CPB Discussion Paper

No 6
June 2002

Financing medical specialist services in the Netherlands: Welfare implications of imperfect agency

Kees Folmer and Ed Westerhout

CPB Netherlands Bureau for Economic Policy Analysis
Van Stolkweg 14
P.O. Box 80510
2508 GM The Hague, the Netherlands

Telephone	+31 70 338 33 80
Telefax	+31 70 338 33 50
Internet	www.cpb.nl

ISBN 90-5833-103-2

Contents

Abstract

From 1995 onward the financing scheme for specialist care in the Netherlands has moved from a fee-for-service scheme to a lump-sum budget scheme. This paper analyses the economic and welfare effects of this policy change. The paper adopts a model that integrates demand and supply considerations and recognizes the potential roles of moral hazard and supplier-induced demand. The model is fully numerical, being estimated and calibrated upon data for the Dutch health care sector. The paper finds that the shift in financing regime has been welfare-reducing. The policy change induced medical specialists to lower the supply of health services which was already too low from a welfare point of view. This conclusion is robust to significant changes in major parameter values.

Keywords: fee-for-service scheme, lump-sum budget, medical specialists, moral hazard
JEL-codes: D60, H21,I18

1 Introduction[1]

For many years, the services delivered by medical specialists in the Netherlands have been financed according to some sort of fee-for-service scheme. Although the financing scheme in place was frequently adjusted, it basically remained fee-for-service as it linked the income of medical specialists to the volume of their output. This financing scheme has been heavily debated. In particular, it has been argued that it induces medical specialists to deliver more services than is in the patients' interests, increases aggregate expenditure on specialist care and makes this expenditure uncontrollable at the macro level.

Six years ago, things have begun to change. In 1995, the Dutch government allowed medical specialists to participate in so-called local initiatives, yearly negotiations between health insurers, hospitals and the medical specialists affiliated with these hospitals, in which a budget for medical specialists is negotiated that is independent of the volume of their services. The specialists were free to participate in these local initiatives. However, if they chose not to participate, they would run the risk that the government would continue to lower the fees for their services as part of a policy of macro budgeting. Tired of political conflicts and faced with the prospect of falling incomes, most medical specialists opted for participation in the new financing scheme.

This paper investigates whether the move from a fee-for-service system to a lumpsum budgeting system was good policies. We approach this question by using a model of specialist care in which consumption decisions reflect both demand and supply factors.[2] In particular, we adopt a principal-agent model in which medical specialists take consumption decisions and in which demand enters these decisions as specialists derive utility from the fulfilment of patient preferences. By including both demand and supply, we do justice to empirical evidence that shows that medical consumption is responsive not only to demand factors like the out-of-pocket price of medical consumption and patient income (*e.g.* Rosett and Huang (1973), Newhouse *et al.* (1993)) but also to supply factors like physician income and the number of physicians relative to the population (*e.g.* Evans (1974), Fuchs (1978), Newhouse (1987)).

Our model for medical specialists is empirically founded. Key parameters have been taken from time series estimation on Dutch data; the remaining parameters have resulted from calibrating the model upon 1995 values of relevant variables. We use this model to calculate the medical consumption effects of the financial reform. Next, we use it to explore the implications of this reform for patient welfare, physician welfare and social welfare.

[1] An earlier version of this paper was presented at the iHEA conference in York, July 22-25 2001. Thanks are due to the participants and to Peter Kooiman, Esther Mot and an anonymous referee for useful remarks. Ton Brouwer prepared the lay out of the graphs.

[2] Our model is part of a larger one that covers almost the complete Dutch health care sector (Folmer *et al.* (1997), Westerhout and Folmer (1997), Canton and Westerhout (1999) and Canton *et al.* (1999)).

Our model recognizes mainly two reasons why the reform of the payment scheme may have changed social welfare. First, it will be seen to have reduced medical consumption. In general, medical consumption may be too high or too low from a welfare point of view. The distortion of the price of medical consumption that is inherent in insurance schemes makes medical consumption too high; supplier-induced demand by medical specialists can make it too high or too low. In our simulations, negative supplier-induced demand is the dominant factor. The reduction of medical consumption that was due to the financial reform has thus reduced patient welfare.

Second, the change in the payment system will be seen to have led to an increase of leisure consumption by specialists, whereas their income has been left unchanged. However, ethical costs - attached to not fully meeting with patient preferences - increased. On balance, the reform has lowered the welfare of medical specialists. Social welfare, defined as the sum of patient and physician welfare, has also declined.

The sensitivity analysis undertaken indicates that our findings are fairly robust: various parameter configurations that differ considerably from our benchmark configuration yield similar results. Furthermore, our calculations suggest that we could even have derived similar results if we had taken the demand for medical services to be an exogenous variable. Thus, general equilibrium effects appear to be of little relevance in our analysis.

Our paper fits into the literature on the welfare effects of medical insurance (*e.g.* Feldman and Dowd (1991), Manning and Marquis (1996), Zeckhauser (1970)), but extends it by recognising the independent role of physicians. This extension has an impact on both the relation between fee and volume changes and the welfare consequences of the latter, as fee changes generally also affect the well-being of physicians. Next, our paper joins the literature on physician responses to fee changes (*e.g.* McGuire and Pauly (1991) and Rizzo and Blumenthal (1994)), but extends it by including the role of patients. This allows us to demonstrate that supplier-induced demand may help to combat the moral hazard that is due to insurance (see also Wedig *et al.* (1989)). Close to our paper are Ellis and McGuire (1990, 1993) which also focus on the interaction between demand and supply considerations. Different from these papers is that our paper explores the effects of a real-world experiment and adopts specifications for the demand and the supply of health care that have a strong empirical base. Furthermore, our paper does not take demand and supply policies as independent, but as related, due to the constraint that the financial reform that is analysed should not change the income of physicians.

Our paper is structured as follows. Section 2 describes some institutional features of specialist care in the Netherlands. Section 3 sets up a framework for evaluating the effects of different financing regimes. Next, section 4 constructs a model that fits into this framework and section 5 gives information about its empirical foundation. Section 6 uses this model to calculate the effects of various financing regimes. It also explores the robustness of these results. Finally, section 7 concludes.

2 Medical specialists in the Netherlands: some important features

This section describes some important features of the Dutch insurance scheme and of the position of the medical specialist in the health care system. These features are important as they motivate many assumptions that underlie our model. Next, it discusses the reform that has been implemented in the financing scheme for specialist care and that will be analysed in this paper. Subsequently, it briefly reviews existing empirical evidence on the effects of the financial reform upon the output of medical specialists.

In the Netherlands, two health insurance plans coexist. First there is a public insurance scheme. The government selects who is eligible for public insurance. Roughly, those with a labour income below about euro 27,500 are covered by the public insurance plan. This applies to about two-thirds of the population. The rest of the population can voluntarily seek insurance on the private market. In the Netherlands, only a negligible part of the population is not insured for health care expenditure. Private insurance policies differ from their public counterpart in risk differentiation in premiums and the use of deductibles. Our model recognizes this distinction between the public and private schemes, and distinguishes the demand of those with public insurance and those who are privately insured.

The Dutch health care sector is rather strongly regulated by the government, although current policies aim to lower the degree of regulation. In particular, the supply of services is controlled by requiring nearly all suppliers to have a permission to open a practice. Next, fees per contact or treatment are bounded from above. Unlike many other West-European countries, most specialists in the Netherlands are free entrepreneurs, although a growing number is employee in a hospital (Van Lindert *et al.* (1999)). Hence, medical specialists in the Netherlands are relatively autonomous, also in relation to hospital managers. The latter have a limited influence on the number of patients that should be treated and what kind of treatment should be chosen. Medical specialists are also relatively autonomous with respect to health care insurers. The reason is that, in the Netherlands, insurers have up to this moment largely abstained from managed care activities. Furthermore, specialists do not face budgets for their prescriptions of pharmaceuticals outside hospitals. Hence, they are also relatively autonomous in their prescription behaviour.

In order to increase the control of specialist expenditure, the Dutch government imposed a budget on expenditure on specialist care at the macro level in 1992. If this budget was exceeded, fees in the next year were reduced. This didn't yield the desired results however, mainly because there was no direct link at the level of the individual specialist between his output and any subsequent fee reduction. In 1995, medical specialists, hospitals and health insurers in a number of regions started to experiment with schemes with lumpsum budgets for medical specialists. The government supported these experiments, first by offering other regions in the Netherlands the possibility to participate in similar experiments, and, second, by stating that specialists that chose to participate would be freed from further fee reductions. As most medical specialists chose to participate, the government in effect enlarged the experiment to the whole field of medical specialists.

The fee-for-service system is still in place, though, but now only plays a role in the financing of the budgets. The direct link between the number of treatments and the income of the specialist has been removed. Moreover, income effects are largely absent, as budgets have been determined such that aggregate specialist income did not change substantially.[3]

An analysis of the consequences of the new financing scheme in the five hospitals involved in the first round of experiments indicates that the financial reform has changed the behaviour of specialists. The new system had effects upon various output variables like the admission probability, the waiting time and the referral ratio (the number of referrals to the general practitioner relative to the number of first visits) (Ziekenfondsraad (1998) and Van den Berg and Mot (2000)). The probability of admission decreased because of the experiment. The shift of financing scheme probably increased the average waiting time (Mot (2001)). In addition, a statistically significant effect has been found for the referral ratio: after the introduction of the new financing scheme, a significantly larger fraction of patients of medical specialists was referred back to a general practitioner (Van den Berg and Mot (2000)). On the other hand, there were no significant effects upon the average duration of stay. This could reflect that the duration of stay is to a large extent driven by technological growth.

Aggregate time series of hospital admissions, outpatients treatments and outpatient visits point in the same direction (Statistics Netherlands (2000)). Hospital admissions grew 1.1 percent a year during the period 1990-1994, but with the onset of the new financing regime the average yearly rate of growth in the period 1995-1999 dropped to -1.7 percent. Outpatient treatments grew firmly during 1990 -1994 at an average yearly rate of 10%, but the yearly growth rate has declined since 1995 to 5,5%. The development of outpatient consultations yields the same evidence. An average rate of increase during 1990-1994 of 1.3% was followed by an average decline in later years: -0.11% during 1995-1999. Yet, the insured population grew in the period from 1995 to 1999 at rates of 0.5 percent per year, whereas the share of people aged 65 and older increased from 13.2 to 13.5 in this period.

During 1997-1998, copayments were in place for the publicly insured. Although the maximum of copayments was relatively small, this may have had a distinctive effect upon the volume of specialist services demanded by the publicly insured. However, it cannot fully explain the developments in the 1995-1999 period: exclusion of the 1997-1998 years gives a similar picture for the 1995-1999 period. Apart from that, no other major policy reforms took place in the period under consideration, which suggests that the developments in hospital production in

[3] This is not to say that individual physicians did not face important shifts in their incomes. Indeed, reduction of income differentials across different specialties was one of the things that was negotiated in the local initiatives.

the 1995-1999 period can at least partly be attributed to the change in the financing scheme for medical specialists.

Summing up, micro evidence on specialist behaviour in the hospitals that were the first to experiment with the new financing scheme points to a negative effect on the volume of specialist services. Macroeconomic time series on various output variables in the 1995-1999 period hint in the same direction. Further, it is useful to remark that currently a new financing scheme, in particular a scheme based upon diagnosis-related-groups, is under discussion in the Netherlands, mainly because it is felt that the current lumpsum budgeting system offers too little incentives for specialist production.

3 A framework for comparing different financing regimes

To be able to compare different financing regimes, we use a model that consists of three blocks. The first two blocks describe the behaviour of patients and medical specialists respectively. The third block evaluates the results for different regimes on the basis of individual welfare functions.

Several elements warrant separate discussion here. The moral hazard effect of insurance schemes is well-documented in health economics. Our analysis takes this into account by setting up a general framework that encompasses a great variety of coinsurance schemes. A second important element is the characterisation of health care markets. We think it is appropriate to view markets for medical services as disequilibrium markets. Information asymmetry and lack of transparency and homogeneity of the services traded suggest that price adjustment as a balancing mechanism is at best only partly successful. The implication of this observation is that supply and demand may be different, both at the individual and at the aggregate level. A third caveat is that we have found demand effects to play a minor role in our analysis. Indeed, our calculations suggest that similar results would have been obtained had we taken demand to be exogenous. Still, we present the model with endogenous demand for we cannot claim a priori that demand effects will be irrelevant.

Having concluded that supply and demand may be different, the question arises which factor(s) determine actual consumption. The strong information asymmetry and the autonomous position of the medical specialist in the Dutch health care sector suggest that the role of medical specialists in determining actual consumption levels is large. For part of the medical markets, we take this to its extreme and postulate that physicians alone determine consumption levels. However, we do recognize the role of demand by letting demand be one of the factors that govern the behaviour of physicians.[4] Indeed, the aggregate supply model we will use is a mix of a principal-agent model in which the medical specialist acts as an agent of the patient and a neoclassical labour supply model which describes the specialist's choice between consumption and leisure.

The model for patient behaviour derives the demand for specialist services from a utility function which includes the consumption of specialist services and other products as arguments. Using the patient budget constraint, utility can be expressed as a function of the patient's income (net of health care premiums), y, and the relative price of medical services, t_c. The patient population is heterogeneous with respect to the need of health care, measured by , ,

[4] Actually, it would be better to speak of notional demand. For it is difficult to define demand given that the consumers of medical care are by assumption imperfectly informed and go to the doctor precisely to get informed. Indeed, demand should be interpreted as the amount of services that consumers would have preferred had they been perfectly informed. For brevity however, we will use the term demand throughout the paper.

and the insurance status, denoted by index *j*. As the allocation over the public and private insurance scheme is on the basis of income, we also distinguish an average income related to the insurance status. Hence, the utility function can be written as $u_{i,j} = u(z_{i,j} \mid y_j, t_c, \theta_i, j)$ where *i* indexes patients and *z* denotes medical demand.[5] Maximisation of utility with respect to the consumption of health care yields a demand function which expresses demand as a function of income and the price of medical services: $z_{i,j} = z(y_j, t_c, \theta_i, j)$. Aggregation yields the corresponding aggregate demand for medical services: $Z_j = N_j \int z_{i,j} dG_j(\theta_i)$ where N_j denotes the number of households in insurance scheme *j* and $G_j(\theta_i)$ is the distribution function of θ_i for the insured of type j. Equating the demand for first consultations to the number of diseases for which patients seek treatment, demand for all medical services can be decomposed into the demand for first consults, $Z^f_j = \eta_j Z_j$, and demand for specialist services, $Z^s_j = (1-\eta_j) Z_j$. Finally, aggregation over the two insurance schemes yields aggregate demand for first consultations and treatments: $Z^f = \sum_j Z^f_j$ and $Z^s = \sum_j Z^s_j$.

The second block describes the behaviour of medical specialists. Medical specialists are assumed to decide only on the supply of specialist services during subsequent consultations of patients, taking as given the number of first consultations (Rutten (1978)).[6] Maximisation of a utility function $v_k(s_k \mid Z^s, Z^f, t, e_k)$ under appropriate constraints yields the supply function for subsequent medical services, s_k: $s_k = s(Z^s, Z^f, t, e_k)$, where *k* indexes physicians. It is a function of demand, Z^s, the number of first consultations, Z^f, the fee for medical services [7], t, and the ethical cost variable e. As long as the specialist acts as a perfect agent of the patient, ethical costs are zero. Costs are positive as soon as she acts in her own interest. The subscript *k* on *e* reflects that physicians differ in their ethical costs. Aggregation over all specialists yields aggregate supply $S = N_s \int s_k dG_s(e_k)$ where N_s denotes the number of physicians and $G_s(e_k)$ is the distribution function of e.

Obviously, there is no mechanism in our model that ensures that demand and supply will coincide. Indeed, S and Z^s will generally deviate. To derive the supply of services for

[5] The absence of the i index with the income variable denotes that our model abstracts from the heterogeneity of patient income. Given that the income elasticity of the demand for medical care is relatively small, this abstraction is not particularly relevant in the context of our analysis.

[6] The distinction in our model between first consultations and subsequent specialist services aims to separate the health care decision in a decision to seek care and a decision on the amount of treatment. Note that our assumption that it is the consumer who decides whether or not to seek care and that it is the specialist who decides on the intensity of treatment conforms to the findings reported in Newhouse (1996) that demand-side cost sharing mainly affects decisions to seek care, whereas supply-side cost sharing mainly operates on the treatment intensity.

[7] As described in section 2, until 1997 different fees for specialist services applied to publicly and privately insured patients. In this paper we have assumed that fees have been harmonised since 1995 to be able to study the impact of the specialist budget separately from that of the fee harmonisation. For a discussion of the effects of fee differentiation, see Canton *et al.* (1999).

individual patients, we adopt the idea of proportional rationing, *i.e.* at the level of the individual patient the ratio of supply and demand equals S/Z^s.

First consultations are beyond the control of the specialist. Hence, the consumption of first consultations X^f is demand-determined: $X^f = Z^f$. However, the consumption of subsequent services X^s is determined by the supply decision of the medical specialist: $X^s = S$. Total medical consumption can then be calculated as $X = X^f + X^s = Z^f + S$. Similarly, total medical consumption for patient *i* of type *j* reads as [8] $x_{i,j} = z^f_{i,j} + s_{i,j} = \eta_j z_{i,j} + (S/Z^s)(1-\eta_j)\, z_{i,j}$.

Next to patients and physicians, two other actors exist in the model. First, two different types of insurance companies collect premiums from the insured which they use to cover medical expenditure net of copayments. By assumption, the two industries make zero profits. Second, the government redistributes income across medical specialists. In particular, patients and insurers pay t_c, but medical specialists collect t per unit of medical services. The difference, $t_c - t$, is distributed by the government to specialists on a lumpsum basis, *i.e.* independent of their medical production. Hence, if $t_c - t$ is negative, the government actually levies a lumpsum tax on specialists. We choose the lumpsum budget of specialists such as to neutralise the income effects of changes in the services fee. This device bears resemblance with the financial reform that is analysed (see section 2). Besides, income neutrality implies that policy changes do not affect the distribution of income over patients and physicians, which allows us to exclude income considerations from our welfare analysis.

The evaluation of patient welfare comes down to recalculating utility, with demand $z_{i,j}$ replaced with actual consumption $x_{i,j}$. Aggregate patient welfare can in turn be calculated as $U = \sum_j N_j \int u_{i,j}\, dG_j(\varepsilon_i)$. The evaluation of physician welfare comes down to calculating the value of utility v_k for the optimal amount of services s_k. Aggregate physician welfare then reads as $V = N_s \int v_k\, dG_s(e_k)$. In order to calculate social welfare, we first calculate the changes in patient welfare and physician welfare which are due to the policy change. Next, we convert these changes into their consumption equivalents, *i.e.* the welfare changes in terms of units of the non-medical consumption good, in order to bring the two welfare changes on the same footing. The change in social welfare is then simply the sum of all consumption equivalents.

All variables can be expressed as a function of the fee for medical services t which serves as a policy instrument. Hence, we can calculate the level of welfare that corresponds to the current regime in the Netherlands and compare it with the welfare level of the pre-reform fee-for-service regime. This allows us to evaluate the welfare effects of the financial reform. In addition, we can derive the financing scheme that is optimal in the sense of maximising patient welfare, physician welfare or social welfare.

[8] The expression for $x_{i,j}$ reflects that we assume η_j or the ratio between the demand for first consultations and the demand for subsequent treatments to be the same for households who have different ε_i. This assumption is made for simplicity, but may be unrealistic. However, it will not greatly affect our conclusions.

Having laid down our general framework, we now make concrete the three building blocks that together make up our model. This involves first the specification of utility functions and constraints for patients and medical specialists. Next, we have to specify the heterogeneity of the patient and physician population. Thirdly, we have to fill in numbers for the various parameters that enter into the models of patient and specialist behaviour. For all these purposes, we rely upon a model for the Dutch health care sector, which was constructed primarily to assess the numerical effects of various policy reforms.

4 A model of specialist care

4.1 Patient behaviour

The patient population consists of two groups, those who are publicly insured and those who are privately insured. The two groups have different average income levels[9], face different copayment schemes and generally differ in their need of health care. Their economic behaviour may be very similar though. Hence, we adopt one model of patient behaviour which we implement separately for the two groups of insured. This section discusses this model.[10] For ease of notation, we omit the insurance index.

The patient derives utility from the consumption of medical specialist services, *z,* and the consumption of other goods and services, *c*. Medical specialist services are sometimes referred to as 'consultations' or 'visits', while in fact they refer to treatments. Patients are assumed to have linear-quadratic additively separable preferences over the two types of goods,

$$u_i \;'\; c_i \;\&\; \tfrac{1}{2}\,(\,c_i^2 \;\%\; ,_i z_i \;\&\; \tfrac{1}{2}\,*\,z_i^2$$
$$*\; > \;0,\;\; ,_i \;\geq\; 0,\;\; 0 < (\, < 1/y \qquad (1)$$

where u_i is patient *i*'s utility and *y* is the income of the patient net of health care premiums.

Our motivation to employ a linear-quadratic specification is twofold. First, we want to allow the marginal utility of medical care to turn negative if medical consumption becomes sufficiently high (*i.e.* the possibility of iatrogenesis, see Lee (1995)). Second, zero marginal utility of medical care is needed to obtain a finite solution for the demand for health services when the out-of-pocket price is set at zero. Note however that negative marginal utility of non-medical consumption is excluded by imposing $(< 1/y$.[11] We allow medical need to be unevenly distributed among individuals. This heterogeneity is reflected in the model by assuming , to differ across individuals.

The patient faces a kinked budget constraint. He pays the cost of medical services up to some copayment maximum *m*; above this maximum, the out-of-pocket price of medical services equals zero. This modelling device covers the copayment systems that feature the two insurance

[9] Our model does not take into account the heterogeneity of incomes within the groups of the publicly insured and the privately insured. Similarly, it does not include the heterogeneity of copayment schemes among the privately insured. These simplifications are not relevant in the context of our analysis.

[10] See Westerhout and Folmer (1997) for a more detailed discussion of the patient models.

[11] As the value of the parameter (is independent of the insurance status of the patient (see section 5), we use $1/\max_j y_j$ rather than 1/y as the relevant boundary.

schemes in the Netherlands. Indeed, private insurance schemes in the Netherlands feature deductibles, whereas in the public insurance scheme copayments are fully absent.

Defining the non-medical product as numeraire, the price of the medical product by t_c[12], we can formalise the patient's budget constraint as

$$c_i \leq y \,\&\, \min(m, t_c z_i) \tag{2}$$

Equation (2) states that consumption of medical services restricts the consumption of other goods and services as long as copayments are below the maximum m.

The demand for medical services follows from the maximisation of utility function (1) subject to budget constraint (2). The nonsatiation assumption $(< 1/y$ implies that the budget constraint is always binding. The description for the demand for medical services depends on whether the copayment maximum m is higher or lower than a critical value $m^{(}$ (defined in Appendix A). In the former case, the expression for the demand for medical services depends upon two critical values for $,_i$, denoted by $,^{(}$ and $,^{((}$ (which are defined in Appendix A), for which $,^{(} > ,^{((}$.[13] The demand for medical services is described by the following system:

$$z_i \;'\; 0 \qquad ,_i \leq ,^{((} \tag{3}$$

$$z_i \;'\; \frac{\&t_c}{*\%t_c^2(}\,(1\&(y)\;\%\;\frac{'_i}{*\%t_c^2(} \qquad ,^{((} \leq ,_i \leq ,^{(} \tag{4}$$

$$z_i \;'\; \frac{'_i}{*} \qquad ,_i \geq ,^{(} \tag{5}$$

In case $m < m^{(}$, the demand for medical services depends upon one critical value for $,_i$, denoted by $,^{(((}$ (which is also defined in Appendix A). The demand for medical services is then

[12] More generally, the out-of-pocket price equals bt_c, where b represents the copayment rate ($0 \leq b \leq 1$). As in the Netherlands for services of medical specialists only the case in which b equals one is relevant , we set b at this value.

[13] In principle, the solution to the demand for medical services can be defined for two different cases. In the polar case $,^{((} > ,^{(}$, the copayment maximum is so low that anybody who consumes a positive amount of medical services pays the maximum amount of copayments. In case $,^{((} < ,^{(}$, a positive fraction of the patient population faces copayments below the maximum. As the former case lacks any empirical relevance, this section elaborates only the latter case. Also in our numerical simulations, the former case does not occur.

described by the following system:

$$z_i \; ' \; 0 \qquad ,_i \leq \; ,^{(((} \tag{6}$$

$$z_i \; ' \; \frac{,_i}{*} \qquad ,_i \geq \; ,^{(((} \tag{7}$$

At the population level, we can express the demand for medical services for each type of insured as a weighted sum of the relevant components. As explained in section 2, the model distinguishes the patient population with respect to type of insurance: public and private insurance. If we use the fact that copayments are absent in the public insurance scheme and that the copayment maximum in private schemes is rather high, we can derive that aggregate demand of the privately insured obeys equations (3) to (5); aggregate demand of the publicly insured is described with equation (7) ($,^{(((} \; ' \; 0$ if $m \; ' \; 0$). If we use subscripts p and n to refer to publicly and privately (non-publicly) insured patients respectively, aggregate demand for services by the two groups can be formulated as follows:

$$Z_p \; ' \; N_p \frac{+_p(,_i)}{*_p} \tag{8}$$

$$\begin{aligned} Z_n \; ' \; & (G_n(,_n^{(}) \,\&\, G_n(,_n^{((}))\, N_n \left[\frac{\& t_c(1 \,\&\, (_n y_n) \;\%\; +_n(,_i \mid ,_n^{((} < ,_i < ,_n^{(})}{*_n \,\%\, t_c^2 \,(_n} \right] \\ & \%\; (1 \,\&\, G_n(,_n^{(}))\; N_n \frac{+_n(,_i \mid ,_i > ,_n^{(})}{*_n} \end{aligned} \tag{9}$$

where $G(.)$ stands for the cumulative distribution function of $,$, $+(x|y)$ is the expectation of x conditional upon y and N is the size of the corresponding population of insured.

Expressions (6) and (7) demonstrate that the demand by the publicly insured is unresponsive to the price for medical services. Expressions (3) to (5) illustrate how the demand for specialist services by the privately insured reacts upon a change in the price of medical services. Obviously, a price increase does not affect the demand of those who do not consume any medical services (equation (3)). It neither affects the demand of those who face a zero out-of-pocket price as their medical expenditure already exceeds the amount of the deductible (equation (5)). On the other hand, an increase in the price of medical services unambiguously reduces consumption of those who have not emptied their deductible (equation (4)). In sum, the direct effect of a price increase is to lower demand.

However, apart from this direct effect, compositional effects play also a role. Indeed, price changes generally influence the size of the three groups of consumers. It can be derived from equations (A1) - (A5) in Appendix A that $,^{*}$ and $,^{**}$ are decreasing and increasing functions of the service fee t_c respectively. In particular the former is relevant since demand z_i is discontinuous at $,_i \ ' \ ,^{(}$. As a price increase implies that consumers will exhaust their deductible with a lower level of medical consumption, the range of consumption over which the out-of-pocket price is zero is increased. This boosts the demand for medical consumption. It is unlikely though that the compositional effects will dominate the direct effect. Indeed, in none of the numerical versions we experimented with, was this the case. Therefore, we will not discuss further the compositional effects.

Assuming that each disease takes one first consultation, we can split demand into the demand for first consultations and subsequent services as follows:

$$Z_p^f \ ' \ 0_p Z_p \qquad Z_n^f \ ' \ 0_n Z_n \tag{10}$$

$$Z_p^s \ ' \ (1 \& 0_p) Z_p \qquad Z_n^s \ ' \ (1 \& 0_n) Z_n \tag{11}$$

where 0_p and 0_n equal the share of the demand for first consultations in total demand of publicly and non-publicly insured patients, respectively. The expressions for aggregate demand at the population level then speak for themselves:

$$Z^f \ ' \ Z^f{}_p \ \% \ Z^f{}_n \tag{12}$$

$$Z^s \ ' \ Z^s{}_p \ \% \ Z^s{}_n \tag{13}$$

4.2 The behaviour of medical specialists

The basic theoretical structure from which the behaviour of medical specialists is derived is the neoclassical labour supply model.[14] To account for some specific features of specialist behaviour, this basic structure is extended along two dimensions. First, in formulating the time and budget constraint, the specialist takes as given the amount of first visits. Patients decide on first consultations so that these are beyond the physician's control.

Second, an agency relationship between the patient and the specialist is imposed. In principle, the presence of imperfect information at the side of the patient makes it possible for the medical professional to deviate from the patient's interest and pursue other (personal) objectives. On the

[14] See also Folmer *et al.* (1997) for a discussion of the physician model.

other hand, we do not think it reasonable to assume that the physician can fully neglect the patients' interests. Reputation considerations, the medical oath, or medical ethics all lead us to regard physicians as agents of patients. This special agency-feature is allowed for in the model by imposing a (fixed) ethical cost on the professional when she deviates from the patient's best interest (in other words: when the action selected by the professional deviates from the action that the patient would have selected under perfect information).

A medical professional *k* derives utility, v_k, from leisure, l_k, and the consumption of other goods and services, d_k. The utility function is specified in CES-format,

$$v_k(d_k, l_k, e_k) = \left(\alpha d_k^{-\rho} + l_k^{-\rho}\right)^{-1/\rho} - e_k \qquad \rho > -1 \tag{14}$$

where *e* refers to the ethical cost which is imposed on the physician when she fails to deliver demand. The assumption that ethical costs of not acting in the patient's best interest are fixed is made in order to simplify the analysis.[15] An implication of this assumption is that physicians have to choose between two options, *viz.* an 'ethical' and a 'financial' option.

The physician maximises her utility subject to a time constraint and a budget constraint. The time constraint says that leisure time is determined from the difference between the total time allotment *T* and the time allocated to consultations. Recognising that the number of first visits per specialist reads as Z^f/N_s (where N_s denotes the number of specialists) and denoting the number of subsequent consultations by *s*, we have,

$$l_k = T - \frac{\mu^f Z^f}{N_s} - \mu^s s_k \tag{15}$$

where μ^f units of time are needed for a first consultation and μ^s units of time for every subsequent visit. We assume that it always holds true that $T - \mu^f Z^f/N_s > 0$.

Consumption equals the professional's income, which is calculated as the revenue from subsequent consultations and an additional income component beyond the control of the physician. Or,

$$d_k = t s_k + a \tag{16}$$

where *t* stands for the fee per service supplied and *a* is income beyond the physician's control. This type of income comprises two items,

[15] The fixed nature of ethical costs may be unrealistic. In particular, ethical costs may be increasing in the deviation between demand and supply. As this would seriously complicate the derivation of a tractable empirical specification, we adopt the dichotomous specification discussed above.

$$a \;'\; h \,\%\, \frac{tZ^{f}}{N_{s}} \tag{17}$$

h is the physician's lumpsum income (net of the fixed costs of running a doctor's practice). The second term in equation (17) determines the physician's income from first consultations.

In order to derive the physician's behaviour, it will be useful to first obtain the solution that maximises the physician's utility from consumption and leisure (*i.e.* the first term at the RHS of (14)). This solution characterises the financial option. The interior solution of the maximisation problem, denoted $\bar{s}^{(}$, reads as

$$\bar{s}^{(} \;'\; \$\frac{T}{\mu^{s}} \,\&\, \$\frac{\mu^{f}}{\mu^{s}}\frac{Z^{f}}{N_{s}} \,\&\, (1\&\$)\frac{a}{t} \tag{18}$$

where the auxiliary variable \$ is defined as:

$$\$ \;'\; \frac{t^{F\&1}(\mu^{s})^{1\&F\text{"}F}}{1 \,\%\, t^{F\&1}(\mu^{s})^{1\&F\text{"}F}} \tag{19}$$

and F=1/(1+D) is defined as the (constant) elasticity of substitution between consumption and leisure. Equation (19) shows that 0<\$<1.

Obviously, labour supply cannot be negative. Hence, labour supply under the financial option is described by the interior solution to the corresponding maximisation problem, $\bar{s}^{(}$, except when $\bar{s}^{(} < 0$, in which case 0 replaces the interior solution. Formally, $\bar{s} \;'\; \max(0, \bar{s}^{(})$. The solutions for non-medical consumption and leisure that correspond to the financial option read as $\overline{d}$ and $\overline{l}$ respectively, and follow from substituting this value of $\bar{s}$ into expressions (16) and (15).

Note that as the service fee does not distinguish between the publicly and the privately insured, the physician is indifferent between supplying medical services to the publicly or the non-publicly insured[16]. For the same reason, she is indifferent to supplying services to consumers that differ in their medical need $,_{i}$. In order to obtain a unique solution for the supply at the level of the individual patient, we assume that the specialist who prefers the financial option divides her services between all her patients so that each patient faces the same discrepancy between demand and supply (proportional rationing).

[16] If fees were different, the specialist that prefers the financial option would supply subsequent consultations only to the group of insured to which the highest fee value applies (see Folmer *et al.* (1997)).

We stress that expression (18) implies that financial physicians may supply more or less medical services than demanded, depending on the physician's particular preferences for consumption and leisure and her incentives to supply medical services. In particular, the physician may supply more than demand to raise her income and supply less than demand to gain leisure time.

Under the ethical option, a specialist accommodates her supply of subsequent consultations, $\tilde{s}$, to the corresponding demand, Z^s / N_s. The only exception occurs when this behaviour would imply negative leisure. In that case, $T - \mu^f Z^f / N_s$ replaces Z^s / N_s. This effect does not occur in any of our simulations, however. Apart from this exceptional case, both the publicly insured and the privately insured receive the amount of medical services they would have demanded in the absence of any information imperfections. We use $\tilde{d}$ and $\tilde{l}$ to denote the consumption of non-medical goods and leisure under the ethical option. These two variables are derived by substituting the solution for $\tilde{s}$ into expressions (16) and (15) respectively.

Now we can calculate the critical level of ethical costs, denoted as e^*, for which the physician is indifferent between the financial and the ethical option by solving the equality:

$$
\begin{aligned}
v_k(\bar{d}_k, \bar{l}_k, e^*) &= v_k(\tilde{d}_k, \tilde{l}_k, 0) \quad \Rightarrow \\
e^* &= \bar{v} - \tilde{v} \\
&= \left(\alpha \bar{d}^{-\rho} + \bar{l}^{-\rho}\right)^{-1/\rho} - \left(\alpha \tilde{d}^{-\rho} + \tilde{l}^{-\rho}\right)^{-1/\rho}
\end{aligned}
\tag{20}
$$

where $\bar{v}$ and $\tilde{v}$ are implicitly defined by the first and last term of the second line of equation (20) respectively.

Using this expression for e^*, the outcome of the optimisation problem can now be summarised as follows:

$$
\begin{aligned}
s_k &= \bar{s} \qquad e_k \leq e^* \\
s_k &= \tilde{s} \qquad e_k \geq e^*
\end{aligned}
\tag{21}
$$

A physician thus prefers the financial option if e^* exceeds her ethical cost variable e. The ethical option is chosen whenever e^* is below this threshold value. By definition, the physician is indifferent between the two options when $e = e^*$.

We now assume that physicians are heterogenous with respect to the value of the ethical cost variable e. Hence, a number of physicians will prefer the ethical option while others will choose the financial option. Denoting the distribution function of e by $G_s(e)$, the supply of medical services at the aggregate level can then be expressed as a weighted average of supply as defined for the two options:

$$S = G_s(e^*)\,N_s\bar{s} + (1 - G_s(e^*))\,Z^s \tag{22}$$

What are the effects of an increase in the service fee? We distinguish a direct effect plus two groups of indirect effects. The direct effect of a fee increase is to change the supply by financial specialists. Here, two factors work in opposite direction. A higher fee makes it financially more attractive to provide medical services (a substitution effect), but also increases the physician's income and thereby her demand for leisure (an income effect).[17] In the numerical version of our model, the elasticity of substitution is sufficiently high for the substitution effect to dominate, so the effect of a fee increase is to increase supply.[18] However, this works only for intermediate fee levels. Too low fee values restrict the supply by financial specialists to be zero. Combined, we should expect the supply of services by financial specialists to be a function of the services fee that consists of two parts: a flat part for low fee levels and a decreasingly upward-sloping part for high fee levels.

Note that the supply by ethical specialists is not directly affected by an increase in the services fee. The total supply function should therefore have properties identical to the function that describes the supply by financial specialists. However, two groups of indirect effects disturb the picture.

First, as will be shown below, the consumer price may decrease when the producer price increases on account of a general-equilibrium effect. This decrease may boost the demand for first consultations and subsequent consultations which basically has two effects. On the one hand, the ethical specialist responds to the increase in demand for subsequent consultations by raising her supply of medical services. On the other hand, the financial specialist responds to the increase in first consultations by reducing her supply of subsequent consultations. Indeed, the increase in first consultations leaves her less time to supply subsequent services and increases her income collected from first consultations. As the effect of an increase in demand is to increase the supply by ethical specialists, but to decrease the supply by financial specialists, the effect upon total supply remains ambiguous.

[17] The income effect relates not only to the supply of subsequent treatments, but also to the supply of first treatments.

[18] Obviously, one may remark that because the lumpsum income component is adjusted in order to compensate for the income effects of fee changes, the income effect is nil. However, this is not entirely correct. The point is that ethical and financial specialists differ in their supply and thus in the income effects due to fee changes whereas the lumpsum income adjustment is the same for the two groups of specialists. As income is relevant only for the supply behaviour of financial specialists, redistribution from ethical towards financial specialists may reduce aggregate supply.

Second, compositional effects play a role too. Indeed, the fraction of specialists that choose to be ethical or financial is endogenous and may thus respond to a change in the services fee. Intuitively, it is clear that the fraction of financial specialists is a positive function of the service fee in deviation from the fee level for which supply and demand are equal. At the latter fee level, 100% of the specialists choose for the ethical option as the two options coincide with respect to levels of consumption and leisure, but only the financial option includes ethical costs. If the fee deviates from this specific fee level, the financial option generally corresponds to a different pair of consumption and leisure. Assuming that the consumption-leisure pair that corresponds to the ethical option does not change,[19] the change in the fee drives a wedge between the consumption-leisure pairs that correspond to the ethical and financial option. As the latter implies the highest utility, it follows that the change in fee must increase the utility differential between the financial and ethical option. This will lead a number of specialists, namely those with the lowest ethical costs, to decide to switch to the financial option. Consequently, the number of financial specialists increases. As the argument relates to the deviation of the service fee from the level for which demand and supply are equal, it follows that the fraction of financial specialists is a U-shaped function of the fee, with a minimum of zero at the fee level for which demand and supply coincide. The role of this compositional effect is generally ambiguous. Whether the change in the fraction of financial specialists that corresponds to the increase in service fee, strengthens or weakens the relation between supply and service fee, depends on whether the financial option or ethical option features the highest supply.

Summing up, we have a direct effect and two indirect effects, of which only the former is unambiguous. As we will see, the supply effects of fee changes in the numerical version of our model correspond to a large extent with this direct effect.

4.3 Closure of the model

We close the model by specifying the calculation of copayments and premiums, of the consumer price of medical services and of lumpsum subsidies to specialists.

For the calculation of copayments, note that patients whose expenditure is lower than the copayment maximum m pay their health care expenditure $t_c x_{i,j}$, where $x_{i,j}$ refers to realised consumption of medical services:

$$x_{i,j} = z^{f}_{i,j} + s_{i,j} \tag{23}$$

[19] Note that we make this assumption only for the purpose of explanation. In our model, the consumption-leisure pair corresponding to the ethical option does change, but less than the pair that corresponds to the financial option. Hence, qualitatively the argument developed above holds true.

Here, z^f refers to demand for first treatments and s equals supply of subsequent treatments.

Copayments by patients with medical expenditure higher than m equal m. Denoting the population shares of the two groups as $G(\hat{\theta}')-G(\hat{\theta}'')$ and $1-G(\hat{\theta}')$ respectively, aggregate copayments, denoted Q, read as follows (see equation (9)):

$$Q = N\left[t_c \int_{\hat{\theta}''}^{\hat{\theta}'} x_i \, dG(\theta) + (1-G(\hat{\theta}'))\, m\right] \tag{24}$$

Note that $\hat{\theta}'$ and $\hat{\theta}''$ differ from their counterparts θ' and θ'' as the former are based upon actual consumption rather than demand. Premiums, P, follow easily from subtracting copayments from total expenditure, $t_c X$, where X reads as $Z^f + S$:

$$P = t_c X - Q \tag{25}$$

The consumer price of medical services, t_c, consists of the fee-for-service, t, plus a tax levied by the government at rate τ:

$$t_c = t(1+\tau) \tag{26}$$

The revenues from this tax, $\tau t X$, are used to finance lumpsum subsidies to specialists, $N_s h$:

$$\tau = \frac{N_s h}{tX} \tag{27}$$

Hence, the specialist receives her income from two sources. The first is the payment for the services she delivers at the fee-for-service rate t. The second is the lumpsum subsidy that is granted by the government (see also equations (16) and (17)).

The policy change that this paper analyses keeps specialists on their initial income level. This is achieved by adjusting lumpsum subsidies. This implies that h is calculated as follows. Combine expressions (16) and (17) to yield the following expression for income per specialist:

$$d_k = h + t\left(\frac{Z^f}{N_s} + s_k\right) \tag{28}$$

Next, by aggregation over all specialists we obtain

$$D = N_s h + t(Z^f + S) = N_s h + tX \tag{29}$$

where $D = N_s \int d_k dG_s(e_k)$ and S is defined in equation (22). This equation can be written as an expression for h:

$$h = \frac{D - tX}{N_s} \tag{30}$$

This expression tells us which value of h keeps D constant when the government changes the fee level t.

The neutrality of specialist income with respect to the fee for medical services implies that health care expenditure is a constant. The same holds true for the sum of health care premiums and copayments, as this equals health care expenditure. The same holds also true for non-medical consumption ***c*** of patients, which equals patient income net of the sum of health care premiums and copayments. The welfare analysis will make use of this property.

4.4 Welfare

For both the publicly insured and the privately insured, we define patient welfare U as the sum of individual utilities, where the summation runs over consumers who differ in their value for θ_i:

$$U_j = N_j \int_0^\infty u_{i,j} dG_j(\theta_i) = N_j \int_0^\infty \left[c_{i,j} - \tfrac{1}{2}\alpha_j c_{i,j}^2 + \theta_i x_{i,j} - \tfrac{1}{2}\beta_j x_{i,j}^2 \right] dG_j(\theta_i) \qquad j = p,n \tag{31}$$

Appendix B provides further details on the elaboration of the expression for patient welfare when health care consumption is lognormally distributed.

Similarly, welfare of medical specialists, V, is defined as the sum of the welfare of specialists choosing for the ethical and the financial option:

$$
\begin{aligned}
V &= \bar{V} + \tilde{V} - N_s E(e_k \mid e_k < e^*) \\
&= N_s\left[(1 - G_s(e^*))\,\bar{v} + G_s(e^*)\,\tilde{v} - \int_0^{e^*} e_k \, dG_s(e_k)\right]
\end{aligned} \tag{32}
$$

where $\bar{V}$ and $\tilde{V}$ are implicitly defined by the first and second term on the second line of equation (32).

In formula (32), $G_s(e^*)$ equals the share of specialists that opt for the financial option. Ethical costs are only relevant for those specialists who choose to bear these costs, *i.e.* the specialists whose e_k is below e^*.

We also want to calculate the effect of the financial reform upon social welfare. Note that simply adding up patient welfare (equation (30)) and specialist welfare (equation (31)) makes no sense due to the ordinal nature of the two welfare functions. Therefore, we first transform the changes in patient and physician welfare into their consumption-equivalent counterparts. The latter are defined in terms of the non-medical consumption good which is consumed by both patients and specialists. The two consumption-equivalent changes can be added up to arrive at the consumption-equivalent change in social welfare. Appendix C gives a detailed specification of the method of consumption-equivalent welfare changes.

Our welfare measure is quite standard in the literature, although one may find the neglect of redistributional concerns a drawback. However, note that our analysis focuses more on efficiency effects than on distribution effects (witness also the assumption of proportional rationing made in section 4). Therefore, it would be difficult to motivate a welfare measure that includes redistributional aspects.

5 Empirical foundation of the model

The empirical foundation of the model is based on time series estimation as well as on calibration. The demand for subsequent treatments cannot be observed, as - according to our model - data on subsequent treatments reflect both supply and demand elements. Therefore we have proceeded in reverse order. We first estimated the supply model. This estimation yields an estimate of demand. This estimate of demand was then used as an input in the calibration of our patient models.

Medical specialists

We have estimated a linearize reduced-form expression for the growth in the supply of specialist services (equation (21)) using time series analysis. Subsequently, we have transformed these estimates into values of the structural parameters of the specialist model. This procedure implies that the parameter values of the model depend on estimated coefficients and exogenous variables as well. That is to say, it is possible to obtain a complete set of structural coefficients for each year in the sample. For details on the estimation and calibration procedure we refer to Folmer (1998).

To promote stability of the share of specialists that prefer the financial option, we have imposed that at least 70% of the specialists always prefers the ethical option. This guess has been based on the computed values of $G_s(e^*)$ in all sample years. For the remainder, the distribution of ethical costs *e* across specialists is chosen to be lognormal. Apart from being fully identified by two parameters, the lognormal form for $G_s(e)$ ensures that the argument *e* is always positive. This choice implies that ln *e* is normally distributed. Its mean and standard deviation are denoted as μ_e and F_e. The values of these parameters can be obtained from estimated coefficients, as the empirical equation for the supply of services is specified in growth rates.

Table 5.1 Parameter values in the specialist model (average 1991/1995)

parameter	description	value
F	elasticity of substitution between consumption and leisure, financial specialist	6.91
"	relative weight of consumption in utility specialist	0.0016
μ_s	mean of ln e	7.35
F_e	standard deviation of ln e	1.35
	minimum fraction of specialists that prefers to be ethic	0.7
variable	**description**	
$G_s(e^*)$	relative preference for the financial option	0.13
$	definition: see equation (19)	0.1
e^*	value of ethical cost at which the physician is indifferent between both options	1272

Table 5.1 summarises the results of our calibration procedure. The elasticity of substitution between consumption and leisure time looks rather high, namely 6.91. However, this value applies to financial specialists only. As a measure of the substitution possibilities between leisure and consumption, the average of the substitution elasticities corresponding to financial and ethical specialists, $G_s(e^*)$ F, may be better. This average elasticity of substitution equals 0.91 in the calibration year, which is in line with estimates in the literature (Rizzo and Blumenthal (1994)). Supply is 10.8% below demand in the calibration year, reflecting a substantial degree of negative supplier-induced demand.

Patients

Canton *et al.* (1999) describe in detail the empirical implementation of the model for both the privately and the publicly insured. Here we will globally discuss this issue.

The procedure consists of four steps. First, recall that the parameter ,$_i$ is stochastic. We characterise its distribution as lognormal. The motivation for doing so is that there is evidence that health care expenditure is lognormally distributed (see Van Vliet and Van der Burg (1996) for the Netherlands and Duan *et al.* (1982) for the United States) and, according to our model, the parameter ,$_i$ is highly correlated with expenditure. As to the latter, the parameter ,$_i$ is proportional to the demand for health care at a zero out-of-pocket price. This means that the correlation between ,$_i$ and expenditure can only be lower than one due to copayments.[20] Given lognormality, the distribution of ln(,) is normal with parameters, say, μ$_,$ and F$_,$. Van Vliet and Van der Burg (1996) have computed the coefficient of variation of the distribution of health care expenditure based on cross section data for 1991 - 1994, from which it is easy to derive an average estimate of F$_,$.

This leaves us with three parameters that remain to be identified: *, μ$_,$ and (. These three parameters are calibrated simultaneously by using information on the aggregate income elasticity of the demand for medical services, the level of average demand and the insurance effect. The latter is defined as the ratio of the demand of a fully insured (with a zero copayment maximum) and an uninsured patient and in turn derives from an estimate of the price elasticity of health care demand (Van Vliet (1998)). This insurance effect equals 1.2, which is in line with estimates in the RAND Health Insurance Experiment (Newhouse *et al.* (1993)). As explained above, the level of average demand is derived from the calibration of the specialist model, which characterises demand as one of the determinants of supply.

For the publicly insured, no information is available on price and income elasticities as the public insurance scheme lacks copayments. We assume therefore that the corresponding

[20] Supplier-induced demand cannot play a role as we assume specialists to ration different consumers proportionally (see above).

structural parameters of the model of the privately insured apply to the model of the publicly insured as well. Consequently, the values of ($_p$ and *$_p$ equal those of ($_n$ and *$_n$ respectively.

Table 5.2 summarises the calibration of the two patient models. The price elasticity of demand is generated by the model itself for each year as it depends on all estimated coefficients, the income *y* and the copayment maximum *m*. Its computed value is quite large. The average price elasticity of demand, which multiplies this price elasticity with the fraction of the patient population that faces a positive out-of-pocket price, equals -0.40.

Table 5.2 Calibration of the patient models (1995)

Data:	privately insured patients	publicly insured patients
Copayment maximum (*m*)	euro 57	0
Demand volume per capita	0.175	0.276
Coefficient of variation	2	1.7
Insurance effect	1.2	
Income elasticity	0.088	
Parameters		
*	25759	25759
µ	7.7	8.2
(	$11.52\ 10^{-6}$	$11.52\ 10^{-6}$
F	1.28	1.17
Variables		
computed price elasticity	-0.69	

The model generates an average yearly increase in the number of treatments of medical specialists of about 2.2% over the period 1996 - 2000. Is this in line with the observed growth rate? To see this one needs to convert observed growth rates for admissions, outpatient treatments and outpatient visits into the rate of specialist treatments. The transformation requires data about developments in the number of treatments per admission, outpatient treatment and outpatient visit. Table 5.3 summarises. It follows that the average realised growth rate (2.4%) in specialist treatments is close to the rate generated by the model.

Table 5.3 Model outcomes and observed trends, 1996 - 2000

		average yearly growth rate 1996 - 2000
(i)	admissions	-1.7
	treatments per admission	1.5
	treatments during admissions	-0.2
(ii)	outpatient visits	-0.1
	treatments per outpatient visit	2.4
	treatments during outpatient visits	2.3
(iii)	outpatient treatments	5.6
	treatments per outpatient treatment	1.6
	treatments during outpatient treatments	7.1
(iv)	total specialist treatments (computed)	2.2
(v)	total specialist treatments (observed)	2.4

6 Evaluation of policy experiments

6.1 A base simulation

To study the impact of the financial reform, we perform a base simulation and then conduct a sensitivity analysis. In the latter, important parameters take values that are different from those that apply in the base simulation. Section 6.2 discusses the sensitivity analysis; this section discusses the base simulation.

The base simulation is fully determined by the empirical foundation discussed in the previous section. It is important to note that until 1997 fees for publicly insured patients were lower than the ones corresponding to privately insured patients. From 1997 onward, fees are equal. As we are mainly interested in the impact of the financial reform, in all simulations we have imposed the harmonisation already in the first year of the simulation, *i.e.* 1995; model outcomes refer to the year 2000.

We evaluate the values of interesting variables at 13 levels for the fee for services, which we vary from zero to euro 1080 in steps of euro 90. This proves more than sufficient to sign the effects of the financial reform, which reduces the fee-for-services from euro 360 to 0.[21] The income of medical specialists, which is uniform for all fee levels, derives from calculating the income that applies if both the fee equals euro 360 (which is close to the actual fee in 1995) and lumpsum subsidies to medical specialists are zero.[22] Note that this rule implies that at a zero fee the income of specialists is fully lumpsum financed. When the fee is equal to euro 360, her income is completely funded by the fee-for-service system. At larger fee values, lumpsum subsidies are negative and are reimbursed to patients (*i.e.* the consumer price is below the fee for services).

The lumpsum income is the same for all specialists, no matter which option (financial or ethical) they prefer. Consequently, if the response by ethical specialists and by financial specialists to the introduction of a specialist budget is different, the income of ethical specialists and financial specialists will change on account of the experiment. However, specialist income averaged over the two groups of medical specialists is a constant.

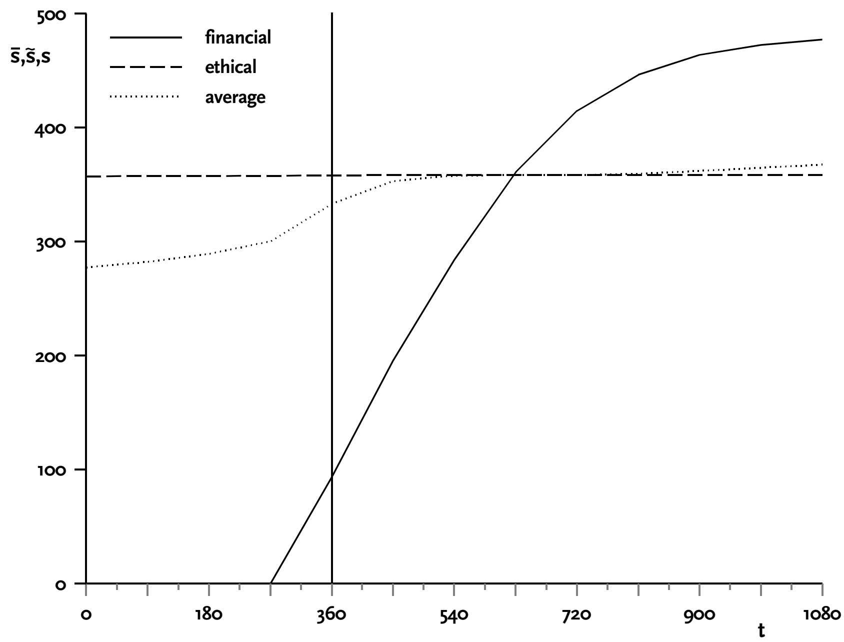
As noted above, we perform calculations using fee values increasing from zero to euro 1080 in steps of euro 90. Figures 6.1 to 6.9 indicate the impact of variations in the fee for specialist services on various variables that relate to specialist and patient behaviour.

[21] For a fee value of euro 360, the model mimics the situation before the financial reform. Note that this pre-reform situation differs slightly from the calibration year, as we perform our simulations for the year 2000.

[22] The fee value of euro 360 is obtained by dividing in the calibration year aggregate expenditure on medical specialist services by a volume index of medical specialist services. It thus averages the fees for a large number of services that are delivered by medical specialists.

We first analyse the supply curves of physicians that prefer the ethical or financial option. As Figure 6.1 shows, the supply curve of the ethical specialist is (almost) flat. As ethical specialists follow demand, there are no direct effects of changes in the fee for services, only indirect effects which operate through the demand for specialist services. Hence, the curve for the supply by ethical specialists reflects the demand curves of the publicly and privately insured (see Figures 6.4 and 6.5 below). The supply curve of financial specialists is more complex. It follows from our discussion of the physician model in section 4 that the supply of specialist services equals zero when the value of the fee *t* is low. When the fee is increased beyond the value of euro 270, supply becomes positive and increasing. Hence, the substitution effects of fee changes dominate the income effects. Due to the curvature of the utility function of the specialist, the increase of labour supply slows down for high values of the fee.

Figure 6.1 **Supply of total treatments**



Supply under the ethical option and supply under the financial option are equal at a fee value of about euro 630. Due to the equality of the former type of supply with demand, aggregate supply and demand for subsequent treatments coincide also at the same fee value of euro 630. Due to the assumption of proportional rationing, this also holds true for both types of insured (see Figures 6.4 and 6.5). As at this fee value, the labour supply of specialists under both options is the same, the utilities for the two options (exclusive ethical costs) also coincide. The consequence of this is illustrated in Figure 6.2 which shows the share of physicians who prefer the financial option. At $t = t^{*} =$ euro 630 ($t^{(}$ defined as the value for t for which supply equals demand), this share equals zero. The more the fee *t* deviates from $t^{(}$, the higher the yield of optimisation and the larger the number of specialists who prefer the financial option.

Figure 6.1 demonstrates that the aggregate supply of subsequent treatments is moderately increasing everywhere. That the slope of the aggregate supply curve is positive for low values of the fee, follows from the decline of the share of financial specialists (see Figure 6.2). This decline increases the weight of ethical specialists who supply a higher level of services. For higher fees, the aggregate supply curve becomes more steep, which reflects that the supply curve of financial specialists is upward sloping.

Figure 6.2 **Share of specialists that prefer the financial option**

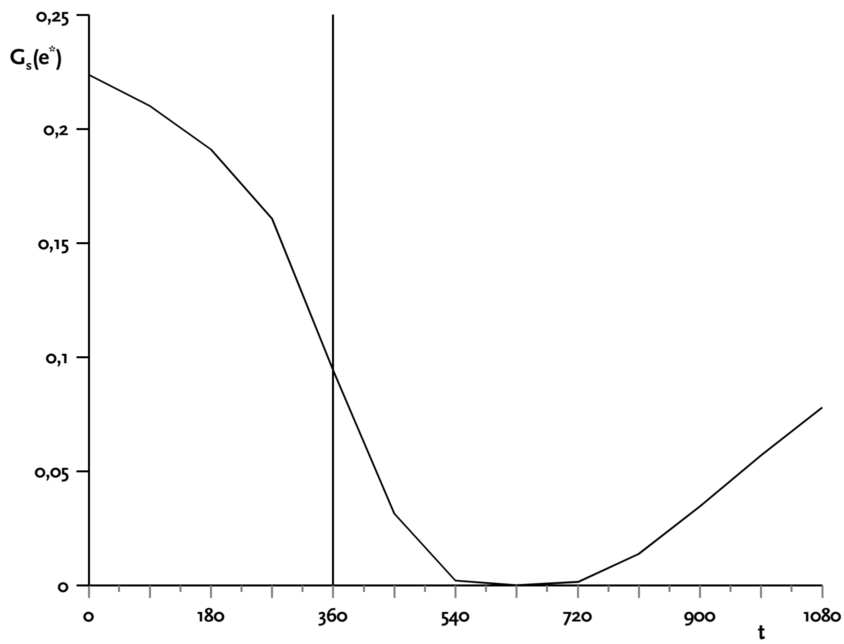


Figure 6.3 **Consumer price**

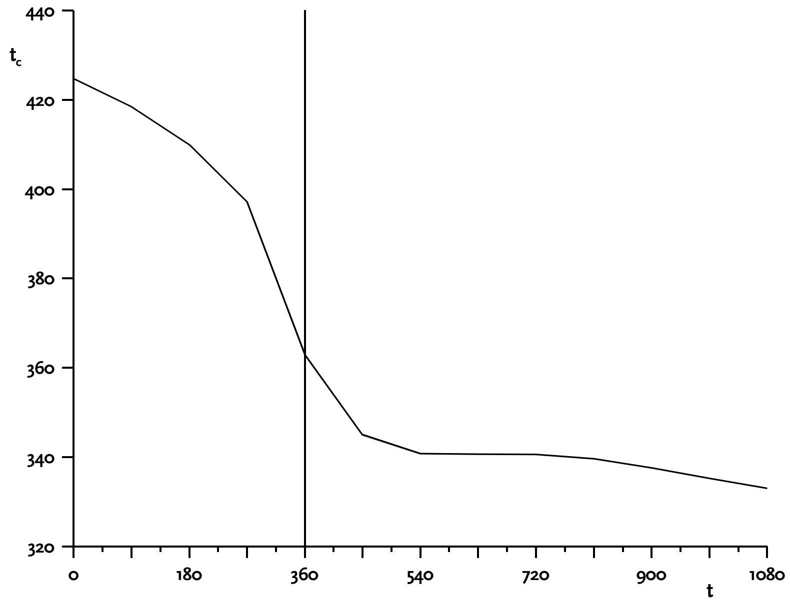


Figure 6.3 displays the relation between the consumer price and the producer fee for medical services. It is negatively-sloped, due to the fact that the supply of services is increasing in the producer fee. Indeed, in order to achieve that fee changes leave unaffected the income of specialists, the consumer price must decline when the volume of specialist services increases.

Figure 6.4 demonstrates that the demand of publicly insured patients is independent of the fee value. This is due to the absence of copayments in the public insurance scheme.

Figure 6.4 **Demand and supply: publicly insured**

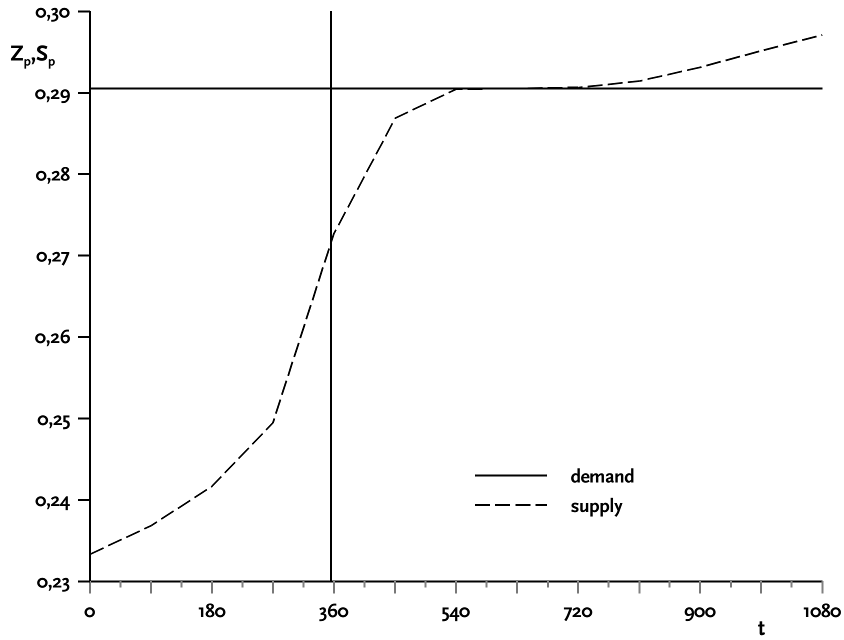


Figure 6.5 shows that the demand of the privately insured increases marginally when the fee increases. This does not mean that the slope of the aggregate demand curve Z is positive. Rather, the fact that an increase in the fee for services causes the consumer price to decline explains why demand is increasing in the fee for services.

Figure 6.6 shows the income of specialists. By construction, average income of financial and ethical specialists is independent of the fee for services. However, the same does not hold true for financial and ethical specialists taken separately. The reason is that the supply of services by financial and ethical specialists is different. Because a fee increase boosts the supply of services, it reduces lumpsum subsidies to specialists. As labour supply in the financial option equals zero for small fees, it thus decreases the income of financial specialists. Similarly, income of ethical specialists increases (as the average budget is constant).[23] For higher fees, the reverse holds true. As soon as labour supply of the financial specialist becomes positive, her income moves up. Now, income under the ethical option declines.

[23] This is an additional explanation for why fee increases decrease the number of specialists who prefer the financial option when the fee is small (see Figure 6.2).

Figure 6.5 Demand and supply: privately insured

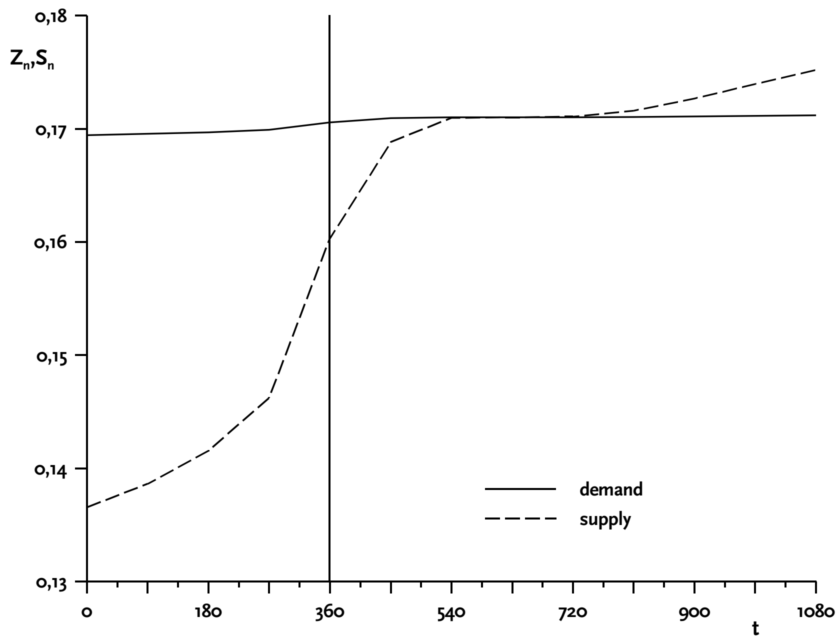
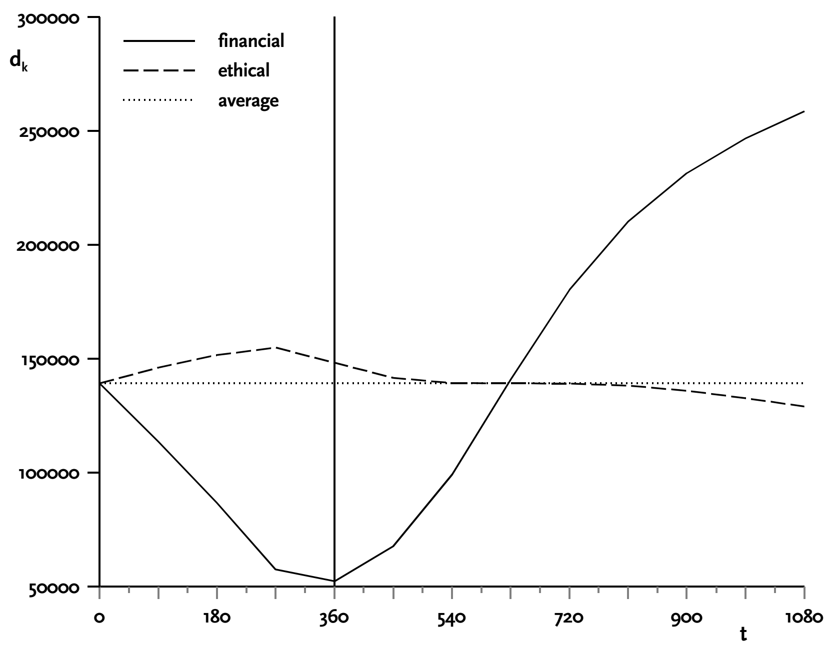


Figure 6.6 Income per specialist



Figures 6.7, 6.8 and 6.9 show the consumption-equivalent changes in specialist welfare, patient welfare and social welfare respectively.

Figure 6.7 Consumption-equivalent change of medical specialist welfare

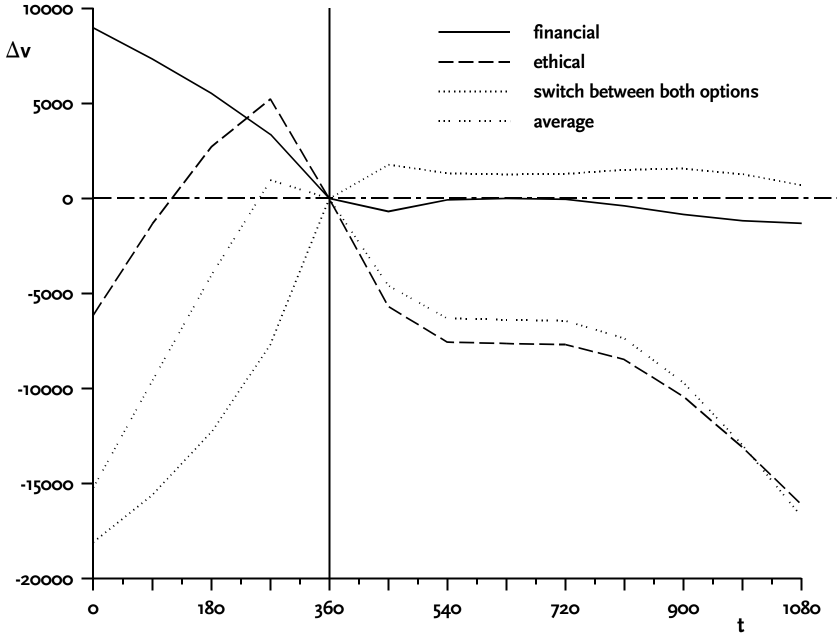


Figure 6.8 Consumption-equivalent change of patient welfare

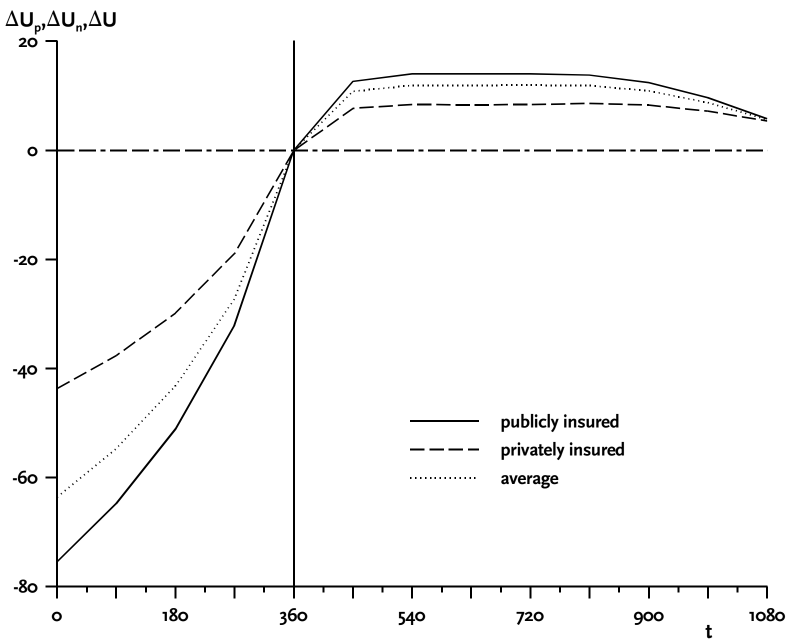
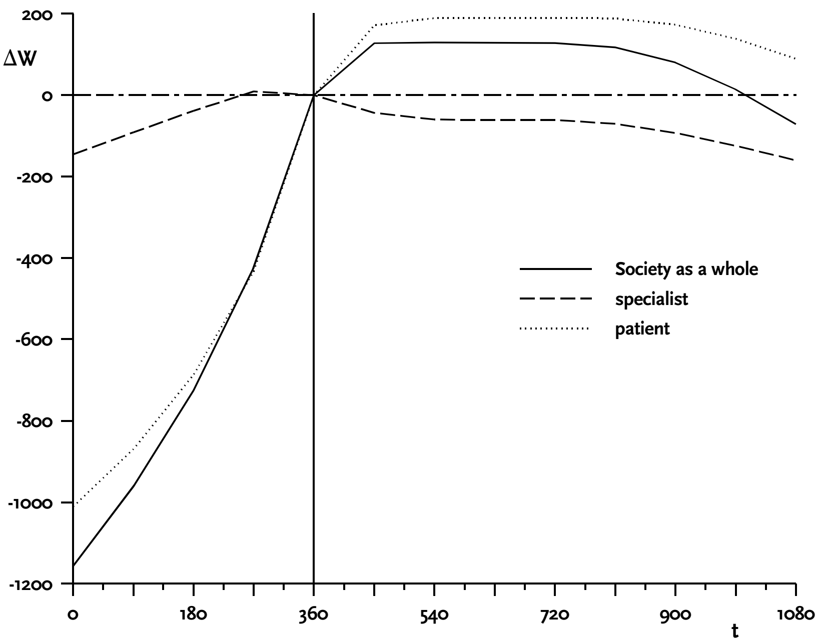


Figure 6.9 **Consumption- equivalent welfare changes**



To assess the effects upon patient welfare, recall that consumption of non-medical services is invariant to fee movements. Hence, changes in patient welfare fully reflect movements in health care consumption. The consumption of health services changes only because of changes in demand, movements in supply of the financial specialist and changes in the share of physicians who prefer the financial option. It appears that patient welfare reaches its peak value at the fee where the relative number of financial specialists is zero.

How can we explain that patient utility is highest when supply meets demand? Obviously, we would expect differences between supply and demand to be welfare-reducing, as the patient chooses his demand such as to maximise his utility. So in this case welfare losses due to imperfect agency are zero. However, other considerations suggest that welfare would be higher if supply were below demand, thereby correcting the excess consumption due to moral hazard. The reason why this does not occur lies in our financing rule for specialist income. The welfare gain from a smaller loss due to moral hazard is a lower health insurance premium and less copayments. However, our simulations keep the sum of premiums and copayments equal to the budget value. The latter is kept constant by adjusting the lumpsum income to compensate for fee changes. Hence, any welfare gain from the reduction of moral hazard is transmitted to the group of medical specialists.

Figure 6.7 demonstrates how aggregate specialist utility depends on the fee value. The utility function reaches a maximum at a fee value of euro 270. Combining the two welfare functions, that of patients and that of specialists, into an aggregate social welfare function, leads us to conclude that maximum social welfare is achieved for a fee of euro 540 (see Figure 6.9). In the optimum, the fee for services is thus below the level for which supply equals demand. Optimal supply policies reduce supply in order to compensate for the moral hazard effect in demand.

As Figure 6.9 shows, the introduction of a budget for medical specialists, which reduced the fee of euro 360 to nil, was welfare-reducing. Both patients and medical specialists saw their welfare decline. Patient welfare decreased because the policy change aggravated an inefficiency that already existed. In particular, patients suffered initially from a too low supply of medical services; the fee reduction only widened the gap between supply and demand. Medical specialists also suffered from the fee decrease. Although the financial reform has benefited physicians as it has increased consumption of leisure time, it has also pushed up ethical costs. The impact of the latter more than offsets the effect of the former.

6.2 Sensitivity analysis

As discussed, we have constructed our model such as to mimic as good as possible the Dutch system of care provided by medical specialists. This implies among other things that the parameter configuration upon which the above model calculations are based, fits closely the Dutch health care sector. Yet, it is obvious that in reality some parameters may take different values. To see whether deviations from our benchmark parameter configuration yield different results, this section conducts a sensitivity analysis.

In particular, this section focuses on three variables that we view as crucial for our numerical results: the price elasticity of supply, the price elasticity of demand and the income of medical specialists. Therefore, this section considers three types of calculations. The first one varies the price elasticity of labour supply by imposing 75% higher and lower values for the elasticity of substitution between leisure and consumption in the utility function of the specialist F. The second varies the price elasticity of demand by making similar changes in the value of (in the utility functions of the publicly and privately insured. Finally, the third examines the consequences of assuming that the budget of the specialist is based on an income linked to fee values of euro 180 and euro 540, respectively. The relation between initial supply and demand may also be relevant, as it determines the size of the supplier-induced demand distortion relative to the moral hazard distortion. However, we did not include this variable in our sensitivity analysis as it is an endogenous variable in our model.

We have deliberately specified the range of parameter values defined by these calculations as very wide and even wider than we judge as realistic. The reason is that if our conclusions pass this test, this strongly supports the conclusion from the base simulation.

(i) Changes in the elasticity of substitution F

In the base scenario the value of F is set at 6.91 (see table 5.1). Here we will analyse effects of changing the value of F to 3.9 (low variant) and 12.1 (high variant).

In the high variant, the fee value that minimises the number of financial specialists falls down from euro 630 to about euro 540. The same holds true for the fee that optimises the welfare function for both types of insured patients: it drops from euro 540 to euro 450. In the low variant (F = 3.9) it appears that the share of financial specialists does not reach its minimum at fee values lower than euro 1080. We observe the same phenomenon when we inspect social welfare. In this case, the model is always in a situation of excess demand. Of course, this is linked to the strong preference of medical specialists for leisure time relative to consumption.

We conclude that our results are quite sensitive to changes in the price elasticity of supply of medical specialists. However, our main conclusion that the introduction of local initiatives was welfare-reducing remains valid in these two cases.

(ii) Changes in the value of (

Changes in (affect the marginal utility of non-medical consumption and the value of demand if total expenditure is below the copayment maximum (see equation (4)). The base line value is $11.52\ 10^{-6}$; here, we take 75% lower and higher values.

In both variants, the absolute value of welfare changes. In particular, lower values of (yield higher values for welfare and vice versa. The same holds true for demand and supply as both demand and supply react to changes in (. Still, excess demand continues to be zero when the fee equals euro 630 (*i.e.* the same value as in the base scenario). The fee that optimises social welfare equals euro 450 in the low variant and euro 540 in the high variant, almost identical to its value in the benchmark case.

We therefore conclude that the results of the model are quite insensitive to changes in the marginal utility of non-medical consumption. Therefore, the changes made to the value of (do not affect our result that the financial reform was welfare-reducing.

(iii) Changes in the value of the specialist budget

The average budget per specialist in the base scenario amounts to euro 129,200, which corresponds to the average income in absence of a budget when the fee equals euro 360. In the low scenario the budget is reduced to euro 71,700 and in the high variant the budget is euro 219,500. These income values are linked to situations where the budget is zero and the fee equals 180 and 540 euros, respectively. Figures 6.10 and 6.11 highlight some differences between the three situations.

In figure 6.10 we see that the minimum of the fraction of specialists that prefer the financial option varies from fee values of (about) euro 640 (low variant) to euro 810 (high variant). Note that there will always be a fee value that makes this fraction exact equal to zero; but, as we evaluate the fee at discrete points, the graphs do not always indicate the exact minima.

Figure 6.10 Share of specialists that prefer the financial option at three different budget values

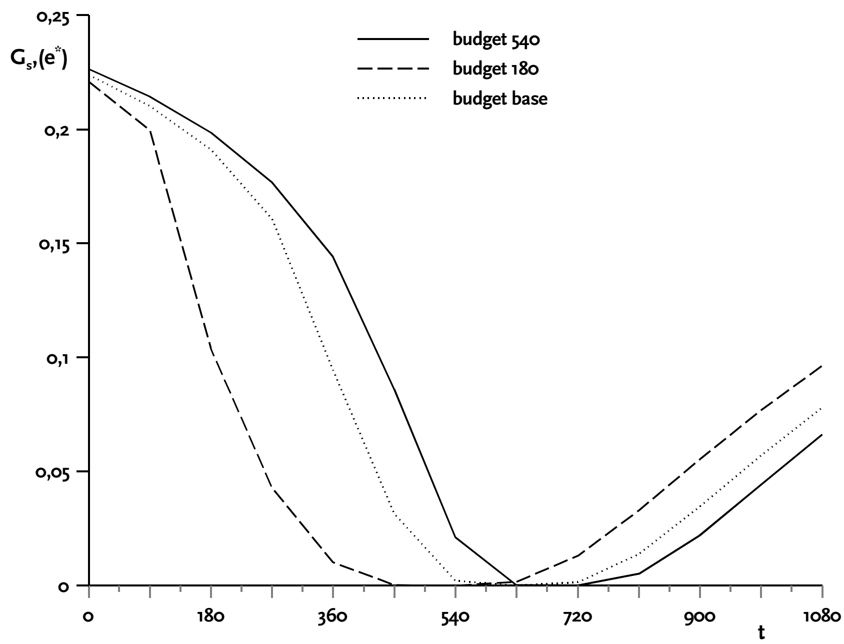
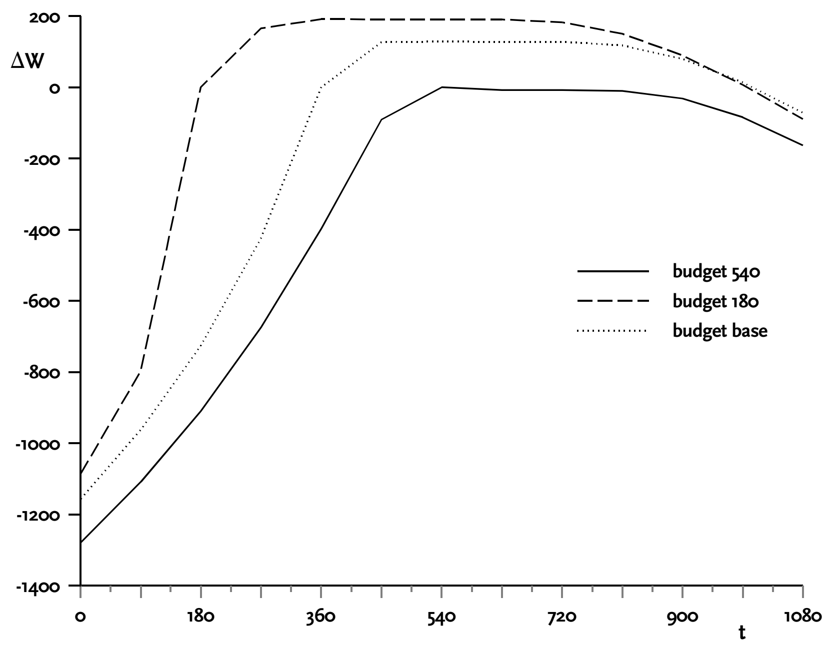


Figure 6.11 Consumption-equivalent change of social welfare at three budgets of specialists



It will be clear now that the optima of patient welfare in the two variants and the points of intersection of demand and supply are linked to each other. Patient welfare is higher when the budget is lower. This phenomenon is also linked to labour supply under the financial option; see Figure 6.11. The lower the budget, the sooner labour supply gets positive. For high values of the fee the supply curves converge: labour supply is bounded by the time constraint (14).

The optimal fee values differ only slightly more than in the two previous cases: euro 540 in case of a high budget and euro 360 in case of a low budget. Again, our main result that the shift in financing arrangement for medical specialists was welfare-reducing continues to hold true.

Table 6.1 summarises our findings.

Table 6.1 Results corresponding to 6 alternative parameter configurations

variant	optimal fee (patients)	optimal fee (specialists)	optimal fee (total)	welfare change from old to new
standard	720	270	540	–1160
high F	540	270	450	–1220
low F	1080	360	1080	–1560
high (	720	270	540	–1360
low (	720	270	450	–1010
high budget	810	360	540	–1280
low budget	630	180	360	–1090

The welfare change from the old to new system corresponds to a fee change of t = euro 360 to t= 0. In the budget variants, the initial fee values are euro 540 and euro 180, respectively.

7 Conclusions

This paper draws two conclusions that differ in their scope. The importance of the finding that the reduction of fees for specialist services in the Netherlands reduced social welfare is obvious. This result follows from the position of the initial fee for services, euro 360, relative to the level that optimises our social welfare measure, euro 540. As the policy change increases the deviation of the actual fee from its optimal level, it aggravates the welfare loss from a suboptimal financing scheme. Note however that our conclusion on the adverse welfare effect of the policy change does not hinge upon the position of the optimal fee. As our sensitivity analysis and Figure 6.9 clearly demonstrate, we would have found the same result when the optimal fee level had been somewhat below euro 360. Indeed, the steep decline of the social welfare measure in the range running from a zero fee to a fee of euro 360 indicates that our result on the welfare-decreasing nature of the shift in financing scheme is robust to minor changes of parameters or initial conditions.

A caveat is in order with respect to our definition of the policy change. We have defined the financing scheme that resulted after the reform as being fully lumpsum, *i.e.* with a zero services fee. Van den Berg and Mot (2000) report that in some hospitals, it is agreed that specialist budgets might be increased if specialist production grows faster than expected. To the extent that these agreements are credible, the financial reform may actually be perceived not as a switch from a fee-for-service scheme to a lumpsum scheme, but as a plain reduction of the fee for medical services. It can be seen from Figure 6.9 that, if this is true, our calculations overstate the impact of the change in financing scheme. However, this would not affect our finding that the introduction of a lumpsum budget for specialist services has reduced social welfare.

A second central result is that the optimal financing scheme features both a fee-for-service element and a budget element. This result may be more universal than the previous one. In particular, in order to be able to meet two targets, *i.e.* an efficient supply of medical services and a given income for medical specialists, one should have at least two instruments. A financing scheme that combines a fee-for-service element with a budget element provides two such instruments. Restricting the financing scheme to one instrument only will generally result in a suboptimal outcome.

Our analysis could be extended further. One option is to explore whether the financial reform may have induced specialists to engage in risk-selection strategies. Indeed, fee reductions may not only have induced medical specialists to reduce their supply of services, but may also have led them to intensify their efforts to dump high-risk consumers, *i.e.* supply no services at all to these patients (Ellis and McGuire (1986), Newhouse (1996)). If this is the case, the financial reform has had more adverse welfare effects than recognized in this paper.

References

Berg, B. van den, and E.S. Mot (2000), Honorering, verwijsgedrag en prikkels, *Economisch Statistische Berichten* 85, pp. 791-793 (in Dutch).

Canton, E.J.F. and E.W.M.T. Westerhout (1999), A Model for the Dutch Pharmaceutical Market, *Health Economics* 8, pp. 391-402.

Canton, E.J.F., C. Folmer and E.W.M.T. Westerhout (1999), *Reforming the Dutch Health Care Sector*, Paper presented at the IIPF Conference, August 24-27 1998, Cordoba (Argentina)

Duan, N., W.G. Manning, C.N. Morris and J.P. Newhouse (1982), *A Comparison of Alternative Models of the Demand for Medical Care*, RAND, R-2754-HHS, Santa Monica.

Ellis, R.P. and Th.G. McGuire (1986), Provider Behavior under Prospective Reimbursement - Cost Sharing and Supply, *Journal of Health Economics* 5, pp. 129-151.

Ellis, R.P. and Th.G. McGuire (1990), Optimal Payment Systems for Health Services, *Journal of Health Economics* 9, pp. 375-396.

Ellis, R.P. and Th.G. McGuire (1993), Supply-Side and Demand-Side Cost Sharing in Health Care, *Journal of Economic Perspectives* 7, pp. 135-151.

Evans, R.G. (1974), Supplier-induced Demand: Some Emperical Evidence and Implications, in M. Perlman (ed.), *The Economics of Medical Care*, MacMillan, London.

Feldman, R. and B. Dowd (1991), A New Estimate of the Welfare Loss of Excess Health Insurance, *American Economic Review* 81, pp. 297-301.

Folmer, C. (1998), Empirische invulling artsenmodellen (Empirical Elaboration of Physician Models), CPB, mimeo (in Dutch).

Folmer, C., J. Stevens, F. van Tulder and E.W.M.T. Westerhout (1997), Towards an Economic Model of the Dutch Health Care Sector, *Health Economics* 6, pp. 351-363.

Fuchs, V.R. (1978), The Supply of Surgeons and the Demand for Surgical Operations, *Journal of Human Resources* 13, Supplement, pp. 35-56.

Lee, Ch. (1995), Optimal Medical Treatment under Asymmetric Information, *Journal of Health Economics* 14, pp. 419-441.

Lindert, H. van, D. Delnoij, P.P. Groenewegen and S. Hofland (1999), Op weg naar een geïntegreerd medisch specialistisch bedrijf - managementparticipatie van medisch specialisten in Nederland, *Acta Hospitalia*, nr. 2, pp. 39-55 (in Dutch).

Manning, W.G. and M.S. Marquis (1996), Health Insurance: the Tradeoff between Risk Pooling and Moral Hazard, *Journal of Health Economics* 15, pp. 609-639.

McGuire, Th.G. and M.V. Pauly (1991), Physician Responses to Fee Changes with Multiple Payers, *Journal of Health Economics* 10, pp. 385-410.

Mot, E.S. (2001), The Influence of the Payment System for Medical Specialists upon Waiting Times in the Netherlands, mimeo.

Newhouse, J.P. and the Insurance Experiment Group (1993), *Free for All?, Lessons from the RAND Health Insurance Experiment*, Harvard University Press, Cambridge.

Newhouse, J.P. (1987), Medical Care Expenditure: A Cross-National Survey, *Journal of Health Economics* 6, pp. 109-128.

Newhouse, J.P. (1996), Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection, *Journal of Economic Literature* 34, pp. 1236-1263.

Rizzo, J.A. and D. Blumenthal (1994), Physician Labor Supply: Do Income Effects Matter?, *Journal of Health Economics* 13, pp. 433-453.

Rosett, R.N. and L. Huang (1973), The Effect of Health Insurance on the Demand for Medical Care, *Journal of Political Economy* 81, pp. 281-305.

Rutten, F.F.H. (1978), *The Use of Health Care Facilities in the Netherlands - An Econometric Analysis*, Doctoral Thesis, Erasmus University, Rotterdam.

Statistics Netherlands (2000), Vademecum of Health Statistics of the Netherlands, Voorburg .

Van Vliet, R.C.J.A. (1998), *Schatting verzekeringseffecten op basis van de CBS Gezondheidsenquete ten behoeve van Ramingsmodel Zorgsector,* Erasmus Universiteit Rotterdam.

Van Vliet, R.C.J.A. and H.G. van der Burg (1996), *Frequency Distribution for the Costs of Health Care Services* (in Dutch: *Verdelingsfuncties voor kosten van zorgvoorzieningen*), Instituut Beleid en Management Gezondheidszorg, Erasmus Universiteit Rotterdam.

Wedig, G., J.B. Mitchell and J. Cromwell (1989), Can Price Controls Induce Optimal Physician Behavior?, *Journal of Health Politics, Policy and Law* 14, pp. 601-620.

Westerhout, E.W.M.T and C. Folmer (1997), *Moral Hazard, Proportional Copayment Schemes and Deductibles,* Paper presented at the Sixth European Workshop on Econometrics and Health Economics, Lisbon.

Zeckhauser, R. (1970), Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives, *Journal of Economic Theory* 2, pp. 10-26.

Ziekenfondsraad (1998), *Evaluatie experiment specialistenhonorering,* rapport no. 783, Amstelveen (in Dutch).

Appendix A Critical values for the patient models

The expressions for $,^{(}$ and $,^{((}$ can be derived as follows. First we determine the expressions for the indirect utility functions corresponding to the two interior solutions (4) and (5) and the corner solution (3). To that end, we substitute the demand equations for medical services into the budget constraints to obtain similar expressions for *c*; subsequently, the expressions for z_i and *c* are substituted into the utility functions. Denote these three indirect utility functions by $u_1(,_i)$, $u_2(,_i)$ and $u_3(,_i)$, where subscript 1 refers to demand equation (4), subscript 2 to demand equation (5) and subscript 3 to the corner solution equation (3). $,^{(}$ can now be determined by setting $u_1(,_i)$ equal to $u_2(,_i)$. Similarly, $,^{((}$ can be derived by setting $u_1(,_i)$ equal to $u_3(,_i)$ and $,^{(((}$ follows by setting $u_2(,_i)$ equal to $u_3(,_i)$. We end up with the following expressions:

$$,^{(} \ ' \ \frac{\&<_b}{2<_a} \ \% \ \frac{1}{2<_a}\sqrt{(<_b)^2 \& 4<_a<_c} \tag{A1}$$

$$,^{((} \ ' \ <_b \tag{A2}$$

$$,^{(((} \ ' \ \sqrt{\frac{\&<_c \& \frac{1}{2}(<_b)^2}{<_a \% \frac{1}{2}}} \tag{A3}$$

where the auxiliary terms are defined as follows

$$<_a \ ' \ t_c^2 \frac{(}{2*} \tag{A4}$$

$$<_b \ ' \ t_c(1\&(y) \tag{A5}$$

$$<_c \ ' \ \&*m(1\&(y) \ \& \ \frac{1}{2}*(m^2 \ \& \ \frac{1}{2}t_c^2(1\&((y\&m))^2 \tag{A6}$$

The value of $m^{(}$ can be found by solving $,^{(} \ ' \ ,^{((} \ ' \ ,^{(((}$. It follows that:

$$m^{(} \ ' \ \frac{1 \ \& \ (y}{(}\left\{\sqrt{1 \ \% \ \frac{t_c^2(}{*}} \ \& \ 1\right\} \tag{A7}$$

Appendix B Welfare

This Appendix explains how the expected values and variances of important variables can be calculated, given that we assume the parameters $,_i$ and e_k to be lognormally distributed.

Some properties of the lognormal distribution

A stochastic variable *x* is said to be lognormally distributed if its logarithm, ln *x*, is normally distributed with parameters, say, μ en σ. The density function *g(.)* of *x* obeys:

$$g(x) = \frac{1}{x\sqrt{2\pi\sigma}} \exp\left(\frac{-(\log x - \mu)^2}{2\sigma^2} \right) \tag{B1}$$

And it follows that:

$$Pr[a \leq x \leq b] = G(b) - G(a) = \int_a^b g(x)\,dx = F\left(\frac{\ln b - \mu}{\sigma} \right) - F\left(\frac{\ln a - \mu}{\sigma} \right) \tag{B2}$$

where *F(.)* denotes the standard normal distribution function.

From (B2) it also follows that the mathematical expectation of *x*, *E(x)*, obeys:

$$E(x) = \int_0^\infty x g(x)\,dx = \exp\left(\mu + \tfrac{1}{2}\sigma^2\right) \tag{B3}$$

Using this expression we may calculate a number of (conditional) moments of *g:*

Conditional expectation:

$$E(x|a<x<b) = \int_a^b x g(x)\,dx = \left(F\left(\frac{\ln b - \mu}{\sigma} - \sigma \right) - F\left(\frac{\ln a - \mu}{\sigma} - \sigma \right) \right) E(x) \tag{B4}$$

Variance:

$$\begin{aligned} \sigma_x^2 &= E(x^2) - (E(x))^2 = \int_0^\infty x^2 g(x)\,dx - \exp\left(2\mu + \sigma^2\right) = \\ &= \exp\left(2\mu + \sigma^2\right)(\exp(\sigma^2) - 1) = (E(x))^2\,(\exp(\sigma^2) - 1) \end{aligned} \tag{B5}$$

Conditional variance:

$$E(x^2|a<x<b) - (E(x|a<x<b))^2 = (E(x))^2 \exp(\sigma^2)\left(\Phi\left(\frac{\ln b-\mu}{\sigma}-2\sigma\right)-\Phi\left(\frac{\ln a-\mu}{\sigma}-2\sigma\right)\right) - (E(x))^2\left(\Phi\left(\frac{\ln b-\mu}{\sigma}-\sigma\right)-\Phi\left(\frac{\ln a-\mu}{\sigma}-\sigma\right)\right) \tag{B6}$$

Coefficient of variation (the ratio of standard deviation and expectation):

$$C_v^2 = e^{\sigma^2} - 1 \tag{B7}$$

So the coefficient of variation of the lognormal distribution only depends on the variance of the corresponding normal distribution

The conditional *n*-th moment now equals:

$$E(x^n|a<x<b) = (E(x))^n \exp\left(\frac{1}{2}n(n-1)\sigma^2\right)\left(\Phi\left(\frac{\ln b-\mu}{\sigma}-n\sigma\right)-\Phi\left(\frac{\ln a-\mu}{\sigma}-n\sigma\right)\right) \tag{B8}$$

Average utility per patient

We may write average utility (see equation (1)) as follows:

$$E(u) = E(c) - \frac{1}{2}(E(c^2) + E(,_i x) - \frac{1}{2} * E(x^2) \tag{B9}$$

This expression can be rewritten as:

$$E(u) = E(c) - \frac{1}{2}((E(c))^2 + E(,_i x) - \frac{1}{2} * (E(x))^2 - \frac{1}{2}((\sigma_c^2 + * \sigma_x^2) \tag{B10}$$

with σ_c and σ_x the standard deviations of non-medical consumption *c* and realised medical consumption volume *x*.

The expected value of non-medical consumption

Non-medical consumption is a stochastic variable as copayments are stochastic. The latter are linked to the distribution of need across patients. Formally,

$$
\begin{aligned}
E(c) &= E(y^g - pr - cp) = y^g - pr - E(cp) \\
E(cp) &= \int_0^m t_c x g(x)\,dx + \int_m^\infty m g(x)\,dx = \\
&\Phi\left(\frac{\ln m - \mu}{\sigma} - \sigma\right) t_c x + \left(1 - \Phi\left(\frac{\ln m - \mu}{\sigma}\right)\right) m
\end{aligned}
\tag{B11}
$$

where m and t_c have been defined in the main text already and y^g denotes gross income.

The variance of non-medical consumption equals that of copayments *cp*. We derive an expression for the expectation of the square of *cp,* following equations (B2) and (B8). The computation of the variance can then easily be computed from formula (B5).

$$
\begin{aligned}
E(eb^2) &= \int_0^m (t_c x)^2 g(t_c x)\,dt_c x + \int_m^\infty m^2 g(t_c x)\,dt_c x = \\
&(t_c E(x))^2 \left\{ (1 + VC^2)\, \Phi\left(\frac{\ln m - \mu}{\sigma} - 2\sigma\right) \right\} + m^2 \left\{ 1 - \Phi\left(\frac{\ln m - \mu}{\sigma}\right) \right\}
\end{aligned}
\tag{B12}
$$

Expected value of medical consumption

We specify the demand for first treatments per patient, *zf,* as a fixed share in total demand per patient z (see equation (10)):

$$
\begin{aligned}
z_p^f &= \phi_p z_p \\
z_n^f &= \phi_n z_n
\end{aligned}
\tag{B13}
$$

where the subscripts (*p*, *n*) refer to publicly and privately insured patients, respectively.

By definition consumption x^f equals demand z^f. Consumption of subsequent treatments per insured person x^s is a weighted sum of the demand for subsequent treatments z^s and corresponding supply. Let $G_s(e)$ denote the distribution function of ethical costs e and e^* the value of e at which the specialist is indifferent between the ethical and financial option. Now expected consumption of total consumption for publicly and privately insured patients equals:

$$
\begin{aligned}
E(x_p) &= E(z_p^f) + E(s_p) = E(z_p^f) + (1 - G_s(e^*))\, E(z_p^s) + G_s(e^*)\, E(\bar{s}_p) = \\
&\phi_p E(z_p) + (1 - G_s(e^*))(1 - \phi_p^p)\, E(z_p) + G_s(e^*)\, E(\bar{s}_p) = \\
&\left(\phi_p + (1 - G_s(e^*))(1 - \phi_p)\right) E(z_p) + G_s(e^*)\, E(\bar{s}_p) \\
E(x_n) &= \left(\phi_n + (1 - G_s(e^*))(1 - \phi_n)\right) E(z_n) + G_s(e^*)\, E(\bar{s}_n)
\end{aligned}
\tag{B14}
$$

Supply of subsequent treatments per specialist under the financial option $\bar{s}$ follows from equation (16). Aggregated supply is allocated to both types of insured using shares in expected demand:

$$
\begin{aligned}
E(s_p) \; ' \; E(\bar{s}) \; \frac{N_s}{N_p} \; \frac{E(Z_p)}{E(Z_p)\%E(Z_n)} \\
E(s_n) \; ' \; E(\bar{s}) \; \frac{N_s}{N_n} \; \frac{E(Z_n)}{E(Z_p)\%E(Z_n)}
\end{aligned}
\qquad \text{(B15)}
$$

with N_s the number of specialists, and N_p and N_n the numbers of publicly and privately insured patients. The allocation of total supply to both types of patients is not unique as the model only generates total supply $\bar{s}$; formula (B15) therefore expresses a rule of thumb needed to obtain supply per type of insured patient.

The next step is to compute expected medical consumption *E(x)*. Note that x is a linear function of demand z, which in turn depends on the parameter $,_i$. As a result, expected consumption consists of an aggregate of conditional first and second order moments of the stochastic variable $,_i$. We will derive the expression in two steps.

The expected value of $,_i$ *z*. If it holds that $,_i^{((} < ,_i^{(}$, then it follows that:

$$
\begin{aligned}
E(,_i z) \; ' \; \Bigg(\; F\left(\frac{\ln ,_i^{(} \& \mu}{F} \& F\right) \& F\left(\frac{\ln ,_i^{((} \& \mu}{F} \& F\right) \Bigg) \\
\frac{\& t_c(1\&(y) \;\; E(,_i)}{* \;\%\; (t_c^2} \;\; \% \\
\Bigg(\; F\left(\frac{\ln ,_i^{(} \& \mu}{F} \& 2F\right) \& F\left(\frac{\ln ,_i^{((} \& \mu}{F} \& 2F\right) \Bigg) \frac{\exp(F^2)\big(\; E(,_i)\big)^2}{* \;\%\; (t^2} \;\; \% \\
\Bigg(1\& \;\; F\left(\frac{\ln ,_i^{(} \& \mu}{F} \& 2F\right) \Bigg) \frac{\exp(F^2)\big(\; E(,_i)\big)^2}{*}
\end{aligned}
\qquad \text{(B16)}
$$

To compute *E(,$_i$ x)* we need expressions (B14). It follows that:

$$
E(,_p x_p) \;\; ' \;\; E(,_p z_p) \; \big(0_p\%(1\&G_s(e^{(}))(1\&0_p)\big) \; \% \; E(,_p) \;\; G_s(e^{(})\,E(\bar{s}_p) \qquad \text{(B17)}
$$

$$
E(,_n x_n) \;\; ' \;\; E(,_n z_n) \; \big(0_n\%(1\&G_s(e^{(}))(1\&0_n)\big) \; \% \; E(,_n) \;\; G_s(e^{(})\,E(\bar{s}_n) \qquad \text{(B18)}
$$

Note that we have used the assumption of stochastic independence of $,_i$ en *e*.

As x is a linear function of z its variance is a linear function of the variance of z. We derive an expression for the latter.

The expectation of the square of z is equal to:

$$E(z^2) = \int_0^\infty z^2 f(z)\,dz = \int_0^{\varepsilon_i^{\tau\tau}} (z(\varepsilon_i))^2 g(\varepsilon_i)\,d\varepsilon_i + \int_{\varepsilon_i^{\tau\tau}}^{\varepsilon_i^{\tau}} (z(\varepsilon_i))^2 g(\varepsilon_i)\,d\varepsilon_i + \int_{\varepsilon_i^{\tau}}^{\infty} (z(\varepsilon_i))^2 g(\varepsilon_i)\,d\varepsilon_i \equiv I_1 + I_2 + I_3 \tag{B19}$$

where:

$$I_1 = 0 \tag{B20}$$

$$I_2 = \int_{\varepsilon_i^{\tau\tau}}^{\varepsilon_i^{\tau}} \left(\frac{-t_c(1-\tau y)+\varepsilon_i}{\alpha+\tau t_c^2}\right)^2 g(\varepsilon_i)\,d\varepsilon_i = \left(\frac{-t(1-\tau y)}{\alpha+\varepsilon_i t_c^2}\right)^2 \int_{\varepsilon_i^{\tau\tau}}^{\varepsilon_i^{\tau}} g(\varepsilon_i)\,d\varepsilon_i$$

$$-\frac{2t_c(1-\tau y)}{(\alpha+\tau t_c^2)^2}\int_{\varepsilon_i^{\tau\tau}}^{\varepsilon_i^{\tau}} \varepsilon_i g(\varepsilon_i)\,d\varepsilon_i + \int_{\varepsilon_i^{\tau\tau}}^{\varepsilon_i^{\tau}} \frac{\varepsilon_i^2}{(\alpha+\tau t_c^2)^2} g(\varepsilon_i)\,d\varepsilon_i =$$

$$\left(\frac{-t(1-\tau_c y)}{\alpha+\tau_c t_c^2}\right)^2 \left\{F\left(\frac{\ln\varepsilon^{\tau}-\mu}{\sigma}\right) - F\left(\frac{\ln\varepsilon^{\tau\tau}-\mu}{\sigma}\right)\right\} + \tag{B21}$$

$$E(\varepsilon_i)\ \frac{-2t_c(1-\tau y)}{\left(\alpha+\tau t_c^2\right)^2}\left\{F\left(\frac{\ln\varepsilon^{\tau}-\mu}{\sigma}-\sigma\right) - F\left(\frac{\ln\varepsilon^{\tau\tau}-\mu}{\sigma}-\sigma\right)\right\} +$$

$$\frac{\left(E(\varepsilon_i)\right)^2(\exp(\sigma^2))}{\left(\alpha+\tau t_c^2\right)^2}\left\{F\left(\frac{\ln\varepsilon^{\tau}-\mu}{\sigma}-2\sigma\right) - F\left(\frac{\ln\varepsilon^{\tau\tau}-\mu}{\sigma}-2\sigma\right)\right\}$$

and:

$$I_3 = \frac{1}{\alpha^2}\int_{\varepsilon_i^{\tau}}^{\infty} \varepsilon_i^2 g(\varepsilon_i)\,d\varepsilon_i =$$

$$\left(\frac{E(\varepsilon_i)}{\alpha}\right)^2 \left(\exp(\sigma^2)\right)\left\{1 - F\left(\frac{\ln\varepsilon^{\tau}-\mu}{\sigma}-2\sigma\right)\right\} \tag{B22}$$

Equations (B19) - (B23) enable us to compute the variance of total demand z. Finally, from equation (B14) it follows that the variance of x equals:

$$\sigma_x^2 = \left(\eta + (1 - G_s(e^*))(1-\eta)\right)^2 \sigma_z^2 \tag{B23}$$

As both types of insured have distinct values for the parameter η corresponding variances of total medical consumption also differ.

Expected utility of specialists can directly be derived from equation (30) in section 4.4. The conditional expectation of the ethical costs directly follows from equation (B13).

Appendix C Consumption-equivalent welfare changes

To calculate social welfare, we have to bring the utility changes of patients and specialists on an equal footing. We use ΔU_p and ΔU_n to denote the utility changes of the populations of the publicly and privately insured respectively. For specialists, we have to define four categories: for specialists can opt for either one out of two options before and after the policy change. The utility change for specialists that choose for the ethical option before and after the policy change reads as $\tilde{v} - \tilde{v}|_0$ where $\tilde{v}$ is defined in equation (23) and the index 0 refers to the initial situation, *i.e.* before the policy change. Similarly, specialists who remain on the financial option undergo a utility change $\bar{v} - \bar{v}|_0$ with $\bar{v}$ also defined in equation (23). The expression for the utility change of specialists who choose to switch to the financial option are a little bit more complicated as the specialist bears ethical costs only after the policy change: $\bar{v} - \tilde{v}|_0 - E(e \mid e \leq e^*)$. Similarly, the utility change undergone by specialists who switch to the ethical option reads as $\tilde{v} - (\bar{v}|_0 - E(e \mid e \leq e^*|_0))$. Note that the latter two categories cannot exist simultaneously. Either some specialists switch to the ethical option or some specialists switch to the financial option; two-way traffic cannot occur due to the one-dimensionality of physician heterogeneity.

To be able to compare these policy changes, we convert them into consumption equivalents, defined as the equivalents in terms of the non-medical consumption good. This procedure bears resemblance to the concept of Pareto efficiency as we analyse whether the utility change of the winners of the policy change is sufficiently large to compensate the losers. Actually, we calculate a sort of equivalent variation, the only difference being that the transfer from the winners to the losers is in the form of non-medical consumption goods rather than income and takes place ex post rather than ex ante. Hence, we use the direct utility function to calculate consumption equivalents rather than the indirect utility function used by the measure of equivalent variations.

In calculating consumption equivalents, we linearize around the initial position. For patients, we abstract from the heterogeneity with respect to health status and assume that all patients have the average health status. Using equation (1), the consumption equivalents of the utility change of the publicly and privately insured can then be calculated as follows:

$$\Delta U_p = \frac{N_p(u_p - u_p|_0)}{1 - ({}_p c_p|_0} \tag{C1}$$

$$\Delta U_n \equiv \frac{N_n(u_n - u_n|_0)}{1 - ({}_n c_n|_0} \tag{C2}$$

u and c thus refer to the patient with health status $E(,)$. The consumption equivalents of the utility changes of the four categories of medical specialists can be calculated using the utility function (13). First, we define:

$$\bar{A} = \frac{N_s}{"}\left(\frac{\bar{d}|_0}{("\,\bar{d}|_0^{-D} + \bar{l}|_0^{-D})^{-1/D}}\right)^{1+D} \tag{C3}$$

and

$$\tilde{A} = \frac{N_s}{"}\left(\frac{\tilde{d}|_0}{("\,\tilde{d}|_0^{-D} + \tilde{l}|_0^{-D})^{-1/D}}\right)^{1+D} \tag{C4}$$

To calculate the consumption-equivalent welfare change for the whole population of specialists, we have to aggregate over all possible values of the ethical cost variable. The change in social welfare can then be calculated as the sum of the consumption-equivalent welfare changes for the two groups of insured and the consumption-equivalent welfare change for the population of medical specialists. We have to distinguish between the case where the population fraction that opts for the financial option increases ($e^* > e^*|_0$), equation (C5):

$$\begin{aligned} \Delta W = {} & \Delta U_p + \Delta U_n + \tilde{A}\,(\tilde{v} - \tilde{v}|_0) \int_{e^*}^{\infty} dG_s(e_k) + {} \\ & \bar{A}\,(\bar{v} - \bar{v}|_0) \int_{0}^{e^*|_0} dG_s(e_k) + \tilde{A} \int_{e^*|_0}^{e^*} (\bar{v} - \tilde{v}|_0 - e_k)\, dG_s(e_k) \end{aligned} \tag{C5}$$

and the case where this population fraction declines ($e^* < e^*|_0$), equation (C6):

$$\begin{aligned} \Delta W = {} & \Delta U_p + \Delta U_n + \tilde{A}\,(\tilde{v} - \tilde{v}|_0) \int_{e^*|_0}^{\infty} dG_s(e_k) + {} \\ & \bar{A}\,(\bar{v} - \bar{v}|_0) \int_{0}^{e^*} dG_s(e_k) + \bar{A} \int_{e^*}^{e^*|_0} (\tilde{v} - \bar{v}|_0 + e_k)\, dG_s(e_k) \end{aligned} \tag{C6}$$

Note that the consumption-equivalents depend on the (conditional) expectations of the ethical cost variable e.

EARLIER PUBLISHED CPB DISCUSSION PAPERS AND RESEARCH MEMORANDA

CPB Discussion Papers

001. De Groot, H.L.F.
On the Optimal Timing of Reductions of CO_2 Emissions: An economists' perspective on the debate on 'when flexibility'; June 2001.
ISBN: 90-5833-073-7

002. Ter Rele, H.J.M. and Van Steen, E.L.T.
Housing subsidisation in the Netherlands: Measuring its distortionary and distributional effects; June 2001.
ISBN: 90-5833-075-3

003. Ruud A. de Mooij and Sjef Ederveen
Taxation and foreign direct investment: A synthesis of empirical research; November 2001
ISBN: 90-5833-081-8

004. Richard Nahuis and Ashok Parikh
Factor Mobility and Regional Disparities: East, West, Home's Best?; January 2002
ISBN: 90-5833-093-1

005. Khaled Moussa Diaw and Joeri Gorter
The remedy may be worse than the disease: A critical account of The Code of Conduct; February 2002.
ISBN: 90-5833-094-X

Research Memoranda

147. Graafland, J.J.
Analyzing the macro economic impact of child care subsidies: an AGE approach; January 1999.
ISBN: 90-5833-004-4

148. Nieuwenhuijzen, V. van
Collapsing controlled exchange rates: how well can we predict and explain them?; January 1999.
ISBN: 90-5833-005-2

149. Tang, P.J.G., B. Jacobs and R. Nahuis
Sectoral productivity growth and R&D spillovers in the Netherlands; February 1999.
ISBN: 90-5833-007-9

150. Oers, F.M. van, R.A. de Mooij, J.J. Graafland and J.Boone
An earned income tax credit in the Netherlands: simulations with the MIMIC model; March 1999.
ISBN: 90-5833-008-7

151. Minne, B., and H.G.A. Noordman
Kwaliteitsstijging Nederlandse export in 1988-1996?; June 1999.
ISBN: 90-5833-009-5

152. Tang, P.J.G., and Adrian Wood
Globalisation, co-operation costs, and wage inequalities; June 1999.
ISBN: 90-5833-011-7

153. Lejour, A.M., and P.J.G. Tang
The informal sector: a source of growth; June 1999.
ISBN: 90-5833-012-5

154. Wiel, H.P. van der
Loondifferentiatie in Nederland na 1969: een sectorale invalshoek; June 1999.
ISBN: 90-5833-013-3

155. Koopmans, C.C., D.W. te Velde, W. Groot and J.H.A. Hendriks
NEMO: Netherlands Energy demand MOdel; a top-down model based on bottom-up information; June 1999.
ISBN: 90-5833-014-1

156. Broer, P., N. Draper and F. Huizinga
The equilibrium rate of unemployment in the Netherlands; July 1999.
ISBN: 90-5833-015-X

157. Douven, R.C.M.H.
Equilibrium rates and wage flexibility in Europe; August 1999.
ISBN: 90-5833-018-4

158. Wiel, H.P. van der
Sectoral Labour Productivity Growth: A growth accounting analysis of Dutch industries, 1973-1995; September 1999.
ISBN: 90-5833-021-4

159. Wiel, H.P. van der
Firm turnover in Dutch business services: the effect on labour productivity; December 1999.
ISBN: 90-5833-025-7

160. Stegeman, H.
Individual remuneration; February 2000.
ISBN: 90-5833-029-X

161. Boone, J. and R.A. de Mooij
Tax policy in a model of search with training; February 2000.
ISBN: 90-5833-031-1

162. Don, F.J.H.
Forecasting in macroeconomics: a practitioner's view; February 2000.
ISBN: 90-5833-033-8

163. Koning, P.W.C. and M. van Leuvensteijn
Duration Dependence in Unemployment Insurance and Social Assistance: Consequences of Profiling for the Unemployed; April 2000.
ISBN: 90-5833-038-9

164. Creusen, H.P.W.A. and B. Minne
Falling R&D but stable investments by oil companies: why?
A study on R&D and investment in fixed assets in the oil industry; May 2000.
ISBN: 90-5833-040-0

165. Mannaerts, H.J.B.M.
STREAM: Substance Throughput Related to Economic Activity Model. A partial equilibrium model for material flows in the economy; May 2000.
ISBN: 90-5833-042-7

166. Jongen, E.L.W., E. van Gameren and J.J. Graafland
The impact of active labor market policies: an AGE analysis for the Netherlands; May 2000.
ISBN: 90-5833-043-5

167. Lejour, A.M. and P.J.G. Tang
The differential impact of the South on wage inequality in the North; June 2000.
ISBN: 90-5833-046-X

168. Lejour, A.M. and R. Nahuis
Openness, Growth and R&D spillovers: Uncovering the Missing Link?; July 2000.
ISBN: 90-5833-048-6

169. Nahuis, R. and J.A. Smulders
The rising Skill Premium, Technological change and Appropriability; August 2000.
ISBN: 90-5833-049-4

170. Ewijk, C. van and P.J.G. Tang
Efficient progressive taxes and education subsidies; September 2000.
ISBN: 90-5833-050-8

171. Douven, R.
Regulated competition in health insurance markets; November 2000.
ISBN: 90-5833-051-6

172. Gorter, J. and A. Parikh
How mobile is capital within the European Union? November 2000.
ISBN: 90-5833-054-0

173. Leuvensteijn, M. van and P.W.C. Koning
The effects of Home-ownership on Labour Mobility in The Netherlands: Oswald's these revisited; February 2001.
ISBN: 90-5833-058-3

174. Hakfoort, J.R. and J.C. Weigand
Commercial Publishing - A quiet life? Market power and performance on the Dutch market for consumer magazines; February 2001.
ISBN: 90-5833-060-5

HOW TO ORDER

To order *the Central Economic Plan, the Macro Economic Outlook, CPB Report and other Special Publications* published by Sdu, please contact:

Sdu Publishers
P.O. Box 20014
2500 EA The Hague
Telephone : +31 - 70 - 378 98 80
Telefax : +31 - 70 - 378 97 83

Special Publications published by Koninklijke De Swart, can be ordered at:

Koninklijke De Swart
P.O. Box 53184
2505 AD The Hague
Telephone : +31 - 70 - 308 21 21
Telefax : +31 - 70 - 308 21 59

'FKSEC, a macro-econometric model for the Netherlands' and 'Challenging Neighbours' can only be ordered by your local bookseller.

CPB Documents, CPB Discussion Papers, Working Papers, Research Memoranda, Monographs and the Special Publications 'Scanning the future', 'Netherlands in Triplo (only in Dutch)' and Aviation growth within environmental preconditions (only in Dutch)' can be ordered at:

CPB Library
P.O. Box 80510
2508 GM The Hague
Telephone : +31 - 70 - 338 34 25
Telefax : +31 - 70 - 338 33 50
Email : library@cpb.nl

Are you interested in data or background information? Visit our Internet site!
The address is: http://www.cpb.nl.