

CHAPTER 7

PROFESSIONAL PRACTICES: THE NOT-ONLY-FOR-PROFIT FIRMS

THE SCALE OF THE NOFP SECTOR IN CANADA: EXPENDITURES AND EMPLOYMENT

In the previous chapter we suggested that it would prove helpful for analytic purposes to categorize health care firms according to the objectives which they appear to pursue. Some health care firms, like most firms in the general economy, are organized for the purpose of earning as large a profit as possible for their owners, *e.g.*, the for-profit (FP) firms which dominate the drug and equipment industries. Others, like most hospitals, have no participants, whether managers or owners, with personal legal title to profits; whatever other objectives they pursue they are not organized and managed with profit as an objective (NFP). Intermediate are not-only-for-profit (NOFP) firms, in which a legal claimant to profits is well-defined, but profits represent only one among several competing objectives of the firm's ownership and management.

The NOFP sector in Canada is represented by a wide range of different types of firms owned and run by self-employed professional practitioners. Medical practices, however, account for by far the largest share of both employment and economic activity. The other large sub-sectors are dental practices and community pharmacies, although, as noted above, the latter shades into the corporate, for-profit sector. Both conceptual and statistical boundaries are difficult to draw with precision.

The scale of an industry may be variously described by its sales, its employment levels, or direct measures of its physical capacity. Sales data in the health context are usually referred to as expenditures; but of course each dollar of expenditure is simultaneously a dollar of revenue for the provider. Expenditures on medical and dental services were displayed in Table 1-1; in 1982 these items amounted to \$4.4 billion and \$1.7 billion respectively. They made up a fifth of all health spending, and nearly 2 percent of GNP. The NOFP sector also includes services provided by other professional practices, chiropractors, optometrists, and others, but these in 1982 totalled under half a billion, and their share has been declining slowly over time.

The treatment of drugs and appliances is more problematic. These accounted for \$3.3 billion in sales in 1982, of which prescribed drugs were \$1.5 billion. But even if one counted all the dispensing activities of pharmacies as part of the NOFP sector, that would represent only somewhat over half of prescribed drug sales. The rest, cost of goods sold, is the output of the (strictly FP) drug manufacturing sector. And of course many non-prescribed drugs (total sales \$1.4 billion) are sold elsewhere in the retail sector, while the "professional" content of their sale even by pharmacies is questionable to non-existent. On balance, then, the most generous addition to NOFP firm sales from the drug and appliance field would be about \$0.8 billion, just over half prescription drug sales, with the true figure being perhaps much less.

Medical and dental practices not only make up the bulk of the NOFP sector, they also display the most variation through time. As Table 1-1 showed, physicians almost doubled their share of national income from 1951 to 1971, moving from 0.71 percent to 1.32 percent as their total sales rose from \$153.0 million to \$1250.4 million. But by 1976 they had dropped back sharply, to 1.10 percent, despite sales of \$2103.2 million, and this share remained fairly constant through to 1981 (1.10 percent and \$3741.0 million), rising to 1.24 percent on sales of \$4414.3 million in 1982. Dentists, on the other hand, moved their share up sharply during the 1970s, after two decades of relative stability. In both 1946 and 1971 they accounted for 0.33 percent of GNP -- recovering after a fall in the late 1940s -- but by 1979 they had added nearly a third to their share, up to 0.42 percent, and in 1982 reached 0.47 percent. Between 1970 and 1982 dentist billings grew by a factor of 6 -- the second fastest increase of any health care component. ("Homes for Special Care" or care of the aged rose over nine times.) These swings in expenditure patterns do not show up in the smaller and/or more questionable components of the NOFP sector, whose shares of total national income have been relatively stable in the post-war period. Medical and dental practices appear to be where the action, as well as most of the money, is.

But the significance of such swings in expenditure/sales data depends on the reliability of the data themselves. And this, of course, varies according to the source. Data on billings by firms is most readily available and reliable under a universal public health insurance scheme; thus expenditure data improved in quality and coverage during the 1960s as the public plans developed. The existence of extra-billing (billing both patient and insurer) and opting-out, as well as various forms of uninsured services, renders this source still incomplete. Taxation statistics from self-employed practitioners form another source of gross and net income data, but they suffer from their own biases. Moreover published data are for net practitioner incomes only, not gross practice receipts, except for special tabulations prepared periodically by Health and Welfare Canada.¹

Expenditure/sales data also reflect changes in relative prices and incomes as well as shifts in actual scale of economic activity. Much of the increase in expenditures on physician services from 1951 to 1971, for example, was a result of increases in the relative price of their services, increases in both fees charged and collections ratios. Industry employment data, on the other hand, provide an indication of the actual resources, at least of human time and skills, used up in the sector. They are at best only a first approximation to real output levels, however, since those depend also on changes in productivity per person employed.

Table 7-1 presents data showing the growth through time in the stock of physicians, dentists, and pharmacists, relative to the total labour force. Almost a third of licensed physicians are employed in other than fee-for-service practice, which places them outside the NOFP sector. Most are employed in NFP firms, hospitals or other institutions. Dentists, by contrast, are almost entirely in fee-for-service practice; of 11095 licensed dentists in 1980, 9877 were reported by Health and Welfare Canada (see Data Sources Appendix) as self-employed practitioners, and some of the remainder would be salaried in private practice. Pharmacists likewise are predominantly in (salaried or self-employed) community pharmacy; about 10 percent would be working in hospitals or other NFP institutional settings, and a small proportion in FP drug manufacture or distribution.

TABLE 7-1
Professional Manpower in the NOFP Sector, 1946-1981,
Relative to the Overall Labour Force ('000)

	Active Civilian Physicians	Physicians in Fee-For-Service Practice	Licensed Dentists	Licensed Pharmacists	Labour Force ('000)
1946	12831 (0.266)	6343 (0.131)	4649 (0.096)	6000 (0.124)	4829 --
1951	14325 (0.274)	8790 (0.168)	5019 (0.096)	6400 (0.122)	5223 --
1956	17871 (0.309)	11868 (0.205)	5549 (0.096)	7856 (0.136)	5782 --
1961	21290 (0.326)	14588 (0.224)	5986 (0.092)	9022 (0.138)	6521 --
1966	26528 (0.358)	15361 (0.027)	6399 (0.086)	9863 (0.133)	7420 --
1971	32942 (0.382)	20742 (0.240)	7453 (0.086)	11330 (0.131)	8631 --
1976	40130 (0.393)	27395 (0.268)	9401 (0.092)	14687 (0.144)	10206 --
1977	41398 (0.394)	26586 (0.253)	10058 (0.096)	15328 (0.146)	10498 --
1978	42238 (0.388)	26819 (0.246)	10451 (0.096)	15709 (0.144)	10882 --
1979	43192 (0.385)	27419 (0.245)	10763 (0.096)	16052 (0.143)	11207 --
1980	44275 (0.384)	28739 (0.249)	11095 (0.096)	16588 (0.144)	11522 --
1981	45542 (0.385)	29633 (0.250)	11484 (0.097)	17039 (0.144)	11830 --

SOURCES: See Data Sources Appendix.

Bracketed figures are percentages of labour force. "Physicians in Fee-For-Service Practice" is measured by taxable returns of self-employed practitioners.

Other than these three groups, self-employed private practitioners whose practices might be included in the NOFP sector are rich in variety but small in overall numbers. As of 1981 there were 74065 licensed physicians, dentists, and pharmacists; by contrast, other health professionals who would often be in self-employed practice were: chiropractors, 2412; optometrists, 2070 (opticians, 2726); osteopaths, 48; podiatrists, 259; veterinarians, 4236 (Canada, Health and Welfare Canada 1983*b*). Table 7-1 thus covers almost all the NOFP sector.

Employment in the NOFP sector excludes professional personnel in academic, government, or administrative roles. But it includes employees in professional practices other than licensed professionals themselves. Table 7-2 shows Census data for 1961 and 1971 on employment in selected professional practices. Unfortunately, the Census data do not distinguish licensed professionals (owners or employees) in professional practices from other types of personnel, so that one cannot directly determine the extent of other employment. An idea of its significance can, however, be derived from the reported sex data. These can be used to calculate a minimum (lower bound) estimate of the numbers of people, other than licensed professionals, working in professional practices. This minimum estimate assumes that all female professionals work in private practice, and that no male non-professionals do -- more realistic assumptions would raise the estimate of non-professional employees.

TABLE 7-2
Canadian Labour Force Working in Professional Practices,
Census Data, 1961 and 1971

	1961	1971
Offices of Physicians and Surgeons:		
Total	25902	43285
Male	14006	17690
Female	11896	25585
Offices of Dentists:		
Total	9385	15445
Male	5143	6120
Female	4242	9325
Offices of Paramedical Practitioners:		
Total	--	7830
Male	--	3070
Female	--	4755
Drug Stores:		
Total	26933	38455
Male	14171	14735
Female	12762	23700

SOURCE: See Data Sources Appendix

In 1971, the Census reported 9325 females working in dental practices, but only 310 licensed female dentists. At least 9015 females, other than dentists, must therefore have been working in dental practices. The 1971 Census also reported 2890 female physicians, and 2170 female pharmacists, implying at least 22695 female non-physician employees in medical practice, and 21530 female non-pharmacists in drug stores. Since not all female professionals will be in private practice, and some males in such practices will be employees who are not members of the profession controlling the practice, one can estimate that there was at least one non-physician employee for each physician in private practice in 1971, and at least three other

employees for each dentist. Indeed, males working in drug stores (14735) substantially exceeds reported licensed male pharmacists (7240) so that even if *all* licensed pharmacists were in community practice, there would be three additional employees for each pharmacist.

Census data from 1961 for physicians' and dentists' practices permit a similar computation, showing at least 4007 female non-dentist employees and 10441 female non-physician. These minimal estimates of practice auxiliaries make up 42.7 percent and 40.3 percent of total practice employment; the corresponding 1971 percentages are 58.4 percent and 52.4 percent. This indicates a significant increase in non-professional employment in these professional practices over the decade of the 1960s, and the 1981 Census data will be of considerable interest to see if the trend continues.

Linking Tables 7-1 and 7-2, moreover, permits one to compare the professional employment patterns. The increase in numbers of physicians as a proportion of the labour force is quite striking; from 1946 to 1971 this proportion rose by 50 percent, though it has been relatively stable since. A substantial increase in share took place from 1961 to 1971, 17.2 percent, but employees of physician practices rose even faster. There were 67.1 percent more people reported as working in medical practices in 1971 than in 1961, while licensed physicians increased 54.7 percent, and physicians in fee-for-service practice rose only 47.5 percent. These comparisons underline the incompleteness of professional manpower data alone as descriptors of the economic significance of the NOFP sector, or of its growth.

Dental practice data make a similar point. Table 7-1 shows that the proportion of the Canadian labour force made up of dentists was identical in 1946 and 1980, having dipped about 10 percent by 1971 and recovered by 1980. Table 1-1 showed that the share of national income accounted for by dental expenditures/sales was similarly more or less constant from 1946 to 1971, but the rise since 1971 greatly exceeded the roughly 10 percent increase in proportion of dentists in the labour force. Similarly, the fall in the share of medical expenditures in national income since 1971 has occurred without parallel labour force shifts. Between 1961 and 1971, however, numbers of employees in dental practices rose 64.6 percent, while numbers of licensed dentists rose only 24.5 percent, so resource input to the sector was expanding much faster than professional manpower data would suggest. Once again, better understanding of the growth of the NOFP sector in the 1970s must wait on 1981 Census data -- no other source reports practice manpower.²

Employment in the various NOFP sectors is an indicator of the opportunity or resource costs, the human time and skills, used up in those activities. Expenditure data, on the other hand, are the product of output and price levels. Output can be expanded through productivity improvement and technological change even while employment is constant. But some discussions of productivity in professional practices follow the tendency of the data sources to neglect information on salaried employees of professionals, with the result that increases in productivity are measured per professional only. If such increases result from more use of other personnel per peak professional, the result may or may not be an increase in productivity from a social perspective. If output per professional can be doubled by adding one auxiliary employee, then (unless such employees are as expensive as professionals) productivity is clearly increased. But if five auxiliaries are needed (in the practice or elsewhere, such as in a hospital) for such doubling, then overall, "total factor productivity" will have fallen. Yet if the output generates revenue in the practice, and the employees can be placed on someone else's payroll, they may still be added.

In Table 7-3 we present indices of "real" volumes of physicians' and of dentists' services; that is, expenditure data adjusted for estimates of fee change. These are then used to derive indices of "real" service use per capita and output per professional -- active civilian physicians

and licensed dentists, respectively. These are *not* indices of output per full-time equivalent private practitioner, which one would prefer, but will deviate only insofar as the proportion of professionals in full-time private practice has changed over time. The dental fee index used in these calculations is simply the Consumer Price Index dental care component; the medical care fee is derived by linking earlier CPI data with later medical care fee schedule data and adjusting for estimated rates of collection (Barer and Evans 1983). Since 1971 it has followed the national fee indices prepared by Health and Welfare Canada, thus failing to capture the effects of changes in extra-billing patterns. Table 7-3 also displays movements through time in medical and dental fees relative to the overall Consumer Price Index.

TABLE 7-3
Indices of Quantity of Physicians' and Dentists' Services,
Adjusted Price Change, and of Relative Prices, 1946-1982

	Physicians' Services			Dentists' Services			Relative Fees (Over CPI)	
	Total	Per Capita	Per Physician	Total	Per Capita	Per Dentist	Medical	Dental
1946	23.5	41.3	60.4	37.5 ¹	60.2 ¹	58.6 ¹	64.8	63.5 ¹
1951	27.4	42.3	63.0	40.4	62.2	60.0	67.7	61.4
1956	39.1	52.5	72.0	54.5	73.1	73.2	71.5	70.1
1961	50.8	60.1	78.5	64.3	76	80.1	81.5	77.7
1966	71.9	77.6	89.4	77.6	83.6	90.4	80.6	87.4
1971	100.0	100.0	100.0	100.0	100	100.0	100.0	100.0
1976	138.1	129.4	113.4	142.7	133.7	113.1	81.8	105.7
1977	139.9	129.6	111.3	157.4	145.8	116.6	82.1	105.0
1978	145.1	132.8	113.2	168.4	154.6	120.1	80.0	103.8
1979	151.0	136.4	115.2	177.2	159.1	122.7	78.9	104.8
1980	159.4	142.7	118.6	187.0	167.5	125.6	78.3	105.0
1981	162.4	143.8	117.5	193.9	171.6	125.8	77.8	103.6
1982	169.5	149.2	117.8	197.3	173.7	123.8	79.4	104.3
Annualized Rates of Change (%)								
1946-56	5.22	2.43	1.77	5.49 ¹	2.81 ¹	3.23 ¹	1.02	1.42 ¹
1956-66	6.28	3.98	2.19	3.60	1.35	2.13	1.18	2.23
1966-76	6.74	5.25	2.41	6.28	4.81	2.27	0.15	1.92
1976-82	3.47	2.40	0.64	0.06	4.46	1.52	-0.50	-0.22

SOURCES: See Data Sources Appendix

¹ Dental fees are from 1949, not 1946..

From these data we can see the steady growth in output, or at least fee-adjusted billings, per professional, at about 2 to 2.5 percent per year over most of the post-war period. Only for dentistry in the 1949-51 period, medicine post-1976, and dentistry since 1980 do we see any break in this pattern. This growth is presumably a result both of increases in other inputs per professional -- capital equipment and auxiliary personnel -- and of "true" productivity increase or

greater output per some index of inputs weighted by their cost.³ The increases in other inputs may in turn be within the practice -- an increase in hired auxiliaries per dentist, for example -- or outside it. The expansion of the hospital system in the 1950s and 1960s clearly increased the levels of output which physicians could generate and bill for. And the sudden drop in rates of increase of output per physician in the late 1970s may be, in part, a result of the restraints on hospital expansion and the consequent reduction in hospital equipment and personnel available per physician.

The prices of medical and dental services, relative to the general price level, both moved substantially in the post-war period. Both escalated rapidly and steadily up to 1971. Estimated medical fees (adjusted for changed collections ratios) rose at an average of 1.75 percent faster than the general price level over the whole thirty-five year period 1946-71, to end 53.4 percent higher in real terms, while dental fees rose even faster, 2.09 percent per year from 1949 to 1971, for a total gain of 57.5 percent.

But after 1971, the pattern changes. Medical fees, presumably under the influence of the public insurance plans, fell sharply from 1971 to 1976, back to their relative level of 1961. They have continued to move down, though much more slowly, in the 1976-81 period. Dental fees slowed their rate of increase, but continued to escalate about 1 percent per year faster than the general price level during the early 1970s. Since then, they have run very slightly behind the CPI.

Utilization levels per capita for medical and dental care have also shown substantial movement, though with somewhat different timing. Per capita dental billings adjusted for fee schedule change moved up steadily in the 1946-66 period, about 1.95 percent per year. But in the latter half of the 1960s they speeded up sharply, and rose even faster in the late 1970s. Medical care use, on the other hand, speeded up earlier. It more or less paralleled dental care use increases in the first post-war decade, ran well ahead after 1956 and up to 1971, grew slightly more slowly 1971-76, and in the late 1970s and early 1980s ran well behind. The combination of continued restraints on physician numbers and billing patterns, compared with dental markets which are controlled only by the profession itself, can be expected to lead to continuing divergence in these trends. In referring to these differences as "productivity," however, we must keep in mind the problematic nature of the linkages between reported billings and care supplied, and between health care and health status.

In any case, a focus on licensed self-employed personnel alone seriously understates both the level of employment and the growth of output, or at least fee-adjusted billings, in the NOFP sector. Moreover, a significant amount of medical, and some dental, service production takes place outside owner-managed practices, in hospitals, public clinics, or other government agencies. The medical services industry and the collection of NOFP firms in medical care are not coterminous; we isolate NOFP firms in order to give a coherent description of their economic behaviour.

ECONOMISTS AND THE NOFP FIRM: THE BLIND MEN AND THE ELEPHANT

Economic analysis of NOFP firms can be traced through three successive phases, each emphasizing different aspects of such firms, or different characteristics distinguishing them from "normal" or text-book firms in private markets (Evans 1980). A broader discussion of the policy issues is provided by the papers in Slayton and Trebilcock (1978). Analysts have focussed in turn on the monopoly or cartel nature of self-regulation, on the self-employed, "labour-managed"

nature of the firm itself, and on the direct influence of the firm over the demand for its own services implied by the professional relationship.⁴ Each stage implies different assumptions about the NOFP firm's objectives and constraints, predictions about its behaviour, and recommendations as to appropriate policy.

PROFESSIONS AS MONOPOLIES: SUPPLY RESTRICTIONS

The monopoly or cartel approach has several variants, but all emphasize the role of licensure in restricting entry to particular professions. It is implicitly assumed that this in turn automatically restricts access to, and output offered in, particular service markets, although in practice that need not follow. The bitter jurisdictional disputes between dentists and denturists or dental mechanics, physicians and chiropractors, and outside health care, lawyers and notaries, chartered and certified general accountants, and engineers and architects, demonstrate that the definition of a profession does not uniquely determine the boundaries of its domain of exclusive practice rights. The "senior" profession in each of the above disputes, *i.e.*, the one claiming the broadest domain, has generally asserted the prerogative of setting its own boundaries. Much economic analysis has implicitly, and rather surprisingly, accepted and supported this political program by taking the scope of professional practice to be whatever the profession declares it to be.

In order to restrict the supply of services which professionals participate in producing, professional licensure must both restrict access to the profession itself and maintain the boundaries around the collection of activities which are defined as "the practice of X." The law may or may not forbid unlicensed persons actually to *perform* such activities, but it clearly forbids firms other than those owned/controlled by licensed persons to *supply* such services to patient/consumers. Depending on the technology, however, restriction of supply may also require regulation of the structure and conduct of professionally-owned firms themselves. If the nature of the production process is such that one or a few licensed professionals may dramatically expand the output of their firms by hiring large numbers of non-licensed personnel, and delegating functions, then control of service output may require direct restrictions on production itself. Conduct regulation may specify maximum ratios of other personnel per licensed professional in each firm, or may specifically prohibit non-licensed personnel from performing certain functions. Similar constraints may be achieved through collective standards for defence in malpractice suits, or negotiated restrictions on insurance reimbursement.

In this framework, the essence of the professional firm is assumed to be supply restriction -- "professional birth control." The market demand for professional services is assumed to be represented by a normal, exogenous demand curve; consumers choose utilization levels of professional services in response to market prices, and their own tastes and incomes. Total output/sales of professional services would then be inversely related to their prices, and all the agency issues of chapter 4 above are assumed away. Similarly, the individual firm faces an externally set demand for its services; the role of the professional in advising patients on utilization is assumed to be either non-existent or completely non-discretionary.

The nature of this externally set demand, however, varies, depending on the market structure assumed for the professional industry. At one extreme, firms might be assumed to be perfectly competitive price-takers, able to sell as much or as little as they choose at a given market price, but unable to affect that price. The restriction on entry would then ensure that this competitive market price was high enough to generate "supra-normal" profits or professional

incomes, *i.e.*, incomes exceeding those available in non-professional occupations requiring similar training, skills, and effort. In a "normal" market, the existence of such supra-normal profits is an incentive to draw additional suppliers into the market, as well as a signal as to social, or at least buyers', priorities. But entry restriction blocks this process.

Behind the protective wall, however, the professional services market need not be perfectly competitive. More plausibly, locational factors and patients' perceptions of practitioner or firm characteristics create distinctions among firms so that each has some discretion over price setting. In such a "monopolistically competitive" environment, the volume of services each firm can sell varies inversely with its own price. For any given own-price, volume varies directly with the prices charged by others and inversely with the numbers of such competitors. By assumption, however, the firm cannot affect its own sales, except through the price it charges.

A variant on this approach, of some practical interest in medical care, is that in which the firm is assumed to be able to discriminate in price-setting, charging different prices to different patients, and perhaps even charging the same patient different unit prices for different amounts of care. In this environment the notion of "a" price for care becomes rather hazy; charges are set to yield whatever the traffic will bear, but the professional/firm's ability to discriminate in pricing will depend critically on the degree of competition it faces from other firms.

Finally, the most extreme form of cartel behaviour involves not only restrictions on entry and possibly scale of firm, but complete collusion, so that the entire group of firms acts like a single monopoly in its price and output-setting behaviour. Such a cartel, to be fully effective, would have to allocate output among its members (presumably by setting a structure of differential prices) and arrange for profit-sharing. A more limited form would merely involve agreement on a single pricing structure (a schedule of minimum fees) with some mutual understanding about appropriate patterns of price discrimination.

PROFESSIONAL EARNINGS: REWARDS OF MONOPOLY?

This simplest concept of a profession as a "conspiracy against the public" to enhance its members' incomes by staking out and enforcing exclusive rights to a market, and limiting competition among its members by regulation or collusion, clearly captures some important features of the industry. We do observe that average incomes in the self-regulating professions, and particularly the senior health professions, medicine and dentistry, are consistently at the top of the occupational lists. Table 7-4 shows average net incomes by year over the past thirty-five years for physicians, dentists, and other "representative" licensed professionals, relative to the average weekly wage, and indicates that while large swings have occurred over time, the top position of the health professions has not been challenged.⁵ Indeed, Table 7-4 suggests that the income position of the leading non-health professionals has eroded somewhat, through time, while physicians rode a roller-coaster up to extraordinary (relative) heights in 1971, followed by a sharp readjustment toward earlier levels. The extent of this adjustment, however, is difficult to determine in light of the significant discrepancies between incomes reported in tax data and those estimated by Health and Welfare Canada. The 1982 HWC estimate for average Canadian physician net incomes is \$97,000, or 4.97 times the annualized average weekly wage.

TABLE 7-4
Earnings of Selected Self-Employed Professionals,
Relative to Average Weekly Wages and Salaries, 1946-1981

	Physicians (NHW Est.) ¹	Physicians	Dentists	Lawyers	Accountants	Architects / Engineers	Average Weekly Wage x 50
1946	--	7466	5289	6529	--	5984	1624
	--	(4.60)	(3.26)	(4.02)	--	(3.68)	--
1951	--	9975	6287	10214	8171	9628	2502
	--	(3.99)	(2.51)	(4.08)	(3.27)	(3.85)	--
1956	--	13053	9230	12617	9940	13640	3222
	--	(4.05)	(2.86)	(3.92)	(3.09)	(4.23)	--
1961	--	17006	12337	15718	11627	14692	3912
	--	(4.35)	(3.15)	(4.02)	(2.97)	(3.76)	--
1966	--	24993	17212	21045	13946	21200	4817
	--	(5.19)	(3.57)	(4.37)	(2.90)	(4.40)	--
1971	--	39555	25828	27862	18631	21648	6882
	--	(5.75)	(3.75)	(4.05)	(2.71)	(3.15)	--
1976	51800	49310	43336	44858	36616	40626	11402
	(4.54)	(4.32)	(3.80)	(3.93)	(3.21)	(3.56)	--
1980	72100	63411	56977	49481	43799	41052	15869
	(4.54)	(4.00)	(3.59)	(3.12)	(2.76)	(2.59)	--
1981p	84000	68604	62497	56798	42113	43804	17736
	(4.74)	(3.87)	(3.52)	(3.20)	(2.37)	(2.47)	--

SOURCE: See Data Sources Appendix.

NOTE: Bracketed figures are incomes relative to Average Weekly Wage x 50.

¹ Unpublished data from Health and Welfare Canada show a growing discrepancy between physicians' incomes reported in taxation statistics, and provincial insurance plan payments. These estimates have been prepared internally by HWC, and exclude Quebec.

Dentists, however, have shown a steady relative advance (discounting the rather questionable 1946 figure) and have narrowed the gap between themselves and physicians since 1971. Indeed, allowing for workload differences and possible tax advantages, dentists may now be as well or better paid, on a per hour basis at least. But the most recent HWC estimates show physicians pulling away again.

Of course, occupations with long training periods and long hours require higher average earnings to compensate. But most studies of earnings in the health professions indicate over-compensation.⁶ Moreover, we find unsatisfied demand for entry to the medical and dental professions in particular, in the form of queues of applicants for limited numbers of training places, and apparently qualified candidates being turned away.⁷ And it is widely observed that professional associations *do* restrict the economic behaviour of their members, discouraging

price competition, forbidding advertising, and placing specific restrictions on practice organization, in all the ways the "monopoly" model would predict.

Moreover, the tradition of sliding-scale billing, or tailoring charges to the patient's resources, either directly or through varying efforts at collection, is readily reinterpreted as price discrimination or charging what the traffic will bear. The provider who attempts to ensure that no one will be denied service through inability to pay may also be maintaining a high level of "sales" by setting fees just below each patient's "reservation price," the level at which the patient would forego care (or go elsewhere) (Kessel 1958).

In the case of physicians' services, of course, the existence of public insurance plans negotiating uniform province-wide reimbursement schedules is the major (in some provinces the only) factor determining fee levels. But it has been the declared policy of several provincial medical associations to encourage their members to bill patients directly above this level. And while upholding in principle the freedom of the practitioner to set her own fees,⁸ medical associations in extra-billing provinces continue to issue fee guides to co-ordinate billing behaviour and to discourage competitive pricing. This sort of co-ordination and guidance can be remarkably effective; internal surveys of the British Columbia College of Dental Surgeons (which doubles as the British Columbia Dental Association, combining the public statutory and the private trade association functions) indicate that over 80 percent of its members follow the (purely voluntary) provincial fee guide unilaterally promulgated by the College).

The effectiveness of such co-ordination, however, depends on the security of a professional group's control over its exclusive domain. Thus the existence of self-governing professions with overlapping claims of competence represents a constant threat of the outbreak of competitive behaviour which cannot be controlled within the professional framework. Friction and bitterness is common on such interprofessional boundaries. The general response of senior professions (physicians, dentists, lawyers, chartered accountants) to "professional insurgency" seems to alternate between denying the legitimacy of competitive professions or the competence of their members, and trying to draw current members into the "professional team," while suppressing new entry to the competitive profession -- exactly as the profession-as-monopoly model would predict (Evans and Stanbury 1981).

The implications of the simple "monopoly model" for economic analysis and public policy are rather straightforward. Professional organizations lead to underprovision of professionalized services -- prices are "too high," quantities are "too low." The opportunity cost of the marginal resources (not) used in producing professional services, their value in alternative activities, is below the value which users would have placed on the professional services foregone. The result is both an allocative distortion -- society's productive resources are not being devoted to the production of things people value most -- and a transfer of wealth (as a result of the elevated prices) from service users (or tax and/or premium payers, if the service is collectively funded) to professionals. The appropriate policy response is to lower entry barriers and expand supply by more of the same or of competing types of professionals. This will lead, on the assumptions of the monopoly model, to more output and a competitive bidding down of prices so that the additional output can be sold. The recent analysis of professional incomes by Muzondo and Pazderka (1979) follows this approach.

If insurance is widespread (dentistry) or universal (medicine), things become a bit more complex; the monopoly presupposes a substantial degree of consumer/patient price responsiveness. In its pure form, however, it encounters no difficulty from insurance. Competition among suppliers could take the form of cash rebates to their customers. If, on the other hand, prices are effectively fixed, more entry will drive down work loads and incomes until "supra-normal" profits are eliminated; whether this represents excess capacity will depend on

where the fixed price is set. Advocates of a competitive market and of the curbing or dissolution of professional "monopoly" powers usually recommend policies to reduce insurance coverage as well, and minimization or elimination of public insurance.⁹

INADEQUACIES OF THE MONOPOLY MODEL

The "professional monopoly" viewpoint, however, is incomplete or misleading in several respects. Most obviously, it assumes that professional associations control entry directly, which is of course untrue (at least in Canada and for the health professions). Provincial governments, acting through universities and other post-secondary institutions, ultimately determine the numbers of training places available. Professional associations may attempt to influence such decisions, but as the decision in the late 1970s to expand the University of British Columbia medical school showed, they may fail. On the other hand, in most provinces most of the time (Quebec being an apparent exception) professional associations have exercised very powerful influence over their own governing legislation, and are able to make regulations under it subject only to scrutiny of varying effect by the Lieutenant-Governor-in-Council. The control over conduct appears more effective than that over professional numbers.

At a more fundamental level, by assuming both the existence and the policy relevance of an independent consumer demand curve, and by assuming away the agency role of the professional, the monopoly model in effect ignores the basic social (as opposed to private or conspiratorial) justification for the *existence* of self-regulating professions. If consumer/patient/clients are assumed to make their own decisions about how much care to use, responding to out-of-pocket prices and consulting their own incomes and "tastes" but not responding to the advice (as opposed to prices) of professionals, and further, if by the assumption of consumer sovereignty, we accept such decisions (however well or badly informed) as the choices which *ought* to govern production and distribution, then it is not surprising that the result is a policy recommendation to dismantle the professions, or at least to reduce significantly their control over the economic conduct of their members.

In this framework, the quantity and mix of services which consumer/patients choose to take off the market is always the "right" one for them, at the current level and pattern of prices which they face. Underuse, or inappropriate use, is defined only in terms of a faulty price structure -- "too high" as a result of restricted entry and/or collusive behaviour, just as in chapter 2 it was "too low" as a result of insurance. The issues of effectiveness (or harm) and the relation between care and health status or other ultimate objectives are simply ignored. The customer is always right, or else is, on average, closer to right than anyone else can be on her behalf.

In a number of currently "professionalized" markets, these full (or at least best available) information and consumer sovereignty assumptions, and their associated policy implications, do seem to be defensible. But in health care, particularly in its more complex and life/limb threatening aspects, they are highly dubious, if not silly. In any case, intellectual consistency requires either that consumers *are* sufficiently informed to, and do, make their own decisions, or that they are not. If they are, and do, then the exogenous demand curve is a valid tool of analysis, but the agency role of the professional is superfluous. If they do not, then the professional has a role as analysed in chapter 4 (which may nevertheless be well or badly performed), but the exogenous demand curve cannot serve either as a description of utilization behaviour (in response *inter alia* to prices) or as a standard for what levels or mixes *ought* to be. One cannot logically have it both ways. (Not that that prevents many analysts from doing so.)

The monopoly model also adopts the conventional assumption that the professional "firm" is a profit-maximizing entity, responding either to fixed input and output prices or to input and output supply and demand schedules. Once past the entry barriers, and except for specific regulations (and perhaps unofficial provincial constraints and threats), the professional firm attempts to minimize its costs and maximize its profits like any other private business. This assumption, in turn, makes the firm's short-run output decisions determinate, matching marginal revenue to marginal cost across each class of output. In the middle and longer run, decisions over use of capital and other labour inputs, choice of specialty, and location and migration decisions are all governed by the same profit-maximizing objective. The firm will always choose the minimum cost (subject to regulatory constraint) way of producing whatever its consumer/patients wish to buy (as represented by the assumed exogenous demand curve) and will go into whatever line or location of production which consumer willingness to pay dictates. Thus the model provides simple answers to a wide class of policy questions. Technical efficiency is, by assumption, always assured, and problems of location and specialization, if they exist, are easily dealt with through manipulation of prices or subsidies. Whatever behaviour is desired, which the market does not automatically yield, can be achieved by the creation of profit opportunities.¹⁰

Again, the model has some practical relevance. It seems clear, for example, that problems of undersupply of services in remote regions *can* be mitigated by "flooding the market" with providers or by fee differentials or subsidies such that incomes are higher in less than in more desirable regions. People do respond to income incentives. But it is a substantial step from there to the assumption that the professional firm "maximizes profits." The unrealism of that assumption emerges at both empirical and theoretical levels.

SYSTEMATIC DEPARTURES FROM COST-MINIMIZATION: THE (NON) USE OF AUXILIARIES

The empirical evidence bearing on profit maximization, which in turn implies (but is not implied by) cost minimization, is of several forms. The most extensively studied area relates to cost-minimization, in particular to the use of auxiliary personnel in medical, dental, and pharmaceutical practice. The comparison of group with solo (and, where permitted, chain or corporate) practice, and the question of the optimal scale and organization of practices, also open the issue of what motivates practitioners and the practices they control. Similarly, concerns over the distribution of practice capacity (usually expressed in terms of distribution of practitioners) across regions or specialties arise from the observation that factors other than income/profit possibilities strongly influence such career choices.

For at least twenty years, there has been a growing interest in the possibility of using "intermediate-level health practitioners," less neutrally described as paraprofessionals, or ancillaries, to perform a number of the functions currently performed by physicians, dentists, or pharmacists. An extensive literature, based in part on experimental and field experience, has developed which demonstrates the technical feasibility of such substitution over a wide range of functions, while maintaining quality standards equal to or better than those achieved by the "peak professionals."¹¹ And the problem of diagnosis of abnormality, of identifying those problems which have gone beyond the skills of the intermediate-level practitioner, turns out in practice not to be a difficulty. Diagnosing the existence of abnormality, of the unusual or complex problem,

is of a lower order of difficulty from determining the nature of, and appropriate response to, the complexity itself.

The monopoly model predicts, of course, that established professions will attempt to suppress the entry of new types of professional "firms" into the service marketplace. The dentist/dental mechanic, the optometrist, the chiropractor, and the clinical psychologist are all selling services which compete directly with some of those offered by physicians or dentists, while self-employed dental hygienists, midwives, nurse practitioners, or dental nurses¹² would dramatically widen the range of contested territory. The pressure by other types of therapists, acupuncturists, for example, represents the same threat -- more firms of diverse types sharing a market already overcrowded with physicians, dentists, and pharmacists, and practitioners with diverse professional backgrounds and cultures such that economic co-ordination would be more difficult. It is not therefore surprising that the entry and/or spread of such competitors has been bitterly contested, either by attacks on their professional legitimacy and competence or, where possible, by direct regulation, without any serious attempt at justification (as opposed to bald assertion) in terms of patient protection. The relatively slow growth, compared with their potential, in numbers and capacity of professional firms run by/employing substitutes for physicians contrasts sharply with the rapid expansion of occupations, particularly in hospitals, which are complementary to physicians in the sense of increasing their ability to serve (and bill for) patients.

What the monopoly model does not predict, however, is the failure of firms run by peak professionals to take up the opportunity to *hire* such personnel. Practitioner time is the principal cost of production of the NOFP firm. The professional owner(s) may hire other professionals as well as other employees, in which case the compensation of the hired professional shows up directly as a cost of production. But the self-employed individual must also be credited with an implicit wage. Profit in the economic (as opposed to accounting) sense is the net revenue (positive or negative) remaining after such deduction. Accordingly, the substitution of less costly auxiliary time for expensive own-time will lower the professional firm's cost of production and raise its profits. A truly profit-maximizing professional sector would still resist with all weapons available the entry of new independent firms as competitors, but existing firms would be willing to hire any and all auxiliaries so long as the resulting additions to total billings (marginal revenue product per employee) exceeded the increases in wage and overhead costs (marginal outlay per employee) including the implicit costs of supervision. Yet the evidence is extensive that this does not happen.

One possible explanation is external constraint. Under the Canadian Medicare program, for example, practitioners are not permitted to bill for services provided by employees, but only for their own services. Parts of a task may be delegated to assistants, but not the task as a whole. Clearly this discourages the use of more highly qualified assistants, such as nurse practitioners or midwives, who are specifically trained to perform entire services. And the measured productivity gains from task sharing are much less than those from task delegation.

If Medicare regulations were the explanation, however, the problem of underuse of auxiliaries would be confined to Canadian physicians. Yet most of the studies identifying underuse are American, and the problem appears at least as serious in dentistry (on both sides of the border), where no universal plan applies. The phenomenon is far too general to be explained by the payment rules of a particular insurance program.

Self-regulation also imposes constraints; dentists, for example, may not legally employ dental nurses even if they wish to. And physicians who go beyond their colleagues' norms in task delegation may be much more vulnerable to malpractice charges, regardless of whether any unfortunate outcome was in fact linked to delegation. Such collective restraint of the individual

can be interpreted in the cartel form of the monopoly model -- practitioners collectively recognize that if as individuals they begin to expand capacity and output, the resulting expansion in total supply will drive down price, while the new auxiliaries add to costs.

But collective constraints, while clearly significant in preventing the deployment of certain types of personnel, are also an incomplete explanation. Underuse is not restricted to regulatory blockage of specific occupations such as midwife or dental nurse; even legally permitted forms of auxiliaries are utilized well short of the cost-minimizing, profit-maximizing level by individual NOFP firms.

ADDITIONAL, OR ALTERNATIVE, OBJECTIVES FOR THE FIRM: BEHAVIOURAL AND POLICY IMPLICATIONS

The available explanations seem to be twofold, and both require us to modify and extend the simple monopoly model.¹³ One is that practitioner/owners have "tastes" for particular practice styles: they enjoy doing things themselves, for example, and dislike directing and administering teams of other people (Reinhardt 1972). And the umbrella of protection against competition, created by restrictions on entry and collective self-regulation of economic conduct, enables practitioners to indulge these tastes without giving up an unacceptable amount of income. The objectives of the firm must then be extended to include not only "profit" but attainment of the practitioner's preferred practice style. The practitioner's utility or level of welfare depends on both. The NOFP firm is now a utility-maximizer; profit being only one contributor to the practitioner's well-being.

The second explanation points in the same direction. The root weakness of the profit-maximizing assumption is that it neglects the identity between the practitioner as practice owner and residual claimant to net revenues, and the practitioner as supplier of labour and skills to the practice (Evans and Williamson 1978; Evans 1980). The major component of cost to the practice, the practitioner's implicit wage, is also the major component of the practitioner's earnings. A "cost-reducing" substitution of auxiliary for professional labour input will in fact lower the practitioner's income; profits rise but implicit wage falls, unless the total practice output can be expanded to maintain the level of employment of practitioner time. Accordingly, we find auxiliaries more extensively employed where professional/population ratios are low, and practitioners feel "overworked" (a situation which is inexplicable for a profit-maximizing firm, at least in equilibrium -- it just raises prices). Conversely, as professional practitioner to population ratios rise, collective hostility to intermediate-level practitioners increases, and their employment falls.

The bias of professional practices towards overuse of the "high-priced help," and underuse of less expensive auxiliaries, is thus explicable if we shift our focus from the profits of the firm to a more plausible objective, the net income of the firm's owner(s). But net income maximization alone is an implausible objective; formally, such an objective predicts twenty-four-hour work-days, 365 (or 366) days per year. Instead, we postulate maximization of a more general set of objectives which include both net income and leisure time, as well as elements of practice style -- hence the NOFP designation. The firm's behaviour will only respond to profit opportunities which pay off in terms of some combination of net income and reduced or more professionally satisfying work load, subject also to the practitioner's preferences for particular working environments or practice styles.

This extended view of the NOFP firm retains the assumption that service demand is exogenous and price dependent, that firms individually and collectively are constrained by a demand relationship which uniquely (and inversely) links the volume of services they can sell with the price the consumer/patient/client must pay for them. It is thus, like the monopoly model, unable to explain the existence of self-regulation as other than a conspiracy against the public interest to constrain supply and elevate price. What it does do, however, is to open more dimensions of firm behaviour in a way which seems more closely to approximate reality. It requires us to re-interpret the "costs" of self-regulation and professional organization.

In the monopoly model, these "costs" were simply allocative distortions, too low total output from the NOFP sector and too high relative prices. Society as a whole would be better off, according to this model, if more resources were diverted from other sectors into expanding the NOFP sector. In addition to the allocative distortion, wealth is transferred to firm or resource owners in the NOFP sector in the form of higher prices and incomes, and away from those who pay for such services.¹⁴

The extended view of NOFP firms' objectives, however, shifts attention away from allocative distortions and towards the issue of the technical efficiency with which such firms use resources. By unnecessarily using the time and skills of the most costly practitioners, and failing to use equally effective, less costly substitutes, NOFP firms appear systematically to waste resources. And this form of waste is on a large scale. Estimates by various authors for several of the NOFP sectors in Canada and the United States suggest that reductions of as much as 40 percent of total expenditures on dental services, pharmaceutical dispensing, and ophthalmic goods could be achieved through rationalization of production and use of less expensive personnel, without loss of quality (Benham 1972; Benham and Benham 1975; Evans and Williamson 1978). Similar savings estimates for medical services are harder to develop, but in Canada are very conservatively put in the neighbourhood of 10-15 percent for all medical care and 16-24 percent for ambulatory care (Denton *et al.* 1982). For ambulatory primary care in the United States they run into billions of dollars (Record 1981). The most extreme estimated costs of allocative distortions, under the assumption that production of services is technically efficient, are by contrast almost certainly less than 10 percent, and quite likely below 5 percent.¹⁵ And the simple monopoly model's policy "solution" to allocative distortions -- increase the supply of professionals -- is likely to exacerbate the more serious technical inefficiencies.

EVIDENCE FROM GROUP PRACTICES: TECHNICAL EFFICIENCY

The issue of the technical efficiency of the NOFP sector goes beyond the use of auxiliaries as substitutes for peak professionals. The discussion of group versus solo practice, which has gone on for at least fifty years, has as one of its themes the issue of economies of scale and technical efficiency of production. It has been argued by proponents that group practices are able to make better use of "lumpy" forms of capital equipment or of auxiliaries for whom a solo practitioner could not provide full-time employment, and generally to gain from efficiencies in the scheduling and co-ordination of personnel and tasks in serving patients with multiple problems and needs. As evidence, one might observe higher rates of service output and of net income per practitioner in group than in solo practice. Yet if the profit-maximizing, cost-minimizing model of firm behaviour were correct, such discussion would be beside the point. Firms would adopt whatever structure they found to be cost-minimizing, and the persistence of both types of firms would indicate that each had a market niche, somehow defined. Promotion of

one style or another rests on an implicit assumption that practitioner-owners of firms are either uninformed about opportunities for cost-minimization (implausible, in a world of rational profit-maximizers) or have tastes and perceptions about styles of practice which influence their behaviour but which could be modified. The latter view, of professionals with perceptions and tastes for particular practice styles which strongly influence practice mode independently of income opportunities, seems both a priori, and on the evidence, the more plausible.

But the focus of the group practice debate has shifted significantly over the last three decades. The earlier arguments for grouping per se, that taking a number of fee-for-service solo practitioners and combining them in a fee-for-service group would somehow increase efficiency, have largely been abandoned. The apparent income and output advantages of professionals in such groups have been shown to be largely a result of differences in product line rather than choice of production technique (Bailey 1970a, 1970b). The technology of particular medical specialties, such as diagnostic radiology and pathology, lends itself to large groupings with high auxiliary and equipment use per professional. But to compare such groups with solo or small group general or family practitioners, and to assume that large groups of primary care specialists would display the same behaviour as radiologists, is clearly illegitimate. Moreover, when primary care practitioners *do* form large groups, apparent increases in output per practitioner may result, not from greater efficiency in primary care, but from the establishment in-house of diagnostic services which were previously refer-red out. From a system-wide perspective, solo practitioners with a well-functioning referral network may be just as efficient as a formal group. Those functions which do display significant scale economies become specialized services available to several practitioners.

Significant advantages in efficiency have been consistently (though not universally) demonstrated for groups which not only assemble a number of practitioners but which are reimbursed on a capitation rather than a fee-for-service basis (Enthoven 1980; Luft 1981). Variously labelled as Community Health Centres or Health Services Organizations (HSOs) in Canada, Prepaid Group Practices, Closed-Panel Practices, or Health Maintenance Organizations (HMOs) in the United States, all such groups provide a combination of both professional and risk-bearing services. They serve a defined population group of enrollees, and their revenues are a fixed dollar amount per enrollee per time period. Enrollees are then provided with all "needed" services, as judged by the group's practitioners, either from the group itself or from external referrals at the group's expense. In the United States, an enrollee who self-refers for services outside the group pays their cost out of pocket; in Canada, the Medicare program relieves the patient of any personal liability. Depending on provincial arrangements, however, the cost of such services may be charged back to the group's budget.

Such organizations have achieved improvements in efficiency in two main ways. First, they have made significantly greater use of intermediate-level practitioners as substitutes for physicians than has the private practice, fee-for-service system. This observation tends to support the utility- rather than the revenue- or net income-maximizing model of the NOFP firm. It is of course true that a group which, when adding or replacing a professional, chooses a salaried nurse practitioner rather than a physician will, if revenues are unaffected, raise the surplus available for distribution among group members. But this is equally true for a fee-for-service group, if reimbursement agencies permit billing on behalf of auxiliaries. Differences in patterns of auxiliary use indicate that the shift in payment mode is associated with changes in management structure, such that practitioner preferences receive less weight. Or practitioners in such groups may be differentially self-selected and socialized. In any case, the evidence suggests that more efficient use of auxiliaries results from change in the reimbursement and management structure,

the pattern of incentives and control, not from grouping and scale per se. Pure scale economics in medical practice, as opposed to more auxiliary substitution, do not seem to be pronounced.

PROVIDER INFLUENCE OVER USE: PREPAID GROUPS AND THE EXERCISE OF AGENCY

Secondly, and most important, the product mix of such prepaid groups differs substantially from fee-for-service practice. Numerous studies of different groups under different circumstances have shown reductions in hospital use of up to 40 percent for group enrollees, compared with similar patients in fee-for-service practices (Luft 1981). These results have accumulated over thirty years, and form one of the principal supports for present United States health policies of trying to contain health care costs by promoting the spread of HMOs. Similar findings in Canada have helped to keep alive an interest in alternatives to traditional self-employed fee-for-service medical practice, although the integration of such alternative forms of practice into a universal public insurance system is much more difficult (Vayda 1977; Luft *et al.* 1980; Stoddart and Seldon 1983),

For understanding the behaviour of NOFP firms and their practitioner-owners, the significance of such variations in output mix is twofold. First, it demonstrates the extent to which practitioners influence the demand for their own services, and thus supports the analysis of agency in chapter 4. The assumption that health care users display an exogenous "demand" for health care, using more when the prices they face are lower and conversely, requires that a difference in mix be associated with a difference in relative prices. If people enrolled in HMOs use less hospital care than patients of fee-for-service practitioners, then, according to that model, it must be because they face higher prices for hospital care, either absolutely or relative to ambulatory care. But significant differences in use are found in settings where patients are fully insured for either form of care. The obvious explanation, which is also that given by practitioners themselves, seems to be the correct one. Practitioners in capitation-based groups adopt, for economic and/or professional reasons, styles of practice which involve admitting patients to hospital less frequently and for shorter stays. The practitioners directly and powerfully influence the level of services which patients receive, independently of the prices which users do or do not pay.

ALTERNATIVES TO THE EXOGENOUS DEMAND CURVE: PROFESSIONAL ETHICS?

Such an inference may seem blindingly obvious, and is generally accepted by students of health care. But to a significant number of economists, principally but not entirely working in the United States, it has been unacceptable. Such intellectual resistance can be readily understood in terms of the fundamental role which the assumption of exogenous demand plays in "closing" the models above, and permitting the construction of consistent, if not necessarily accurate or enlightening, explanations of NOFP firm behaviour. Economists have a professional predilection for explaining individual or organizational behaviour as resulting from attempts to maximize or optimize with respect to some desired objective, subject to external constraints, and the external demand curve is a critical constraint. If practitioners can generate demand directly, why not do

so indefinitely, and then raise prices to choke off that demand, thus achieving higher incomes at the same work load? What constrains utilization, if *not* the external demand curve?

An obvious answer is, professional ethics. Practitioners' concepts of "practice style" include views on how medicine, dentistry, etc., *ought* to be practiced, and a large part of their training is intended to inculcate such values. Providing services which they believe to be unnecessary, or even harmful, is clearly a source of professional dissatisfaction, as is imposing substantial economic burdens on their patients. The objectives of the NOFP firm must then be extended to include, not just practitioner net income, work load, and practice setting, but also the degree to which practice activity conforms to the professional, ethical principles of the individual practitioner (Evans 1972, 1974, 1976; Wolfson 1976; Reinhardt 1978; Wolfson and Tuohy 1980). Removing healthy organs, or drilling healthy teeth, has a negative impact on the practitioner's overall satisfaction, even if it is profitable and the patients, believing the organs/teeth were diseased, are satisfied. Professional ethics bound and constrain the demand-generation process.

But this approach is weakened by the second implication of hospital utilization differentials between capitation-based groups and fee-for-service practitioners. There is no observable difference in patient health or well-being between high-use and low-use groups. The inference is that the additional use is not "needed" in a technical, medical sense; such an inference obviously underlies all policies aimed at expanding the role of such groups. In terms of Figure 1-3, it appears that hospital use in aggregate has reached the flat of the health status curve, if not actually a downward sloped segment. The same inference can be drawn from observations of wide variations across regions in surgical rates, patterns of diagnostic testing, and hospital use within the self-employed fee-for-service system. Again, there is no evidence of impaired health status in low-use regions, and indeed in some cases there are grounds for believing that populations in high-use regions suffer as a result (Roos and Roos 1981). Thus if practitioners' ethics and preferences for practice style constrain them in influencing utilization, these preferences do not seem always to be linked to objective information on need. They may, of course, be linked to subjective assessments, and probably are. Most practitioners undoubtedly believe that the services they recommend and provide are genuinely needed in the sense of contributing to health status. Since the subjective views of different practitioners lead to such widely divergent (and expensive) results, however, the source of such views becomes a critical issue.

ARE ETHICS EXOGENOUS?

Hence the argument, which has considerable support from both expert opinion and research, that the appropriate response to problems of overservicing and unnecessary use is better information and education for practitioners. The focus here is on overuse relative to objective criteria of need, not relative to willingness to pay; and in terms of modelling the NOFP firm the assumption is that the non-economic dimensions of the firm's objectives, the "professional" aspects of ethics and response to needs, are powerful enough motivators that better information can lead to better performance.

In contrast, a "structuralist" view focusses on the pattern of economic incentives implicit in fee-for-service practice. From this perspective, the educational approach involves a fallacy of composition. A convincing demonstration of the inefficacy of a procedure will eventually lead to its abandonment (although perhaps very slowly) as tonsillectomy indicates. (But circumcision?)

But the overall level of activity will not be affected; NOFP firms will simply shift time and resources to other sectors. Since the objectives of the firm include economic as well as non-economic components, it will be prepared to trade off dollars against professional pride, or put another way, if income and work load shrink (say because physician-to-population ratios are rising), the practitioner will change the criteria which govern her recommendations to patients, and will generate more demand to keep the patient at the door and the wolf from it. The consistent response of utilization to numbers of practitioners, regardless of price levels, all over North America, is indicative of this trade-off in action. Moreover, standards of practice adapt over time in response to current practices, one's own and others. Thus a shift to recommending more servicing, which may initially be a response to economic factors, becomes over time the new standard.

THE ROLE OF FEES WHEN DEMAND IS ENDOGENOUS

In this process, price plays a role not so much in clearing markets, equating separately determined supplies and demands, but as an alternative instrument for trying to reach practitioner income objectives. The often-observed *positive* correlation of physician-to-population ratios with fee levels (though not incomes) is consistent with physicians reacting to the pressures of oversupply and falling incomes by pushing up prices as well as promoting utilization.

In the Canadian context, however, the fee schedule structure is critical. Many of the opportunities for increasing utilization without working longer hours have been removed. Practitioners are not paid for work done by auxiliaries or (except to a limited extent, by prior agreement, and varying by province) for privately provided diagnostic services. Further, the fee schedules provide limited opportunities for procedural reclassification. Thus, the very restrictive fee increases of the early and mid-1970s in Canada led to significant drops in practitioner real incomes, as shown in Table 7-4. They were offset to a limited degree by increased output per practitioner, but generating additional utilization imposed more time and energy costs on the practitioner. On the other hand, the expansion of practitioner numbers translated directly into increased utilization, as new NOFP firms built up "markets" for themselves. Table 7-5 displays price and utilization patterns during the 1970s, indicating the relatively steady 1-2 percent annual increase in billings per physician (adjusted for fee increases) over this period. Substantial increases in utilization are associated with increases in the physician supply. Since 1975, curtailment of physician immigration has reduced this growth somewhat, but physician-to-population ratios continue to climb.

TABLE 7-5
Selected Measures of Physicians' Services,
Expenditures, and Use
Annual Rates of Increase, 1971-72 to 1980-81

	1971-72 1974-75 %	1974-75 1979-80 %	1979-80 1980-81 %	1971-72 1980-81 %
Aggregate Fee Payments	10.1	11.7	15.9	11.6
Insured Population	1.4	1.1	1.3	1.2
Price	2.3	6.8	10.7	5.7
Utilization	7.6	4.6	4.7	5.6
Utilization Per Capita	6.2	3.5	3.4	4.4
Physician Supply	5.5	3.1	3.1	3.9
Output Per Physician	2.2	1.4	1.6	1.7
Consumer Price Index (All Items)	7.7	8.9	10.1	8.6

SOURCES: Canada, Health and Welfare Canada (1981, 1983a).

The central significance of the organization and payment of professionals, and of physicians in particular, for utilization patterns is also illustrated by the experience of other countries. Sweden, for example, like Canada, experienced a dramatic increase in physician supply during the 1970s, 5.6 percent per year from 1970 to 1978. Yet the number of physician visits rose only 1.5 percent per year, and visits per physician fell 3.8 percent annually, or 27.2 percent over the period (Jönsson 1981). Sweden's physicians are salaried; when the population per physician drops, so apparently does the work load. Canada's fee-for-service physicians provide more services per patient. In the United States, an intermediate response seems to have occurred. Visit rates per capita do not appear to have responded to the increase in physician supply (which may reflect the fact that Americans pay a significant proportion of physician costs out of pocket), but the intensity of servicing per visit, billings per visit adjusted for fee changes (which the physician controls) has risen steadily.¹⁶ In none of these settings does the independent decision-making of the patient appear to play a significant role.

We have moved a long way from the simple "monopoly model" of professional firms, gaining in richness and realism at the expense of complexity and increasing uncertainty of predictions. The monopoly model addressed only a single issue, the role of professional entry restrictions/conduct control in elevating service prices above, and depressing outputs below, their socially optimal levels. Profit-maximizing firms were assumed to ensure that production was fully efficient and that regional, specialty, or other types of sub-markets were all served equally adequately or inadequately, relative to consumer willingness to pay, while informed consumers ensured that whatever was produced and used was appropriate. The only policy issues in such a world concern who, if anyone, should impose restraints on entry and conduct, and on what criteria? Is any regulation necessary? If so what kind, and by whom? Expansion of entry and limitation of self-regulatory power are anticipated to lead to increased output and lower prices.

While identifying a part of the elephant, the monopoly model is unable to enlighten or even to express a number of aspects of the real world or of the concerns of health policy. Opening up

the supply side of the model, by abandoning the profit-maximizing assumption, reveals serious concerns about the technical efficiency of professional service production and about the responsiveness of professional firms to the needs/demands of various sub-markets. From this broader perspective, it appears that professionalization generates unnecessarily high costs of production, and that these excess costs far outweigh those of the allocative distortions created by monopoly *per se*. Moreover, such excess costs appear to be *increased* by increases in the supply of professionals, so long as legal restraints permit only professionals to own/manage professional services firms. More physicians/dentists trained leads to less use of auxiliaries, and higher costs of production.

In this environment, the appropriate response may be to open up the licensure process in one of two ways. If restrictions on practice ownership were removed, so that non-professionals, including public corporations, could own and manage practices (though not performing the activities and functions of professionals), then the economic disincentive to auxiliary use would disappear. Such a policy would operate in conjunction with licensure of more intermediate-level professionals themselves (midwives or dental nurses, *e.g.*) to lower the cost of providing professional services. Alternatively, one could adopt a policy of licensing more intermediate-level professionals to compete directly for parts of the professional market, as dental mechanics or optometrists now do.¹⁷ The evidence is more than ample that such practitioners in either setting are capable of performing a significant range of additional functions. Professions do indeed function in part as monopolies, restricting supply and elevating price, but the simple response of training more peak professionals is likely to do more harm than good.

The influence of the preferences of professionals-as-persons on the behaviour of professional-practices-as-firms also opens the issue of service distribution and the adequacy of supply in sub-markets. Profit-maximizing firms go wherever effective demand is sufficient to cover costs. But people have preferences as to location and specialization, such that equal income opportunities may lead to very unequal distributions of service. In theory, a competitive market system would lead to the emergence of price differentials with lower prices and incomes for the same time and effort level in more desirable areas or specialties, to equalize net advantages. And as a rising tide lifts all the boats, expansion of supply should increase availability in both desirable and undesirable areas.

The latter effect does indeed seem to occur, but the differential prices, lower in desirable, overserved areas, do not emerge. This is perhaps unsurprising, since much professional activity is devoted to ensuring that markets for such services are not price-competitive. The implication of extending the firm's objective beyond profit-maximization, however, is that redistribution of services across regions or specialties, if desired on whatever social grounds, may be sought through other mechanisms than the price-income link. Price differentials are not the only, or even necessarily the most effective, way to redistribute capacity or personnel. Changes in the education system, either recruitment or training processes, may be more successful and/or cheaper. Yet ironically, the possibility of a systematic test of the role of price differentials in allocating professionals across regions or specialties is greater in a system of uniform negotiated fee schedules, as in the Canadian provinces, than among independently price-setting practices. One can ensure that the desired differentials are in fact established.

Once we abandon the demand-side assumption that the utilization of professional services is determined by informed consumers responding to price signals, however, the world becomes a good deal more complex. Changes in organization which promise to increase efficiency and competition in service supply, by introducing for-profit corporate ownership, *e.g.*, may also degrade the performance of the agency function. Efficient overservicing is not an obvious gain. Bailey (1977, 1979) argues that this has occurred in the United States commercial laboratory

industry, and similar developments may be underway in Canada. And the general presumption that "monopoly" leads to too little service at too high a price becomes misleading, irrelevant, or just plain wrong, if a public policy-induced oversupply of practitioners leads to both unnecessary servicing and inefficient production. Neither expanding professional throughput nor opening the industry to for-profit corporate enterprise is likely to improve the situation.

The model of the NOFP firm, from this more general perspective, suggests two distinct classes of policies (Evans 1980). One, a combination of research, education, and regulation, would seek to influence practitioner behaviour directly to improve the effectiveness of the services being provided. This would be combined with restrictions on the supply of new practitioners, plus removal of self-regulatory barriers to, and perhaps provision of direct incentives for, the use of more intermediate-level personnel. The second approach would involve attempting to redefine the nature of the services or products being provided, as the American HMOs do in offering packages of insurance plus "needed" services at a fixed price, rather than fees per service. For such redefined products, a combination of consumer information and/or regulation of product (as opposed to provider) characteristics may make a competitive market more effective. Both of these approaches will be dealt with in more detail below.

If the generalized model of the NOFP firm is realistic, however, any sort of policy response to the problems of appropriate servicing levels and efficiency in production will be made extremely difficult by the manpower policies of the past fifteen years. In the late 1950s and early 1960s, decisions were taken which substantially increased the supply of physicians in Canada. In part these were a response to population forecasts which turned out to be in error -- the Hall Commission forecast 28.2 million Canadians in 1981, about four million too high -- and in part they reflected some combination of perceptions of "unmet" need which public insurance would reveal, and of physician "shortages" indicated by rising (relative) prices and incomes. Increases in training places, plus immigration, led to an increase of about 50 percent in the physician-to-population ratio between 1962 and 1979, while output per physician was expanding as well (Barer and Evans 1983). This flood of new supply effectively foreclosed the opportunity to create and deploy less costly substitutes for physicians and led to the expansion of per capita utilization rates which has continued through the 1970s, as displayed in Figure 7.5.

In February 1975, immigration of physicians was cut back sharply in recognition of the problem of emerging oversupply. But present numbers of training places and attrition patterns, combined with current population forecasts, indicate a steady, though slower, rise in the physician-to-population ratio into the far future. On the other hand, the downtrend in death rates at the end of the 1970s and early 1980s could lead over the long run to both a larger and a significantly older population than that implied by population forecasts at the turn of the decade. Several studies (Boulet and Grenier 1978; Denton and Spencer 1983) have shown that if one assumes constant age-sex specific physician utilization rates, and calculates "needs" on the basis of population forecasts from the late 1970s, the aging of the population would not balance the growth of physician stock. But more recent data show relatively rapid increases in survival rates among the elderly, which may eventually justify current rates of production. In any case, one should not take for granted the appropriateness of current utilization patterns, and it remains true that such patterns could be provided by a very different mix of personnel.

Dentistry is experiencing similar dramatic increases in supply, an increase of 43 percent in active dentists per capita or over 3 percent per year, between 1969 and 1981 (Canada, Health and Welfare Canada 1980, 1983*b*). This increases the pressure on competitive auxiliary occupations and leads to efforts to promote prevention and cosmetics as ways of expanding servicing levels.¹⁸ The environment of most NOFP firms in health care for the foreseeable future thus appears to be

one of steadily expanding supply, and the implications for efficient resource allocation, given the present institutional framework, are not encouraging.

NOTES

¹ A comprehensive discussion of the physician income data is provided by Wolfson *et al.* (1980).

² The Census data are, of course, far from perfect. In particular, they are not full-time equivalents, a weakness which may be even more severe for practice employees than for practitioners themselves.

³ Whether the increase in billings corresponds to improvements in health status, or even the extent to which it represents increased servicing as opposed to re-labelling of existing services, is a difficult and important question. Here, we sidestep it.

⁴ Of course the process of intellectual evolution does not correspond to a strict temporal sequence; one can still find analyses addressing the monopoly aspect in the context of assumed exogenous demand and profit-maximizing behaviour.

⁵ The data, of course, *have* been challenged. In the course of physician fee negotiations, a great deal of nonsense has been talked about what physicians do or do not earn, involving confusion of full- and part-time practitioners, and some extraordinary manipulations of fringe benefit concepts. Wolfson *et al.* (1980) attempt to sort out the issues involved. It should be kept in mind that the data in Table 7-4, drawn from National Health and Welfare releases as of February 1983, are net incomes of *all* taxable self-employed practitioners, after expenses but before tax. They thus include part-timers, exclude salaried personnel, and exclude any expenses for tax purposes which re-emerge as income through other channels (as well as, of course, any undeclared income).

⁶ There is a line of argument to the effect that the effort and other costs of entry adjust to dissipate any monopoly gains. But the required informational/expectational assumptions are very specific, asymmetric, and quite implausible.

⁷ Of course the question of what constitutes appropriate qualification is eminently debatable.

⁸ Indeed, a direct attempt to regulate fee-setting behaviour by a medical association or College might be a criminal offence under federal anti-combines law.

⁹ Some analysts advocate roll-back and privatization of insurance *without* removal of the professional "monopoly," which seems, at the very least, intellectually inconsistent.

¹⁰ Some might argue that whatever the market does not yield, should not be desired; but that is an ideological input to economic analysis, not a conclusion from it.

¹¹ As examples only: Reinhardt 1973, 1975; Rafferty 1974; Spitzer 1978; Evans and Williamson 1978; Record 1981; Denton *et al.* 1982.

¹² A dental auxiliary with about twenty months post-secondary training, who can interpret X-rays and "drill and fill" teeth requiring plastic restorations, to quality standards at least as high as those of a general dentist.

¹³ An "explanation" sometimes offered by professional associations, that patient/consumers will not accept services from intermediate-level personnel, has been shown to be without foundation whenever it has been tested.

¹⁴ Wealth-transfer effects may or may not be viewed as a "cost," depending on one's distributional objectives.

¹⁵ Assume, for example, that monopoly power in an industry has advanced prices by 50 percent, so that removal of the distortion would drop prices by a third. Further assume unit elasticity of demand, so that the price decrease would increase quantity sold by 50 percent (total sales remain constant). The allocative burden as a percent of total expenditure would be $(\frac{1}{3})P(\frac{1}{2})Q(\frac{1}{2}) \div PQ = 8.3$ percent. For the health care field as a whole, unit elasticity and 50 percent price enhancement are extreme assumptions.

¹⁶ From 1950 to 1979, expenditures on physicians' services in the United States rose on average 9.7% per year, while the Consumer Price Index Physicians' Services component rose 5.22%, and population rose 1.37% (United States, Department of Health and Human Services 1980, Tables 64 and 69; Gibson *et al.* 1983). Thus physicians' services per capita rose about 2.85% per year over this period, to finish about 125% higher, while visit rates per capita were relatively stable (Andersen and Anderson 1967; Aday and Andersen 1975; Aday *et al.* 1980; United States, *op cit.* 1980). The growth in services per visit proceeds relatively evenly over this whole period. Over the same period, physician supply per capita was increasing at an average rate of 1.06% per year, total increase 35.9%, while physician services prices (unadjusted for changes in collection rates) outstripped the general price level by 1.34% per year or 47.3% over the whole period. Fee-adjusted billings per physician, "productivity," rose 63.4% or 1.7% per year. The substantial increase in relative prices thus accompanied a dramatic increase in total output, and the trends show no sign of abating.

¹⁷ One would also have to permit, and indeed encourage, open advertising and price competition if such a policy were to be effective.

¹⁸ This observation may appear inconsistent with data presented above, which show *increases* in numbers of auxiliaries per dentist in private practice. The concepts of complementarity and substitution in production provide the key. Classes of personnel which substitute for the skills of the dentist -- the expanded function auxiliaries or dental nurses -- which appeared in the early 1970s to be "the wave of the future" are now in the past; but complementary auxiliaries which add to the product line (hygienists) or expand the productivity and billing capacity of the dentist, are still popular. In the same way, physicians support the expansion of complementary hospital personnel who increase their billing opportunities, but resist competitive substitutes.