hospitals as in the preparatory year, leaving them free to move the remaining 20 percent if they so wished." (Coulter and Bradlow, 1993) Any empirical analysis of the effects of fundholding must consider these factors, particularly if it uses data from the first waves.

The structure of the GP fundholding scheme interacts with the new GP contract and with broader factors governing GP practice to generate a set of incentives. These incentives are often not transparent--either to GPs or to the investigator--and they may occasionally encourage unintended responses. Predicting the effect of these incentives on

GP behaviour, moreover, entails assumptions--implicit or explicit--regarding the objectives of GPs.

3 THE OBJECTIVES OF GPs

In asserting that the incentives of the fundholding scheme would allow GPs to "develop their own practice *for the benefit of their patients*" (DoH, 1989b, emphasis added) the Government implicitly assumes a particular model of GP behaviour. The Government's claims may imply, for instance, that GPs respond to financial incentives while maintaining an overriding concern for patient welfare. The reactions of GPs to fundholding depend fundamentally on the objectives of GPs.

Empirical analyses of GP fundholding, however, lack any explicit assumptions regarding GP behaviour. The early studies of the fundholding scheme were purely descriptive, examining areas such as the variation in fundholding budgets (Day and Klein, 1991), the views of GPs about fundholding (Glynn, *et al*, 1992; Newton, *et al*, 1993), and the experiences of GPs and regions in instituting the scheme (Glennerster, *et al*, 1992). Descriptive studies may be useful in suggesting what research questions to ask, but, without any hypotheses, they cannot yield conclusive answers. These early studies give little indication of how the incentives of fundholding affect GP practice patterns.

Now, more than two years after the introduction of the scheme, a few analytical studies examining the effects of fundholding on GP referral and prescribing rates are emerging. (Coulter and Bradlow, 1993; Maxwell, *et al*, 1993; Bradlow and Coulter, 1993) These studies, however, lack theoretical foundations. They content themselves with exploring whether their results corroborate the conventional wisdom about the effects of fundholding. This approach suffers from several weaknesses:

- When the results contradict the conventional wisdom, the investigators have to resort to *ad hoc* explanations. Coulter and Bradlow (1993), for instance, attribute their unforeseen results to factors such as "the attempt to maintain a 'steady state' in the first year

of the reforms" and "the determination of these pioneering fundholders to ensure the success of the scheme". Although these justifications may be valid, they yield little insight about how GPs respond to incentives. A theoretical framework forces investigators to explicitly identify the assumptions behind their initial hypotheses. If the results subsequently contradict these hypotheses, we can begin thinking about which assumptions--about the incentives of fundholding or about the objectives of GPs--were inadequate.

- The lack of an explicit analytical structure can lead to confusion when interpreting empirical results. For example, Coulter and Bradlow observe that fundholders had lower referral rates than nonfundholders before the introduction of the scheme and similar rates afterwards. From these results they conclude that fundholding does not affect referral rates and that fundholders did not "artificially" increase their referrals in the preparatory year. In this argument, they inexplicably ignore the role of nonfundholders as a control group. A more explicit theoretical framework, establishing the implications of the various possible results *before* conducting the data analysis, may have precluded this flawed reasoning.

- Formulating a conceptual model can yield interesting hypotheses contradicting the conventional wisdom.

For these reasons, sound analysis of GP responses to the incentives of the fundholding scheme requires an appropriate theoretical model of GP behaviour. The first step in developing such a model is to examine the objectives of GPs. The next step is to establish which of these objectives we need to include in our models to explain GP behaviour.

3.1 What are GPs trying to maximize?

In this section, we consider which GP objectives are likely to influence GP behaviour. We focus on the ones that we find most relevant to the fundholding scheme: profit, workload, autonomy, patient welfare, and risk.

3.1.1 Profit

Although GPs do not profit directly from holding a budget, they can do so indirectly in several ways:

- They can increase their income from capitation by providing services that attract patients to their practices,
- They can hire additional auxiliary personnel to provide more fee-for-service procedures such as immunizations, health promotion clinics and family planning services,
- They can invest in their practice facilities, increasing their capital.

GPs seeking to increase their incomes must balance their response to these conflicting financial incentives. The ability to retain budget savings to expand their practices encourages them to minimize costs, while the prospect of increased capitation fees leads them to provide services that will attract patients. Advocates of the fundholding scheme emphasize the advantage of these trade-offs. In one of the earliest proposals of GP fundholding, Maynard, *et al*, (1986) assert: "Any profit is owned by the practice, which, as a consequence, has an incentive to economise. Excessive economy that may affect the quality of care will cause patients to go elsewhere... The benefits of such a system are that it is likely to be equitable, economical, sensitive to consumer needs, and meet professionally negotiated goals for health care." The Government proposals echo these arguments.

The increase in the proportion of GP income obtained through capitation and the recent changes allowing patients to switch GPs more easily strengthen the financial incentive to attract patients. Nonetheless, the capitation incentive may have limited power to counteract the pressures towards cost containment. Supporters of the fundholding scheme stress the advantages of allowing fundholders to act as informed purchasers of health care for the patient. The benefits of this agency relationship draw attention to the inability of patients to judge the appropriateness of health care services, an asymmetry of information widely recognized in the health economics literature. It may be wishful thinking, then, to declare that patients can compare GP practices in terms of quality of care, mitigating the fundholders' incentives for cost containment. The competition for patients may be restricted to those few aspects of the practice that patients can judge. Many

patients, moreover, attach high costs to severing links with their physician;[1] other GPs may need to offer markedly (and visibly) superior services to induce them to change practices.

We may gain an idea of the power of financial incentives to alter physician behaviour by examining the American experience with Health Maintenance Organizations. HMOs have a fixed budget from which they must provide a defined range of services to their enrolled members. They have been the subject of extensive analysis. A large body of evidence indicates that HMOs are cost-effective systems of care compared to traditional fee-for-service systems. (Weiner and Ferriss, 1990) HMOs provide higher levels of ambulatory care and preventive services but have 5 to 35 percent lower hospitalization rates. (Luft, 1981) Expenditures per person are 10 to 40 percent lower in HMOs than in fee-for-service plans, with no clear impact on the health status of the enrolled population. (Luft, 1981; Ware, *et al*, 1986; Manning, *et al*, 1987) These positive experiences with HMOs provided one of the principal inspirations for the fundholding scheme.

We must use care, however, in drawing lessons from studies of HMOs. HMOs employ a bewildering array of financial arrangements and incentive systems. (see Hillman, 1987) Weiner and Ferriss (1990) warn that we are still unclear about what role "risk arrangements such as capitation, withholds, and fund structure, play in HMO efficiency," and they point out that these financial incentives "have not always worked in the manner anticipated by their architects."

Hillman, *et al*, (1989) study the relationship between the presence of financial incentives and hospitalization rates in HMOs. They conclude that only the basic method of paying physicians--fee-for-service, salary, or capitation--has a significant impact on HMO hospitalization rates. More complex mechanisms, such as sharing the surplus in the referral fund with physicians, do not significantly affect these rates. Some of the complex incentive mechanisms analysed by Hillman, *et al*, resemble the incentive structure of the fundholding scheme, in which the physician profits indirectly from budget surpluses. Hillman's analysis,

---

[1]Phelps (1986) points out that in medical care "extended association between a physician and patient often can provide...economies in diagnosis, by allowing full comparison of current with past history of the patient, including mood, subtle appearances, or recall of facts left unrecorded in the most comprehensive written record."

then, suggests that the fundholding incentives may not have much impact on physician behaviour.

The experience of the SAFECO Life Insurance Company with the United Health Care Corporation leads to similar conclusions. United Health Care was an Individual Practice Association, in which independent practices contracted with the HMO to provide care for its members. This organization shared many features with the fundholding scheme.

> The innovative feature of the United Health Care plan was its requirement that primary-care physicians act as gatekeepers in controlling the costs and use of medical care in the rest of the system. When enrollees chose United Health care, they were required to choose one primary-care physician and to seek that physician's referral before going to specialists, emergency rooms, or hospitals... The primary-care physician was at limited financial risk for the enrollee's total health care cost. The loss or gain was limited to 10 percent of his reimbursed charges. At the end of the year the primary-care physician retained 50 percent of any surplus in his accounts (up to a maximum of 10 percent of his own fees). He was penalized up to 10 percent of his fees if there was a deficit... Catastrophic costs greater than $5,000 per patient were not paid from the physician account, but rather were borne entirely by the plan's reinsurance account. (Moore, *et al*, 1983)

Despite a promising start, United Health Care's hospital referral costs soon spiralled upward, making it unprofitable and leading to its termination in 1982. In their analysis of the SAFECO experience, Moore, *et al*, conclude that the financial incentives of the plan were insufficient to modify physician behaviour. The risk to the physicians' income was too small to induce changes in practice patterns. Moreover, consultants, who had control over expenditures after the primary-care physician made the referral, had no incentive to control costs.

Because the United Health Care plan differed in a few important ways from the fundholding scheme,[2] the lessons from the SAFECO experience are merely suggestive. Yet, the parallels between the United Health Care plan and GP fundholding, and the work of Hillman, *et al*, indicate that the financial incentives facing fundholders may also fail to have their intended effect. The American experience with HMOs, rather than support the fundholding scheme, cautions against taking for granted the power of financial incentives to

[2]Most significantly, only a small fraction of the patients of United Health Care physicians were enrolled in the HMO.

modify behaviour. Various considerations, moreover, suggest that the financial incentives of GP fundholders may be weaker than those facing HMOs:

- As discussed in section 2, confusion regarding the mechanisms for setting budgets, unclear policies about budget savings, and the Government's pressure to maintain a "steady state" blunt GP responses to the scheme's incentives.

- Studies of HMO performance generally compare HMOs with fee-for-service arrangements, which encourage cost escalation and over-provision of services. The scope for savings in the NHS is likely be narrower than in the US health care system. (Brazier, *et al*, 1990)

- For-profit HMOs can retain any savings as profits. Fundholders, on the other hand, can only profit indirectly from the scheme, since they are required to invest savings in their practice. (Healey and Yule, 1993)

Given the evidence suggesting that the financial incentives of the fundholding scheme may be weak, it becomes particularly important to consider other potential GP objectives besides profit.

3. 1.2 Workload

Most models of physician behaviour, mainly developed to investigate supplier-induced demand, recognize that physicians trade off leisure and income. Early surveys of fundholders identify GP concerns about the increased workloads associated with budget holding. (Glennerster, *et al*, 1992; Glynn, *et al*, 1992) Since fundholding can affect both the income and the workload of GPs, incorporating a measure of workload in our behavioural models may be necessary to explain the responses of GPs to the scheme.

3.1.3 Autonomy

The desire for greater professional autonomy constitutes one of the primary reasons GPs choose to become fundholders. In interviews with GPs from ten fundholding practices in the Northern region, the most common reason GPs gave for applying for fundholding status was the "opportunity to be independent", expressed "in terms of freedom to do

certain things (refer to hospitals of their choice; spend money on new staff and equipment) and freedom from externally determined policies and procedures (for example, contracts set by the district health authority)." (Newton, *et al*, 1993) Glennerster, *et al*, (1992) interviewed a cross-section of fifteen first-wave fundholders from three unidentified regions. The reasons GPs gave for applying to become fundholders included a desire to retain freedom of referrals after the institution of the NHS reforms, greater autonomy in changing staff, and the "opportunity to develop their practices in new ways and have more independence and control over their professional lives." An earlier pilot survey by the Kent Family Health Services Authority obtained similar responses. (Glynn, *et al*, 1992)

The responses of GPs to interviews do not necessarily reflect their true motivations. Nevertheless, it may be reasonable to assume that GP fundholders value their professional independence, and that they may be willing to trade off financial gain for greater autonomy. The preference for greater autonomy may manifest itself in organizational decisions such as contractual arrangements or the provision of services in-house. Thus, attempting to incorporate a measure of autonomy in a theoretical model of GP behaviour may prove useful.

3.1.4 Patient Welfare

GPs are unquestionably concerned with improving patient welfare, although they may choose to sacrifice patient welfare to some extent to increase their income. The relevant issue in our context, however, is not whether improving patient welfare is one of the objectives of GPs, but whether the inclusion of patient welfare in the physician's utility function is necessary to explain the responses of fundholders to the scheme's incentives. In analysing supplier-induced demand, several economists have found it useful to incorporate a patient welfare parameter in their models (under various guises, such as "ethical standards", "professional standards of performance", or "desire to cure the sick"). An analysis of fundholding may require a similar extension to the profit-maximization model.

3.1.5 Risk

Before the implementation of fundholding, some observers believed that the risks facing budget holders would be so high as to cripple the scheme. An American expert on HMOs admonished that the potential for adverse selection represented "the Achilles heel of the NHS reforms." (Scheffler, 1989) He argued that "the population in the UK proposal, 11,000, seems too small to absorb the impact of adverse selection within the budget targets. In the USA, HMOs with fewer than 50,000 enrollees have difficulty surviving and most are much larger." He failed to explain, however, why certain fundholding practices would attract disproportionate numbers of high users, a questionable assumption given the evidence that HMOs experience *favourable* selection. (Hellinger, 1987) Crump, *et al*, (1991) suggested that even without adverse selection, random variations in the need for hospital services would have a significant impact on practice budgets, particularly for smaller practices. Their simulation model predicted that for practices with a list size of 9,000 patients, the yearly costs of hospitalization would vary by 27.5%. They concluded that "given the random variation, it will be difficult for a practice to know whether lower spending in a given year is the result of prudent use or not; general practitioners may therefore be reticent about committing funds to improvement of services."

These omens have not become reality; most practices have been able to stay within their budgets. (Newton, *et al*, 1993) Measures such as the £5,000 stop-loss provision (ignored in the simulation by Crump, *et al*), the restriction of the scope of the budget to elective surgeries, and the use of historical costs in setting budgets (which compensates for adverse selection) mitigate the risk. Generous budgets from a Government intent on making the scheme work, moreover, may have further shielded the first waves of fundholders from risk.

Nevertheless, the risk involved in fundholding remains one of its central features. Fundholders respond to this risk in a variety of ways. Practices may be less willing to use their budget savings for practice changes that entail long-range financial commitments, such as hiring additional professionals. The GPs interviewed by Newton, *et al* (1993) explained that such arrangements "required a more reliable source of income than they could

currently envisage, and a clearer indication from the region (or the Department of Health) concerning policy about the retention and spending of savings." Selecting particular forms of contract may also reduce the risk for fundholders. Block contracts with hospitals or laboratories shift some risk to the providers, although the fundholder sacrifices some flexibility. Additionally, since the budget covers only elective procedures, fundholders can control expenditures by determining the timing of treatment.

Finally, fundholders can reduce their risk by pooling their budgets to form consortia. The Government encourages such an approach for practices with fewer than the required 7,000 patients on their lists. Consortia not only reduce the risk faced by individual fundholders, but they also give the fundholders greater purchasing power and potential economies of scale in administrative costs. These advantages, however, come at the expense of independence and flexibility. (Serra, 1992) Pooling budgets, moreover, may weaken the financial incentives of fundholders: as the size of the consortium increases, the behaviour of each GP will have a relatively smaller impact on the total budget. Risk may play an important role in explaining institutional responses to the fundholding scheme.

3.2 Which GP objectives are useful in explaining GP behaviour?

Traditional economic analysis begins with the assumption of profit maximization. In most cases, the assumption of profit maximizing behaviour represents a deliberate simplification of the complexities of the real world. The power of the profit maximization paradigm lies in its ability to generate empirically validated hypotheses from a few assumptions and from minimal amounts of institutional information. (Evans, 1976) Even if profit is not the sole argument in GP's utility function, then, we may not need to replace the profit maximization framework by more complex formulations. If profit maximization yields predictions that are confirmed by empirical tests, more complex models of physician behaviour may not be justified.

The assumption that physicians are profit maximizers has been frequently useful in generating such validated predictions. (Hornbrook and Berki, 1985) Profit maximization, however, may not be able to explain some prominent features of the health care market.

Explaining the evidence for supplier-induced demand (SID), for example, has led some health economists to abandon the profit maximization model. According to the SID hypothesis, physicians exploit the information asymmetry of the health care market by recommending treatment levels differing from what a fully informed consumer would choose. SID is not inconsistent with profit maximization. We would *expect* profit-maximizing physicians to increase their incomes by taking advantage of their information monopoly. However, some evidence--such as positive correlations between health care utilization rates per capita and physician/population density--suggests that physicians do not always choose to maximize the induced demand. To account for these observations, health economists have developed alternatives to the profit maximization model.

A common explanation for the SID evidence abandons the maximization framework altogether and postulates that physicians seek to achieve a target income, inducing demand to compensate for changes in the market that reduce their income. Analysts seldom attempt to explain the origin of this target. Yett, *et al*, (1983) suggest somewhat evasively that it is "based either on estimates of their peers' earnings or on subjective estimates of their own fair, reasonable or appropriate earnings capabilities." Despite its intuitive appeal, the target income hypothesis may be of limited use. It does not lend itself to precise formulation and it tends to generate indeterminate predictions.

Other alternatives preserve a maximization framework but expand the arguments in the physician's objective function. In justifying such expansion, Evans (1976) remarks that "there is no automatic presumption that the physician wearing his owner's hat will always impose his objectives over himself wearing his manager's or worker's hat; yet owner-domination is required by the profit maximizing model." He argues for inserting medical ethics "into the physician objective function, making his utility depended not only on his own income and leisure but also on his degree of conformity to some partially exogenous standard of professional performance and on his impact on (his perception of) the patient's best interests." In developing an operational model, he specifies the physician's utility "as a positive function of his income and a negative function of his workload and of the extent to which he exerts discretionary influence to increase demand." (Evans, 1974)

Other analysts follow a similar approach. In his model of physician behaviour, Zweifel (1981) incorporates ethical concerns as well as income and leisure objectives. Woodward and Warren-Boulton (1984) formulate a model presuming "that each physician derives additional utility both from positive consumption of the product of his leisure activities...and from providing additional medical care per patient...up to the 'appropriate' amount..." The medical care per patient argument in the physician's utility function seeks to represent physicians' "desire to cure the sick".

The above models, developed principally to examine the SID hypothesis, share a belief that the assumption of profit maximization cannot adequately explain some prominent features of the health care market. Before abandoning the profit maximization framework for more unwieldy models, however, we must consider carefully whether profit maximization suffices to explain physician behaviour. In expanding the objectives of behavioural models beyond profit maximization, we incur substantial costs. Introducing nebulous objectives such as "psychic costs" or "ethical concerns" opens the doors to *ad hoc* explanations, and translating such objectives into adequate operational definitions may not be feasible. Moreover, once we depart from profit maximization, we can no longer rely on the results of neoclassical economic theory.[3]

In modelling supplier-induced demand, Dranove (1985) shows that, even if physicians are profit maximizers, they may not maximize demand inducement in every setting. Natural limits to inducement can arise if excessive "aggressiveness" in recommending treatment shifts the demand curve facing physicians inward--by leading patients, for instance, to reduce the number of visits, to change physicians or to seek second opinions. Dranove's argument implies that the analyst need not appeal to concepts such as "patient welfare" to explain the evidence for SID.

In addition to patient welfare, profit maximization may accommodate the other GP objectives we identified previously. Clearly, a profit maximization model can incorporate risk aversion. We can also argue that autonomy does not represent a distinct objective--GPs may desire greater autonomy purely because it allows them to run their practices in the

[3]A limitation that entails psychic as well as analytical costs--see Reinhardt (1985).

most profitable way. Workload can be included within a profit maximization model by defining the implicit wages of GPs as practice costs, with the implicit wages increasing with increasing GP workloads.

We must avoid both the temptation to discard profit maximization without compelling reasons and the instinctual reaction of some economists to preserve the profit maximization framework at all costs. In some cases, the assumption of profit maximization may prove inappropriate. For instance, if the empirical evidence suggests that GPs are insensitive to the financial incentives of fundholding, explaining GP behaviour may require models including other objectives besides profit. Similarly, if expanding the arguments in the GP's utility function yields simpler, more useful models, we may want to undertake such an expansion.[4] We should be open to the possibility of abandoning the profit maximization framework in certain circumstances. Nevertheless, profit maximization represents a useful and proven starting point for our models.

4 MODELS OF BEHAVIOUR

As we argued before, modelling the behaviour of GPs yields several benefits for studies of fundholding: it induces us to explicitly identify our assumptions; it illuminates the interactions between incentives and GP objectives; and it can suggest interesting empirical questions. In this section we develop five models of GP behaviour. We begin with the simplest model, which focusses only on referral rates and assumes that GPs are profit maximizers who face an exogenous budget constraint. In the second model, we make the size of the budget depend on the behaviour of GPs. In the third, we distinguish between referrals and prescriptions. The fourth model incorporates GP workload. Finally, the fifth model adds a patient welfare argument to capture the potential conflict for GPs

[4]Evans (1976), for example, outlines the analytical obstacles of incorporating an implicit wage function in the definition of profits. Because of these difficulties, he recommends that the researcher adopt "a two variable function embodying a labour-leisure trade-off."

between being a perfect agent for the patient and considering the opportunity costs of the resources they commit. To avoid repetition, after each model we limit our discussion to the features that differ significantly from those of previous models. These models seek to illustrate how we can build conceptual frameworks for researching various aspects of fundholding. We conclude with a detailed example of how these models can be useful in designing an empirical study.

4.1 Profit maximization

First, we assume that GP fundholders behave as profit maximizers. They profit by attracting patients to their practices and by investing budget surpluses in their practice, increasing their capital. We do not distinguish between referrals to hospital and prescriptions; we combine both in the term "referral". GPs select the level of referrals that maximizes their profits. We posit that the number of referrals affects the number of patients on GP lists. This assumption is implicit in arguments suggesting that the incentive to attract patients checks incentives towards cost minimization. We express this behaviour symbolically as:

$$\max_{R,S} \pi$$

$$\pi = p_l L(R) + p_c C(S)$$

subject to the constraint

$$p_r R + S = B$$

where

π = profits
R = number of referrals
S = budget surplus
L = list size
p_l = profit to GP for each additional patient on his/her list
C = capital investment in the practice
p_c = profit to GP from each additional pound invested in the practice
p_r = average cost to the budget each additional referral
B = total budget

This is a static framework, in which the budget is exogenously determined--changes in list size, for example, do not affect the budget. The budget constraint, moreover, ignores the particular contractual arrangements of GPs with providers. We can justify this

simplification by pointing out that even if GPs have block contracts, not incurring the cost of additional referrals, future contracts will likely reflect their present referral levels. Thus, GPs eventually "pay" for increasing their referrals, even though they may avoid doing so temporarily.

We assume $L_{RR} < 0$ and C is an increasing linear function of S, so that $C_S > 0$ and $C_{SS} = 0$.[5] The assumption regarding list size implies that GPs can attract patients to their practice by increasing referrals, but that this ability to attract patients decreases as the referral levels increase (diminishing marginal ability to attract patients). For high enough referral levels, L_R can become negative--list size can eventually *decrease* with increasing referrals.[6] The assumption about capital investment holds that GPs turn a fixed fraction of their surpluses into practice capital. We also assume that $L_R(R/L) < 1$: list size is relatively inelastic to changes in referrals.

To obtain the first-order conditions for maximization, we construct the Lagrangean:

$$\ell = p_l L + p_c C + \lambda [B - p_r R - S]$$

yielding the conditions:

$$p_l L_R = \lambda p_r$$

$$p_c C_S = \lambda$$

The bold lettering denotes the optimal values of the variables. Solving for $\mathbf{L_R}$, the change in list size with an additional referral, *ceteris paribus*, we obtain the following expression:

$$\mathbf{L_R} = (p_c p_r / p_l)k$$

where $k = C_S > 0$.

A sufficient condition for the existence of a maximum is:

$$p_r^2 p_c C_{SS} + p_l L_{RR} < 0$$

---

[5]The notation A_B denotes the partial derivative of A with respect to B, that is, $A_B = \partial A/\partial B$.

[6]We avoid questions about the patients' ability to assess the adequacy of referrals. The model implicitly assumes that the perceptions of patients about optimal referrals are not biassed--*on average* patients do not over or underestimate the benefits of additional referrals for their health. The referral levels may need to vary greatly, however, before patients decide to change GPs.

The above assumptions that $C_{SS} = 0$ and $L_{RR} < 0$ are thus sufficient to guarantee a maximum.

For nonfundholders, the model simplifies to:

$$\max_{R} \pi$$

$$\pi = p_l L(R)$$

The condition for optimization is $\mathbf{L}_R = 0$.

From this model, we can derive several propositions:

Fundholding reduces the optimal referral rates.

Since $(\mathbf{L}_R)_{\text{fundholder}} > 0$ and $(\mathbf{L}_R)_{\text{nonfundholder}} = 0$,

$$(\mathbf{L}_R)_{\text{fundholder}} > (\mathbf{L}_R)_{\text{nonfundholder}}$$

Given the concavity of L, this inequality implies that

$$\mathbf{R}_{\text{fundholder}} < \mathbf{R}_{\text{nonfundholder}}$$

Because L is relatively inelastic to R,

$$(\mathbf{R/L})_{\text{fundholder}} < (\mathbf{R/L})_{\text{nonfundholder}}$$

where R/L = referral rate.

The intuition underlying this proposition is straightforward. If GPs are only concerned with maximizing profits, then nonfundholders will increase referrals up to the point where no more patients are attracted to their list. Fundholders, on the other hand, trade off the benefits of attracting patients with the benefits of investing budget surpluses in their practices. Thus, they refer less than nonfundholders. The difference in referrals between fundholders and nonfundholders broadens with increases in p_r, p_c and C_S and narrows with increases in p_l. These parameters determine the relative profitability of attracting more patients or increasing budget surpluses.

An increase in capitation payments (p_l) increases the number of referrals of fundholders but does not affect the referrals of nonfundholders .

Using the implicit function theorem[7], we find for fundholders

$$\partial \mathbf{R}/\partial p_l = - (p_c p_r k/p_l^2)/\mathbf{L}_{RR}$$

Since the nominator is negative and $\mathbf{L}_{RR} < 0$, $\partial \mathbf{R}/\partial p_l > 0$.

The intuition is again clear. For fundholders, higher capitation payments make attracting patients more profitable relative to investing budget surpluses. For nonfundholders, however, the capitation rate is irrelevant, since they always seek to maximize then number of patients on their lists.

Fundholders have an incentive to decrease the average costs of their referrals as much as possible.

A decrease in p_r leads to a lower optimal level of L_R for fundholders, implying higher **R** and higher profits. Since the model excludes quality of referrals[8], we must hold quality constant in this discussion. This proposition suggests, for example, that fundholders have the incentive to bargain with providers to obtain the lowest prices possible *for a given set of services*, or that fundholders will choose to substitute generic drugs for *clinically equivalent* but costlier brand-name drugs. Note that the cost of referrals does not enter the maximization problem of nonfundholders.

4.2 Profit maximization with an endogenous budget

The previous model assumed that fundholders faced an exogenous budget constraint. In reality, however, the budget of fundholders depends on their list size, so fundholders can increase their budgets for the following year by attracting more patients to

---

[7]The implicit function theorem states that if $F(x_1...x_n) \equiv 0$, then $\partial x_i/\partial x_j = - (\partial F/\partial x_j)/(\partial F/\partial x_i)$

[8]The "quality" of referrals may include elements such as shorter waiting lists, faster laboratory results, or more luxurious consultant accommodations.

their list. We model this mechanism by making the budget level dependent on list size. The model for fundholders in now given by:

$$\max_{R,S} \pi$$

$$\pi = p_l L(R) + p_c C(S)$$

subject to

$$p_r R + S = B(L)$$

From the first-order conditions, we obtain:

$$\mathbf{L}_R = p_r / [(p_l / p_c k) + \mathbf{B}_L]$$

For nonfundholders, the model remains the same, with $\mathbf{L}_R = 0$. We assume that $\mathbf{B}_L = K$: the budget increases by a fixed amount K with each additional patient on the practice list. By checking the second order conditions, we can show that our assumptions are sufficient to guarantee a maximum.

These results suggest the following proposition:

Fundholding reduces the optimal referral rates. This reduction, however, becomes less pronounced as the budget allotment for each patient increases.

From the first-order conditions, we observe that as $\mathbf{B}_L$ increases,

$$(\mathbf{L}_R)_{\text{fundholder}} \rightarrow (\mathbf{L}_R)_{\text{nonfundholder}} = 0$$

implying that $\mathbf{R}_{\text{fundholder}} \rightarrow \mathbf{R}_{\text{nonfundholder}}$.

As in the previous model, fundholders trade off some of the profits of increasing list size for the profits of investing budget surpluses, leading to lower optimal referral rates than nonfundholders. The incentive to attract patients, however, is stronger in this model. With higher list size, fundholders now profit directly from capitation payments and indirectly from higher budgets. Moreover, as the effect of higher list size on budgets becomes greater, the incentive of attracting patients becomes stronger relative to the incentive of retaining budget surpluses, making the optimal referral rates of fundholders approach the rates of nonfundholders.

With our assumption that each additional patient increases the budget by a fixed amount, we have assumed that past referral rates or surpluses do not affect the present

budget, except through their influence on list size. The validity of this assumption depends on the budget setting mechanisms adopted by each region, or, more precisely, on the GP perceptions of these budget setting mechanisms. Comparing the implications of different methods of budget setting (such as using historical costs or a capitation formula) may entail abandoning this static model for dynamic or temporary equilibrium frameworks.

4.3 Profit maximization, separating referrals and prescriptions

We expand the first model by including separate decision variables for referrals and prescribing. The model for fundholders becomes:

$$\max_{R,P,S} \pi$$

$$\pi = p_l L(R,P) + p_c C(S)$$

subject to

$$p_r R + p_p P + S = B$$

where P = number of prescriptions, p_p = average cost to the budget of a prescription, and the other variables are defined as before. Note that list size now depends on both referrals and prescriptions. To the assumptions made in the first model, we add $L_P(P/L) < 1$, $L_{PP} < 0$, $L_{RR} < L_{RP} < 0$, and $L_{PP} < L_{RP}$. The last two assumptions state that referrals and prescriptions are imperfect "substitutes". For example, they imply that the effect of increasing referrals on list size (L_R) diminishes with higher prescribing levels, although L_R is more sensitive to referral levels. We also assume that $p_r > p_p$: the average price of referrals is greater than the average price of prescriptions. The above assumptions are sufficient to guarantee a maximum.

From the first-order conditions for fundholders, we obtain:

$$L_R = (p_c p_r / p_l)k$$

$$L_P = (p_c p_p / p_l)k$$

where $k = C_S > 0$.

The model for nonfundholders is:[9]

$$\max_{R,P} \pi$$

$$\pi = p_l L(R,P)$$

[9]We ignore the incentives generated by *indicative* prescribing budgets for nonfundholders.

yielding the first-order conditions:

$$\mathbf{L}_R = 0$$

$$\mathbf{L}_P = 0$$

This model suggests the following proposition:

In some circumstances, if the difference in cost between referrals and prescriptions is sufficiently large and if there is a sufficiently wide scope for substitution, fundholding increases prescribing rates.

If there is no possibility of substituting prescribing for referrals (that is, if $L_{RP} = 0$), we can argue as before that

$$(\mathbf{L}_P)_{\text{fundholder}} > (\mathbf{L}_P)_{\text{nonfundholder}} \Rightarrow \mathbf{P}_{\text{fundholder}} < \mathbf{P}_{\text{nonfundholder}}$$

because in this case $\mathbf{L}_P$ depends only on **P**. However, where $L_{RP} \neq 0$, that is, where prescribing and referrals are substitutes, this argument is no longer valid, since $\mathbf{L}_P$ now depends on both **R** and **P**. Fundholding may *increase* prescribing rates if fundholders substitute cheaper prescriptions for expensive referrals because the profit from greater budget surpluses compensates them for decreases in list size.

To prove that fundholding does not necessarily decrease prescribing rates we construct a counterexample. Let

$$L = a_r R + a_p P - b_r R^2 - b_p P^2 - cRP$$

To meet our previous assumptions about L, we restrict the value of the parameters to $b_r > 0$, $b_p > 0$, $c > 0$, $b_r > c$, and $b_p > c$. Note that $L_{RP} = c$, so that c represents the degree of "substitution" between prescribing and referrals. We can solve analytically for the optimal prescribing and referral levels. We obtain:

$$\mathbf{P}_{\text{fundholder}} = (c^2 - 4b_p b_r)^{-1}[(a_r c - 2a_p b_r) + (p_c k/p_l)(2p_p b_r - p_r c)]$$

$$\mathbf{P}_{\text{nonfundholder}} = (c^2 - 4b_p b_r)^{-1}(a_r c - 2a_p b_r)$$

We are looking for situations in which fundholders have higher prescribing levels than nonfundholders, that is, where $\mathbf{P}_{\text{fundholder}} - \mathbf{P}_{\text{nonfundholder}} > 0$. This inequality will be satisfied when $p_r c > 2p_p b_r$. Consequently, if the price of referrals (p_r) is sufficiently higher

than the price of prescriptions (p_p) and if referrals and prescribing are sufficiently good substitutes (c sufficiently large), fundholders will prescribe more than nonfundholders.

This model contradicts the common assumption that fundholding will necessarily reduce prescribing rates. The possibility of substituting cheaper prescriptions for more expensive referrals creates an incentive for fundholders to increase prescribing.

4.4 Profit maximization and workload

We now expand the first model to incorporate GP workload. In the definition of profit, we take into account an implicit wage for GPs, by subtracting an implicit wage function from the gross profits and by making list size dependent on GP workload as well as referral rates. We begin with a general model, in which workload depends both on demand (list size) and on supply (the influence of GPs on their workload).

The model for fundholders becomes:

$$\max_{R,S,e} \pi$$

$$\pi = p_l L(R,e) + p_c C(S) - H(t)$$

$$t = eL$$

subject to

$$p_r R + S = B$$

where t = total time spent in patient care, e = effort = time/patient·year, H(t) = the cumulative implicit wage of the GP, and the other variables are defined as before. In this context, the variable e represents the aspects of workload that GPs can control--it is the product of average length of consultation and the number of consultations per year for each patient on the practice list.

We make the additional assumptions:

- $L_{tt} < 0$: The effect on list size of each additional minute of GP workload diminishes with higher levels of workload.
- $H_t > 0$, $H_{tt} > 0$: The cumulative implicit wage of GPs increases with increasing workload. Moreover, the marginal implicit wage increases with higher levels of workload. For example, the implicit wage for the first hour of work is lower than the implicit wage for the tenth hour of work.

- $L_{tR} < 0$, $L_{tR} > L_{RR}$, $L_{tR} > L_{tt}$: GP workload and referrals are imperfect "substitutes". For example, the effect of increasing workload on list size (L_t) diminishes with higher referral levels, although L_t is more sensitive to workload levels.

This model can accommodate both demand-led and supply-led views of GP workload. Unfortunately, its generality results in indeterminate predictions. To make the model useful, we must make further assumptions about GP workload. We illustrate by assuming that GPs have full control over their workload.

The model now becomes:

$$\max_{R,S,t} \pi$$

$$\pi = p_l L(R, t) + p_c C(S) - H(t)$$

subject to

$$p_r R + S = B$$

In contrast to the general model, GPs can now directly choose their level of workload t.

From the first-order conditions, we obtain:

$$\mathbf{L_R} = p_r p_c C_S / p_l$$

$$p_t \mathbf{L_t} = \mathbf{H_t}$$

We omit checks of the second-order conditions. Following the approach for previous models, we can show that our assumptions are sufficient to guarantee a maximum.

For nonfundholders, the model becomes:

$$\max_{R,t} \pi$$

$$\pi = p_l L(R,t) - H(t)$$

giving the first-order conditions:

$$p_l \mathbf{L_R} = 0$$

$$p_t \mathbf{L_t} = \mathbf{H_t}$$

These results suggest the following proposition:

Fundholding increases the optimal workload of GPs.

As fundholding decreases the optimal number of referrals, the marginal benefits of increasing workload ($p_t \mathbf{L_t}$) become greater than its marginal costs ($\mathbf{H_t}$), inducing GPs to

increase workload. The magnitude of the difference in workload between fundholders and nonfundholders depends on the magnitude of L_{Rt} (the extent of substitution between workload and referrals) and the difference in referrals between fundholders and nonfundholders.

Showing this result mathematically, we first define $F \equiv p_t\mathbf{L}_t - \mathbf{H}_t \equiv 0$. Applying the implicit function theorem yields

$$\partial \mathbf{t}/\partial \mathbf{R} = - (p_l\mathbf{L}_{tR})/(p_l\mathbf{L}_{tt} - \mathbf{H}_{tt}) < 0$$

Thus, $\mathbf{R}_{fundholder} < \mathbf{R}_{nonfundholder} \Rightarrow \mathbf{t}_{fundholder} > \mathbf{t}_{nonfundholder}$.

We can find various sources of support for the assumption that GPs have full control over their workload. In a review of GP workload estimates, Thomas, *et al*, (1989) argue that "the independent status of general practitioners in the UK and the present remuneration system...allow general practitioners considerable flexibility in how they structure the service they provide and the level of workload which they choose to undertake." Additionally, studies of the effects of list size on workload find low positive correlations between these variables, indicating that other factors besides list size (demand) may be important determinants of workload. (Wilkin and Metcalfe, 1984; Knight, 1987; Calnan and Butler, 1988; Groenewegen and Hutten, 1991) Nevertheless, we may wish to consider other models defining workload as demand-led, with list size and patient characteristics being its main determinants.[10] We may also wish to expand the definition of workload to encompass other aspects of GP work, such as practice administration and teaching.

---

[10] Butler and Calnan (1987), for example, implicitly assume that list size determines the allocation and use of GP time, even though they find that "the overall variances [of workload measures] explained by list size....were small." In a subsequent paper, they explicitly investigate the hypothesis that workload is demand led, concluding that "the demand led model provides only a partial explanation for variations in time allocation in general practice." (Calnan and Butler, 1988)

4.5 Profit maximization and patient welfare

Until now we have avoided explicitly incorporating a patient welfare argument in the GP's utility function. Patient welfare, however, did enter indirectly into our models. The effects of referrals, prescribing or workload on list size depend on the patients' perceptions of these services on their welfare. Thus, by seeking to maximize list size, GPs maximize patient welfare as perceived by the patients. In some contexts, however, GPs and patients may define patient welfare differently. We explore one such scenario in the following model, which captures the conflict for GPs between being a perfect agent for the patient and considering the opportunity costs of the resources they commit. It describes the utility of GPs as a function of their income and of their perception of patient welfare.

The model for fundholders is now given by:

$$\max_{R,S} U(\pi, W)$$

$$\pi = p_l L(R) + p_c C(S)$$

$$W = W(R)$$

subject to

$$p_r R + S = B$$

where U = utility, W = patient welfare as defined by GPs, and the other variables are defined as above.

To the assumptions of the previous model, we add $U_\pi > 0$, $U_{\pi\pi} < 0$, $U_W > 0$, $U_{WW} < 0$, and $W_{RR} < 0$. We allow $W_R < 0$ for large enough R. The first four assumptions hold that GP utility increases with both profits and patient welfare, but the rates of increase diminish with higher levels of each of these variables. The assumptions regarding W indicate that the impact of increasing referrals on patient welfare (as defined by GPs) diminishes for higher numbers of referrals, and that, after some point, further increases in referrals can decrease patient welfare.

We make the further assumption that W_R becomes negative at a lower level of referrals than L_R. When patients assess the optimal level of GP referrals, they do not consider the opportunity costs of the resources committed by the referrals. They value any referrals that have a positive impact on their health. We assume here that GPs include in their definition of patient welfare (W) *some* consideration for the cost of referrals to the

NHS.[11] We allow for the possibility that GPs give more weight to the benefits for the patient in front of them, as long as they attach nonzero weight to the forgone benefits for other NHS patients. In other words, we assume that $L_R > 0$ up to the point where MB = 0, and $W_R > 0$ up to some unspecified lower point where $0 < MB \leq MC$ (MB = marginal benefit to the patient of an additional referral, MC = marginal cost to the NHS of an additional referral). Consequently, after some point, additional referrals have a positive health benefit for the patient, increasing list size, but decrease patient welfare as defined by GPs because the resource costs outweigh the benefits. By differentiating between the definitions of patient welfare used by patients and GPs, we capture the potential conflict for GPs between being a perfect agent for the patient in front of them and considering the foregone benefits to other patients. We evaluate the implications of these assumptions about L and W in our discussion of the model's predictions.

Constructing the Lagrangean, we obtain:

$$\mathcal{L} = U(\pi,W) + \lambda [B - p_r R - S]$$

giving the first-order conditions,

$$\mathbf{U_\pi} p_l \mathbf{L_R} + \mathbf{U_W W_R} = \lambda p_r$$

$$\mathbf{U_\pi} p_c \mathbf{C_S} = \lambda$$

The above assumptions regarding U, C, and L are sufficient for a maximum.

Solving for $\mathbf{L_R}$, and letting $\mathbf{C_S} = k$ yields:

$$\mathbf{L_R} = (p_c p_r / p_l)k - (1/p_l)(\mathbf{U_W}/\mathbf{U_\pi})\mathbf{W_R}$$

For nonfundholders, the problem becomes:

$$\max_R U(\pi,W)$$

$$\pi = p_l L(R)$$

$$W = W(R)$$

giving the first-order condition,

$$\mathbf{L_R} = -(1/p_l)(\mathbf{U_W}/\mathbf{U_\pi})\mathbf{W_R}$$

From this model, we can derive the following propositions:

[11]If we assume that GPs attach zero weight to the foregone benefits to other NHS patients, then this model would resemble the simpler profit maximization model.

Nonfundholders make more than "appropriate" referrals.

We denote as "appropriate" the number of referrals that maximizes patient welfare as defined by GPs, following the approach of Woodward and Warren-Boulton (1984). Given our assumption about the relationship between L_R and W_R, the first-order condition for nonfundholders implies that $\mathbf{W}_R < 0$: nonfundholders can increase patient welfare by reducing their number of referrals.

Nonfundholders increase their utility both from attracting patients to their practice and from improving patient welfare. After some point, additional referrals still attract increasing numbers of patients (increasing GP profits) even though these referrals become detrimental to patient welfare as defined by GPs. GPs will increase the referrals beyond the "appropriate" level as long as the gains from increased profits compensate them for the disutility of reducing patient welfare.

For fundholders, the sign of $\mathbf{W}_R$ is indeterminate; it depends on the sign of $(p_c p_r C_S - p_l \mathbf{L}_R)$. Besides list size and patient welfare, fundholders consider the profits to be derived from investing budget surpluses. If surpluses are sufficiently profitable relative to capitation payments, or if referrals are sufficiently costly, fundholders may choose to refer below the "appropriate" level ($\mathbf{W}_R > 0$). The policy implications of this possibility depend on the GPs' definition of patient welfare. If GP decisions do not fully reflect the opportunity costs to the NHS, then referral rates at the "appropriate" level will be inefficiently high from an economic perspective.

Fundholding reduces the optimal referral rates.

The proof parallels the one for the profit maximization model. From the first-order conditions, we obtain

$$(\mathbf{L}_R)_{\text{fundholder}} > (\mathbf{L}_R)_{\text{nonfundholder}}$$

Given the concavity of L, this inequality implies that $\mathbf{R}_{\text{fundholder}} < \mathbf{R}_{\text{nonfundholder}}$.

The low elasticity of L with respect to R, in turn, implies lower referral rates for fundholders. Again, profits from budget surpluses provide fundholders with an additional inducement to decrease referrals.

An increase in capitation payments (p_l) can either increase or decrease the number of referrals of fundholders and nonfundholders.

For nonfundholders, we use the first-order condition to define:

$$F \equiv \mathbf{L}_R + (1/p_l)(\mathbf{U}_W/\mathbf{U}_\pi)\mathbf{W}_R \equiv 0$$

From the implicit function theorem, we obtain

$$\partial \mathbf{R}/\partial p_l = - (\partial F/\partial p_l)/(\partial F/\partial \mathbf{R})$$

in which

$$\partial F/\partial p_l = - (1/p_l^2)(\mathbf{U}_W/\mathbf{U}_\pi)\mathbf{W}_R(1 + p_l\mathbf{L}\mathbf{U}_{\pi\pi}/\mathbf{U}_\pi)$$

$$\partial F/\partial \mathbf{R} = \mathbf{L}_{RR} + \mathbf{U}_{WW}\mathbf{W}_R^2/p_l\mathbf{U}_\pi - \mathbf{U}_W\mathbf{U}_{\pi\pi}\mathbf{L}_R\mathbf{W}_R/\mathbf{U}_\pi^2 + \mathbf{U}_W\mathbf{W}_{RR}/p_l\mathbf{U}_\pi$$

Since all the terms in $\partial F/\partial \mathbf{R}$ are negative, $\partial F/\partial \mathbf{R} < 0$. The sign of $\partial F/\partial p_l$, however, is indeterminate, since we do not know the sign of $(1 + p_l\mathbf{L}\mathbf{U}_{\pi\pi}/\mathbf{U}_\pi)$.[12]

The reasons for this indeterminacy are analogous to the income and substitution effects. Higher capitation payments make the opportunity costs of increasing patient welfare higher, inducing GPs to substitute away from patient welfare by increasing referrals. However, increases in p_l also imply that a given number of referrals yields higher profits--the marginal utility of the last pound of profits becomes smaller than the marginal utility of the last unit of patient welfare, inducing GPs to decrease referrals. The above equation for $\partial \mathbf{L}_R/\partial p_l$ indicates whether the balance of these forces will lead to increases or decreases in referrals.

For fundholders, we obtain:

$$\partial F/\partial p_l = (p_c p_r k/p_l^2) - (1/p_l^2)(\mathbf{U}_W/\mathbf{U}_\pi)\mathbf{W}_R[1 + p_l\mathbf{L}\mathbf{U}_{\pi\pi}/\mathbf{U}_\pi]$$

$$\partial F/\partial \mathbf{R} = \mathbf{L}_{RR} + \mathbf{U}_{WW}\mathbf{W}_R^2/p_l\mathbf{U}_\pi - \mathbf{U}_W\mathbf{U}_{\pi\pi}\mathbf{L}_R\mathbf{W}_R/\mathbf{U}_\pi^2 + \mathbf{U}_W\mathbf{W}_{RR}/p_l\mathbf{U}_\pi$$

In this case, the signs of both $(1 + p_l\mathbf{L}\mathbf{U}_{\pi\pi}/\mathbf{U}_\pi)$ and $\mathbf{W}_R$ are indeterminate, so that again $\partial \mathbf{R}/\partial p_l$ can be either positive or negative. Note that this proposition about the effects of changes in capitation payments differs from the one we obtained in the profit maximization model.

---

[12]Note that $|p_l\mathbf{L}\mathbf{U}_{\pi\pi}/\mathbf{U}_\pi| \equiv |\pi\mathbf{U}_{\pi\pi}/\mathbf{U}_\pi|$, Arrow's coefficient of relative-risk aversion. (see Arrow, 1974) Thus, the sign of $\partial \mathbf{R}/\partial p_l$ depends on whether the coefficient of relative risk aversion is less than or greater than one. If this coefficient is greater than one, then $\partial \mathbf{R}/\partial p_l > 0$; if it is less than one, then $\partial \mathbf{R}/\partial p_l < 0$.

4.6 The substitution of prescribing for referrals: a sample application of the models

We illustrate here how we can apply the above models to design an empirical study of fundholding. Model 4.3 suggested that for fundholders the incentive to substitute cheaper prescribing for more expensive referrals may counteract the incentive to reduce prescribing costs. Based on this insight, we propose the following hypothesis: *in therapeutic categories with large scope for substituting prescribing for referrals, the difference in prescribing rates between nonfundholders and fundholders will be smaller than in categories with no scope for substitution.*

To explore this hypothesis further, we expand model 4.3 by sub-dividing the referral and prescription variables into n therapeutic categories. The model for fundholders now becomes:

$$\max_{\underline{R},\underline{P},S} \pi$$

$$\pi = p_l L(\underline{R},\underline{P}) + p_c C(S)$$

subject to

$$\Sigma p_{ri} R_i + \Sigma p_{pi} P_i + S = B$$

where

R_i = referrals in therapeutic category i
p_{ri} = average price of referrals in therapeutic category i
P_i = prescriptions in therapeutic category i
p_{pi} = average price of prescriptions in therapeutic category i
$\underline{R} = (R_1,..., R_n)$
$\underline{P} = (P_1,..., P_n)$

and the other variables are defined as before.

We make the further assumptions $L_{RiRj} = 0$, $L_{RiPj} = 0$, $L_{PiPj} = 0$ for $i \neq j$: there is no substitutability between treatments in different therapeutic categories.

From the first-order conditions for nonfundholders, we obtain:

$$\mathbf{L_{Ri}} = (p_c p_{ri}/p_l)k \qquad \text{for all i}$$

$$\mathbf{L_{Pi}} = (p_c p_{pi}/p_l)k \qquad \text{for all i}$$

For nonfundholders, we obtain:

$$\mathbf{L_{Ri}} = 0 \qquad \text{for all i}$$

$$\mathbf{L_{Pi}} = 0 \qquad \text{for all i}$$

To examine our hypothesis, we focus on two therapeutic categories: category h, with a large scope for substitution, and category k, with no scope for substitution. Our hypothesis posits that:

$$\mathbf{P}_{\text{nonfundholder, h}} - \mathbf{P}_{\text{fundholder, h}} < \mathbf{P}_{\text{nonfundholder, k}} - \mathbf{P}_{\text{fundholder, k}}$$

Under what conditions does this inequality (our hypothesis) hold? To solve analytically for **P**, we resort to our earlier parametrization, letting:

$$L_i = a_{ri}R_i + a_{pi}P_i - b_{ri}R_i^2 - b_{pi}P_i^2 - c_iR_iP_i \qquad \text{for all i}$$

with a_{ri}, a_{pi}, b_{ri}, $b_{pi} > 0$ for all i. Since c_i represents the extent of substitution between prescriptions and referrals in category i, in our case $c_h > 0$ and $c_k = 0$. Note that our assumptions allowed us to separate the effects on list size of referrals and prescriptions in the various therapeutic categories.

Using this parametrization of L with the optimizing conditions for fundholders and nonfundholders, we can establish that our hypothesis implies:

$$c_h(2b_{pk}p_{rh} - c_hp_{pk}) + 4(b_{rk}b_{ph}p_{pk} - b_{rh}b_{pk}p_{ph}) > 0$$

Without further information, we cannot determine whether this inequality will hold. This indeterminacy suggests that our hypothesis implicitly entails additional assumptions. By searching for conditions that satisfy the above inequality, we can uncover these assumptions, namely:

- The relationship between list size and referral/prescription levels is the same for all therapeutic categories. Thus, $b_{rk} = b_{rh}$, $b_{ph} = b_{pk}$. [13]
- The price difference between referrals and prescriptions is "sufficiently" larger than the price difference between prescriptions in the various categories.

After incorporating these assumptions into the model, we can assess its validity by empirically testing our hypothesis.[14] If the results corroborate the hypothesis, they will

[13] Alternatively, we can make the weaker assumption that variations in the list size function among therapeutic categories are uncorrelated with variations in the scope for substitution. In this case, our empirical analysis will require choosing enough therapeutic categories in the "high substitution" and "no substitution" groups to control for random differences between the two groups in the list size function.

provide support for our model of GP behaviour, suggesting that for fundholders the incentive to substitute prescribing for referrals counteracts the incentive to reduce prescribing costs. On the other hand, if the results reject our initial hypothesis, they will highlight the relative insensitivity of GP fundholders to the financial incentives of the fundholding scheme, or the ignorance of GPs about the scope for substitution between prescribing and referrals within the various therapeutic categories. Additionally, these results would establish the inadequacy of our assumptions about GP behaviour. In either case, the study will increase our knowledge about GP responses to financial incentives.

This example emphasizes some of the benefits of modelling GP behaviour explicitly:

- We began with a simple model of referral and prescribing behaviour. This model suggested a testable hypothesis contradicting the conventional wisdom about the effect of fundholding on prescribing rates.

- After choosing to concentrate on the substitution of prescribing for referrals, we developed the models further to focus on this question. In the process, we uncovered additional assumptions implicit in our hypothesis.

- By developing a conceptual foundation, we ensured that the study will contribute to our knowledge of GP responses to fundholding whether or not the results corroborate our initial hypothesis.

4.7 Future directions for modelling

The models developed here illustrate how we can construct a theoretical foundation to address various empirical questions. Many other dimensions of behaviour may be amenable to modelling. These may include:

- Incorporating a quality dimension for referrals, entering into the profit, list size and patient welfare functions. Fundholders may seek to attract patients, or improve patient welfare, by spending additional funds for improvements in hospital services, such as shorter

---

[14]We have developed our hypothesis only in the context of a particular parametrization of the list size function. We should consider as well whether additional assumptions are necessary for our hypothesis to hold in the general case.

waiting lists. They may trade off the added costs of higher quality with the benefits of attracting more patients, implying as well trade-offs between quantity and quality.

- Modelling the *variation* in referral rates, perhaps by allowing for differing practice styles or clinical uncertainty. For referral rates, their variation, rather than their average level, crystallized concern about the efficiency of GP behaviour. The construction of these models may begin with the existing literature addressing variation in referral rates.

- Incorporating various contract forms in the models. The financial incentives of fundholding affect the choice of contractual arrangements. These arrangements, in turn, generate their own sets of incentives influencing behaviour.

- Exploring the interaction of fundholders with providers, manifested, for example, in the types of contracts chosen or in the decision to provide services in-house. In these models, GP objectives such as autonomy, workload or risk reduction may be relevant.

- Modelling the organization of fundholding practices. Specific areas for analysis may include the formation of consortia and the "optimal" list size.

5 CONCLUSIONS

Given the potentially broad repercussions for the NHS of the fundholding scheme, more extensive evaluation of its effects is imperative. This work lays the foundations for modelling the behaviour of GPs, a necessary first step for sound empirical analysis. We have outlined several models incorporating various incentives and GP objectives. The model we choose ultimately depends on the empirical question we seek to answer. We must use care to include all the relevant factors, and to exclude irrelevant ones that add needless complexity. For example, if we want to investigate the effect of fundholding on prescribing rates, we should consider including referral rates in our model (to account for the possibility of substitution) but we may not need to make the budget endogenous (as this may not affect our predictions). By testing the predictions our model empirically, we can assess the adequacy of its assumptions. Whatever our findings, we will gain a greater understanding of how GPs respond to the incentives of the fundholding scheme.

REFERENCES

Arrow KJ (1974), *Essays in the Theory of Risk Bearing*, North Holland, Amsterdam.

Audit Commission (1993), *Practice Makes Perfect: The Role of the Family Health Services Authority*, HMSO, London.

Bosanquet N and Leese B (1988), "Family doctors and innovation in general practice", *British Medical Journal*, 296: 1576-80.

Bradlow J and Coulter A (1993), "Effect of fundholding and indicative prescribing schemes on general practitioners' prescribing costs", *British Medical Journal*, 307: 1186-9.

Brazier J, Hutton J and Jeavons R (1990), *Analyzing Health Care Systems: The Economic Context of the NHS White Paper Proposals*, NHS White Paper, Occasional Paper 10, Centre for Health Economics, University of York.

Butler JR and Calnan MW (1987), "List sizes and use of time in general practice", *British Medical Journal*, 295: 1383-6.

Calnan MW and Butler JR (1988), "The economy of time in general practice: an assessment of the influence of list size", *Social Science and Medicine*, 26, 4: 435-41.

Coulter A and Bradlow J (1993), "Effect of NHS reforms on general practitioners' referral patterns", *British Medical Journal*, 306: 433-7.

Crump BJ, *et al* (1991), "Fundholding in general practice and financial risk", *British Medical Journal*, 302: 1582-4.

Day P and Klein R (1991), "Variations in budgets of fundholding practices", *British Medical Journal*, 303: 168-70.

Department of Health (1989a), *Working for Patients*, White Paper, HMSO, London.

Department of Health (1989b), *Practice Budgets for General Medical Practitioners*, Working Paper 3, HMSO, London.

Department of Health (1989c), *Indicative Prescribing Budgets for General Medical Practitioners*, Working Paper 4, HMSO, London.

Department of Health and Social Security (1987), *Promoting Better Health*, HMSO, London.

Dranove D (1985), *Demand Inducement of the Physician-Patient Relationship*, Working Paper, University of Chicago, Chicago, IL.

Evans RG (1974), "Supplier-induced demand: some empirical evidence and implications", in Perlman M (ed.), *The Economics of Health and Medical Care*, Stockton Press, New York.

Evans RG (1976), "Modelling the Economic Objectives of the Physician", in Fraser RD (ed.), *Health Economics Symposium*, Queen's University and Kingston, Canada.

Glennerster H, Matsaganis M and Owens P (1992), *A Foothold on Fundholding*, King's Fund Institute, London.

Glynn JJ, Murphy MP and Perkins DA (1992), "GP practice budgets: an evaluation of the financial risks and rewards", *Financial Accountability & Management*, 8(2): 149-61.

Groenewegen PP and Hutten JBF (1991), "Workload and job satisfaction among general practitioners: a review of the literature", *Social Science and Medicine*, 32: 1111-19.

Healy A and Yule B (1993), "The economics of GP fundholding", paper presented at the Health Economists' Study Group, July 1993.

Hellinger FJ (1987), "Selection bias in health maintenance organizations: analysis of recent evidence", *Health Care Financing Review*, Winter, 9(2): 55-63.

Hillman AL (1987), "Financial incentives for HMOs: is there a conflict of interest?", *New England Journal of Medicine*, 317: 1743-8.

Hillman AL, Pauly MV and Kerstein JJ (1989), "How do financial incentives affect physicians' clinical decisions and the financial performance of Health Maintenance Organizations?", *New England Journal of Medicine*, 321: 86-92.

Hornbrook MC and Berki SE (1985), "Practice mode and payment method: effects on use, costs, quality and access", *Medical Care*, 23(5): 484-511.

Knight, R (1987), "The importance of list size and consultation length as factors in general practice", *Journal of the Royal College of General Practitioners*, 37: 19-22.

Luft HS (1981), *Health Maintenance Organizations: Dimensions of Performance*, Wiley-Interscience, New York.

Manning WG, *et al* (1987), "Health insurance and the demand for medical care: evidence from a randomized experiment", *American Economic Review*, 77(3): 251-77.

Maxwell M, *et al*, (1993), "General practice fundholding: observations on prescribing patterns and costs using the defined daily dose method", *British Medical Journal*, 307: 1190-4.

Maynard A, Marinker M and Grey DP (1986), "The doctor, the patient and their contract; alternative contracts: are they viable?", *British Medical Journal*, 292: 1438-40.

Moore SH, Martin DP and Richardson WC (1983), "Does the primary-care gatekeeper control the costs of health care? Lessons from the SAFECO experience", *New England Journal of Medicine*, 309: 1400-4.

"New drug budgets penalize efficient GPs" (1993), *Fundholding*, 21 Jan.: 5.

Newton J, *et al* (1993), "Fundholding in Northern region: the first year", *British Medical Journal*, 306: 375-8.

Phelps CE (1986), "Editorial. Induced demand--Can we ever know its extent?", *Journal of Health Economics*, 5, 355-65.

Reinhardt U (1985), "Editorial. The theory of physician-induced demand: reflections after a decade", *Journal of Health Economics*, 4, 187-93.

Roberts SJ and Harris MC (1993), "Age, sex, and temporary resident originated prescribing units (ASTRO-PUs): new weightings for analysing prescribing of general practices in England" *British Medical Journal*, 307: 485-8.

Scheffler R (1989), "Adverse selection: the Achilles heel of the NHS reforms", *The Lancet*, April 29: 950-2.

Serra A (1992), "Where GP fundholding is the only topic in town", *Fundholding*, June 7: 12-13.

Thomas K, *et al*, (1989), "Estimates of general practice workload: a review", *Journal of the Royal College of General Practitioners*, 39: 509-13.

Ware JE, *et al* (1986), "Comparison of health outcomes at a health maintenance organization with those of fee-for-service care", *The Lancet*, May 3: 1017-22.

Weiner JP and Ferriss DM (1990), *GP Budget Holding in the UK: Lessons from America*, Research Report No. 7, King's Fund Institute, London.

Wilkin D and Metcalfe DHH (1984), "List size and patient contact in general medical practice", *British Medical Journal*, 289: 1501-5.

Woodward RS and Warren-Boulton F (1984), "Considering the effects on financial incentives and professional ethics on 'appropriate' medical care", *Journal of Health Economics*, 3: 223-37.

Yett DE, et al (1983), "Physician pricing and health insurance reimbursement", *Health Care Financing Review*, 5(2): 69-80.

Zweifel P (1981), "'Supplier-induced demand' in a model of physician behavior", in Van der Gaag J and Perlman M (eds.), *Health, Economics, and Health Economics*, North-Holland, Amsterdam.