Health care expenditure in the last months of life

Stefan Felder a,*, Markus Meier b, Horst Schmitt a

a Health Economics Department, Faculties of Medicine and Economics, Otto-von-Guericke University Magdeburg, Leipziger Strasse 44, 39120 Magdeburg, Germany
b Department of Economics, Socioeconomic Institute, University of Zurich, Zurich, Switzerland

Received 12 May 1997; received in revised form 11 May 1999; accepted 6 January 2000

Abstract

In OECD countries, a considerable share of health care expenditure (HCE) is spent for the care of the terminally ill. This paper derives the demand for HCE in the last 2 years of life from a model that accounts for age, mortality risk and wealth. The empirical tests are based on data of deceased members of a major Swiss sick fund. The empirical evidence confirms most of the hypotheses derived from the model, i.e., (i) HCE increases with closeness to death, (ii) for retired individuals, HCE decreases with age, and (iii) low-income individuals, as compared to high-income individuals, incur lower HCE in the last months of life. © 2000 Elsevier Science B.V. All rights reserved.

JEL classification: I12; I31; D61

Keywords: Mortality risk; Willingness to pay for survival; Demand for health care services

1. Introduction

The cost of treating the terminally ill is sometimes blamed for the steady increase in health care costs. Indeed, the so-called cost of dying is substantial. As Lubitz and Riley (1993) report for the United States, 30% of the total Medicare
budget is paid out on behalf of persons in their last year of life. Spending per decedent is about seven times larger than per survivor. For Switzerland, health care expenditure (HCE) of the terminally ill is comparably high, as samples of two health insurance companies show (cf. Zweifel et al., 1999). Payments for persons in their last year of life constitute 22% and 18% of the total health care cost of retired persons, depending on the sample. The average ratio of per capita expenditure between decedents and survivors equals 5.6:1.\(^1\)

Besides its size, the path of HCE in the last year of life is exceptional as well. One observes a steady increase, accentuated by a surge in the last quarter of life. In the study by Lubitz and Riley (1993), 62% of all Medicare costs referring to the last year of life were incurred in the last quarter, compared to 10% in the first quarter. In Switzerland, the difference is less accentuated (42–48% as opposed to 14–18%, depending on the sample (cf. Zweifel et al., 1999; Table 1).

A third characteristic of HCE in the last months of life is age dependency. Baker et al. (1995) reported that cost in the last 2 years of life for persons who died at 101 years of age or older was 37% of those incurred by persons who died at 70. In the Swiss samples, we found a similar pattern with respect to age and HCE in the last 2 years of life.

The present paper presents a model designed to explain these three characteristics of HCE in the last months of life. It derives the demand for health care services in the last months of life from a life-cycle model that formalises the trade-off between utility of living and utility of consumption. Section 2 deduces the willingness to pay for a change in the mortality schedule and shows how willingness to pay is affected by age, wealth and the risk of death. Section 3 presents the data and the hypotheses referring to the impact of age, closeness to death, wealth and insurance coverage on HCE. Section 4 discusses the empirical results, and Section 5 concludes.

\section{The model}

We assume that consumption and age are the only arguments in the utility function of a representative household. By \(u(c(t))\), we denote the utility of being alive at age \(t\), given consumption \(c\). The twice differentiable utility function \(u\), determined up to a multiplicative factor, is positive, increasing and strictly concave. Furthermore, we assume that there is no bequest motive, and that a

---

\(^1\) The Swiss figures are lower compared to the US ones because Swiss hospitals are heavily subsidised by the government. For that reason, payments by the insurance companies for inpatient care do not cover the total cost incurred. Since most people die in hospital, the actual (relative) cost of dying will be higher and thus close to the US Medicare figures.
household of age \(a\) chooses its consumption path by maximising expected residual lifetime welfare

\[
EU(a) = \int_a^{\infty} u[c(t)] p_a(t) e^{-r(t-a)} dt, \tag{1}
\]

where \(r \geq 0\) is the rate of time preference assumed to be equal to the interest rate. The conditional probability \(p_a(t)\) of surviving until age \(t\), given that the household has survived until age \(a\), is a function of the age-specific (positive) death rates \(q(t)\):

\[
p_a(t) = e^{-\int_a^t q(s) ds}. \tag{2}
\]

Substituting Eq. (2) for \(p_a(t)\) in Eq. (1) shows that the forces of mortality increase the effective rate of time preference. The household discounts future utility more strongly because there is a positive probability that it will not survive. Suppose that there is a market for actuarially fair annuities so that the lifetime budget constraint can be written as (cf. Yaari, 1965):

\[
w_a + \int_a^{\infty} (l(t) - c(t)) p_a(t) e^{-r(t-a)} dt = 0, \tag{3}
\]

where \(l(t) \geq 0\) is labour income at age \(t\) and \(w\) denotes non-human wealth at age \(a\). With the purchase of an annuity, the household insures against the risk of longevity and achieves an increase in welfare. The optimal consumption path follows from the optimal control problem

\[
\max_c \text{EU}(a)
\]

subject to Eq. (3).

The household will trade human and non-human wealth against an annuity that guarantees a fixed consumption stream until the date of death. Along the optimal consumption path, marginal utility \(u'[c(t)]\) is equalised across periods:

\[
u'[c(t)] = u'[c(a)] \quad \forall t \geq a. \tag{4}
\]

Due to the properties of \(u\), the optimal consumption path \(c\) is also constant in time. Hence, the solution \(c\) and the performance functional \(\text{EU}(a)\) are given by

\[
c = \frac{w_a + \lambda_a}{\mu_a} \quad \text{and} \quad \text{EU}(a) = u(c) \mu_a, \tag{5}
\]

respectively. Here, the discounted remaining life expectancy \(\mu_a\) and human wealth \(\lambda_a\) are defined as

\[
\mu_a := \int_a^{\infty} p_a(t) e^{-r(t-a)} dt \quad \text{and} \quad \lambda_a := \int_a^{\infty} l(t) p_a(t) e^{-r(t-a)} dt. \tag{6}
\]

In what follows, we are interested in a change of mortality, i.e., in an increasing mortality risk, occurring, for instance, due to a particular sickness. This change is
caused by an infinitesimally small positive change $\delta q(t)$ of the hazard rate $q(t)$. Due to Eq. (2), we find\footnote{See Rosen (1988, Section 3.3).}

$$\frac{\delta p_a(t)}{\delta q(t)} = -p_a(t) \int_a^t \delta q(s) \, ds < 0, \quad (7)$$

which ensures that the probability to survive decreases. When the mortality schedule changes, the household will adjust its consumption path, and expected residual lifetime welfare will change. First, we derive for the expected marginal lifetime welfare of wealth at age $a$:

$$\frac{\partial EU(a)}{\partial w_a} = u', \quad (8)$$

a result that is not dependent on the choice of $\delta q(t)$. Furthermore, we find from Eq. (5)

$$\frac{\delta c}{\delta q(t)} = \frac{1}{\mu_a} \left( \frac{\delta \lambda_a}{\delta q(t)} - c \frac{\delta \mu_a}{\delta q(t)} \right), \quad (9)$$

and

$$\frac{\delta EU(a)}{\delta q(t)} = u' \left[ \frac{\delta \lambda_a}{\delta q(t)} + \left( \frac{u}{u'} - c \right) \frac{\delta \mu_a}{\delta q(t)} \right]. \quad (10)$$

A change in the mortality schedule affects expected lifetime welfare through the change in non-human wealth $\delta \lambda_a/\delta q(t)$ and in life expectancy $\delta \mu_a/\delta q(t)$, respectively. The latter is weighted by $(u/u') - c$, the consumer’s surplus from being alive in any year with consumption level $c$.

The wealth equivalent of the change in lifetime welfare due to a change in the mortality schedule, $WE(a)$, is defined as the marginal rate of substitution between $w_a$ and $q(t)$:

$$WE(a) := \frac{\delta EU(a)}{\delta q(t)} = \frac{\delta \lambda_a}{\delta q(t)} + \left( \frac{u}{u'} - c \right) \frac{\delta \mu_a}{\delta q(t)}. \quad (11)$$

Since the utility function $u = u(c)$ is assumed to be normalised (cf. Rosen, 1988), we have $u'(c) < (u(c)/c)$ for all $c > 0$, therefore,

$$WE(a) \leq 0. \quad (12)$$
The wealth equivalent of an increase in mortality risk is negative. Households will thus have a positive willingness to pay for health care services, which restore the original mortality schedule.

For the empirical work, we are interested in three first derivatives of the wealth equivalent function \( WE(a) \), namely, \( \frac{\partial WE(a)}{\partial w_a} \), \( \frac{\partial WE(a)}{\partial q(t)} \) and \( \frac{\partial WE(a)}{\partial a} \). The result

\[
\frac{\partial WE(a)}{\partial w_a} = -\frac{\delta \mu_a}{\delta q(t)} \frac{u u' c}{u'^2} \frac{\partial c}{\partial w(a)} < 0
\]

(13)

follows from \( \delta c/\delta w_a = 1/\mu_a > 0 \) and was first derived by Jones-Lee (1976).

For the derivative of \( WE(a) \) with respect to a change in the mortality schedule, we derive

\[
\frac{\delta WE(a)}{\delta q(t)} = \frac{\delta^2 \lambda_a}{\delta q(t)^2} - \frac{u u' c}{u'^2} \frac{\delta c}{\delta q(t)} \frac{\delta \mu_a}{\delta q(t)} + \left( \frac{u}{u'} - c \right) \frac{\delta^2 \mu_a}{\delta q(t)^2}.
\]

(14)

For the change of \( WE(a) \), when age increases, we find

\[
\frac{\partial WE(a)}{\partial a} = \frac{\partial}{\partial a} \left( \frac{\delta \lambda_a}{\delta q(t)} \right) + \left( \frac{u}{u'} - c \right) \frac{\partial}{\partial a} \frac{\delta \mu_a}{\delta q(t)}.
\]

(15)

The sign of the two derivatives (Eqs. (14) and (15)) cannot be determined a priori without fixing a special choice of the perturbation function \( \delta q(t) \). In the literature, there are in principle two different approaches. In Arthur (1981), Rosen (1988), and Shepard and Zeckhauser (1984), the Dirac mass function \( \delta q(t) := \delta(t) \), which forces an increasing mortality risk mainly at age \( a \), is analysed.

We choose the change of the mortality risk in the sense of Johansson (1996) by a permanent change of the hazard rate. We set

\[
\delta q(t) := q(t),
\]

which might be interpreted as a parametric change \( \partial / \partial \gamma \) of \( q(t, \gamma) := (1 + \gamma) q(t) \). Additionally, we assume that we are dealing with the Gompertz’s model, i.e.,

\[
q(t) = ae^{bt} \quad (\alpha, \beta > 0).
\]

For this model, we can show that \(^3\)

\[
\frac{\delta WE(a)}{\delta q(t)} < 0
\]

(16)

holds, provided that \( w_a \geq 0 \) and \( l(t) \leq 0 \) for all \( t \geq a \). Hence, if labour earnings remain constant or fall and non-human wealth is non-negative, willingness to pay for the original mortality schedule will therefore increase when the risk of death

\(^3\) See Lemma 1 in Appendix A for a detailed proof.
This result is intuitively appealing and follows from the fact that an increase in mortality risk shortens the remaining life expectancy, which, in turn, increases the level of consumption, given that the two conditions hold. The strict concavity of the utility function then yields the desired result.

On the other hand, if non-human life is negative or \( l'(t) \) increases, it can not be excluded that an increase in mortality risk decreases the consumption level so that the value of life decreases. This adds another example to the list of unexpected results in the literature on the value of life (cf. Bergstrom, 1982; Rosen, 1988).

The sign of \( \delta \text{WE}(a)/\delta a \) is also governed by the path of labour income. In Appendix A, Lemma 2 shows that \( l'(t) \leq 0 \) and the positiveness of \( (\partial/\partial a)(\delta \mu_a/\delta q(t)) \) imply the positiveness of \( (\partial/\partial a)(\delta \lambda_a/\delta q(t)) \). In this case, we therefore obtain

\[
\frac{\delta \text{WE}(a)}{\delta a} > 0. \tag{17}
\]

Hence, if labour earnings remain constant or fall, \( \text{WE}(a) \) increases with age. Since the value of \( \text{WE}(a) \) is negative, this means that the absolute value of \( \text{WE}(a) \) decreases with age. When a person ages, his/her remaining life expectancy diminishes. Total consumer surplus, as well as the value of life will thus decrease with age, provided the labour income path is not decreasing. The absolute value of \( \text{WE}(a) \) may increase if labour income is growing with age. This will most likely be the case in middle age. In short, one expects that willingness to pay for life saving follows an inverted U-shaped profile over the life cycle (see also Jones-Lee et al., 1985).

It should also be mentioned that the result (Eq. 17) remains true if we use the Dirac mass function \( \delta(t) \) instead of \( q(t) \) itself for increasing the mortality risk (cf. Rosen, 1988).

Finally, it remains unclear whether the function \( (\partial/\partial a)(\delta \mu_a/\delta q(t)) \) can be positive. For this, we choose \( r := 0.02 \) and use the empirical data of Johansson

![Fig. 1. (a) Discounted remaining life expectancy \( \mu_a \). (b) The change in remaining life expectancy \( (\partial/\partial a)(\delta \mu_a/\delta q(t)) \).](image)
(1996), where $\alpha = 0.000081$ and $\beta = 0.087$ were proposed. As we see, for age classes above 55, a positive sign can be observed (see Fig. 1b).

**Remark 1.** The results (Eqs. (16) and (17)) are also true under less restrictive assumptions on the function $l$. It is sufficient that the behavior of $l$ is not too exotic, i.e., $l'(t)$ is positive in $[a,T]$ and negative in $[T,\infty)$ where $T$ should not be too large. The corresponding mathematics, which is very technical, is presented in an appendix that is available from the authors upon request.

### 3. Data and hypotheses

The present study is based on individual data of 415 deceased persons from a sample of more than 6000 members of a major Swiss health insurance company. The data include HCE, gender and age, as well as specifics of the insurance policy of the individuals. A majority of the individuals (346) were aged 65 or more at the time of their death (see Table 1). For simplicity, for each individual the expenditure records covering the last 2 years of life were aggregated into eight quarterly observations. This led to 3320 observations in the full sample and 2768 in the subsample of patients who died aged 65+, respectively. The distribution of this quarterly HCE proved to be highly skewed to the left. For this reason, for one set of estimations we excluded non-positive observations from the sample, transforming the remaining observations into logarithms. A sample selection test indicated that the inverse of Mill’s ratio ($\lambda$) has to be included as an additional variable in the regression (see Table 2).

A classical result of health economics is the finding that demand for health care services depends on the form of the insurance contract (cf. Pauly, 1968). With full

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Data</th>
<th>All individuals</th>
<th>Age at death 65+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of individuals</strong></td>
<td></td>
<td>415</td>
<td>346</td>
</tr>
<tr>
<td>Female</td>
<td>203</td>
<td>177</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>212</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td><strong>Period of observation</strong></td>
<td></td>
<td>1987 I–1992 IV</td>
<td></td>
</tr>
<tr>
<td>Cases of death</td>
<td></td>
<td>1985 I–1992 IV</td>
<td></td>
</tr>
<tr>
<td>Quarterly HCE</td>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Number of quarterly HCE (observed for each individual)</td>
<td>3320</td>
<td>2768</td>
<td></td>
</tr>
<tr>
<td>Total number of observations</td>
<td>2770</td>
<td>2392</td>
<td></td>
</tr>
<tr>
<td>$&gt; 0$</td>
<td>550</td>
<td>376</td>
<td></td>
</tr>
</tbody>
</table>
insurance coverage, the consumer price of medical services is zero, leading to an increase in demand compared to a situation without insurance. Usually, difficulties arise when deriving a relationship between willingness to pay and demand for health care services, since — abstracting from transaction costs — every positive willingness to pay is met when insurance companies pay the full bill. However, the patients in our sample have insurance policies that include either a franchise or a proportional co-insurance rate. These risk sharing schemes imply positive consumer prices for health care services and guarantee that persons with very low willingness to pay will have zero demand. Consequently, a negative relationship between price and demand for health care services can be postulated.

The previous section has shown that under certain conditions, willingness to pay for changing the mortality schedule increases with the risk of death. Though the data available do not give any information on the health status of the patient, and in particular do not include any diagnosis code indicating the risk of death, it seems fair to assume that health status deteriorates and the risk of death increases over the last quarters of life. Moreover, in those cases where death does not accidentally occur, one can expect patients to consider costs and benefits when deciding on health care services. This leads to

**Hypothesis 1.** HCE in the last months of life increases with closeness to the time of death.

Hypotheses 2 and 3 follow from the profile of willingness to pay for change in the mortality schedule over the life cycle. When labour earnings cease to increase, willingness to pay for survival decreases with age. For retired people, willingness to pay for life-saving health care services is therefore likely to decrease with age. On the other hand, young people’s willingness to pay for survival increases with age. From that, we derive

**Hypothesis 2.** HCE in the last months of life is higher for young (64-) than for retired persons (65+).

**Hypothesis 3.** Within the age class 65+, HCE in the last months of life decreases with age.

Some individuals in the two samples receive government transfers targeted to reduce health insurance premiums for low-income households. Our data include information on subsidised premiums, allowing us to differentiate between low- and high-income classes and to derive

**Hypothesis 4.** Since willingness to pay for survival increases with wealth, low-income patients have lower HCE in the terminal months of life than high-income patients.
The insurance policies show different coverage with respect to inpatient care. While most patients have coverage for general hospital care, some have supplementary insurance that guarantees preferred treatment in the so-called private section of hospitals. Because the cost of treatment for a person receiving privileged treatment is higher and the hospital can charge higher fees, this gives rise to

**Hypothesis 5.** Supplementary hospital insurance results in higher HCE in the terminal months of life.

In order to estimate the relationship between HCE in the last 2 years of life and the different variables mentioned in Hypotheses 1–5, the following equation was specified

\[
\frac{\text{HCE}}{\log \text{HCE}} = \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 \text{SexF} + \beta_4 \text{ASexF} + \beta_5 D_{65} + \beta_6 \lambda
\]

\[
+ \beta_7 \text{Subs} + \beta_8 \text{Ins} + \sum_{q=1}^{7} \gamma_q \text{Quar}_q + \sum_{t=1986}^{1992} \delta_t \text{Year}_t + \epsilon. \quad (18)
\]

Therein \( A \) denotes calendar age in years and \( A^2 \) the squared value of age (to consider possible non-linearities). SexF is a dummy for sex (= 1, if female; = 0, if male). The product ASexF considers possible interactions between age and sex in the determination of HCE. \( D_{65} \) is again a dummy variable, taking on a value of unity, if a person is 65 or older (= 0, otherwise). \( \lambda \) is the inverse of Mill’s ratio, capturing the impact of potential sample selection when excluding zero observations. Subs is a dummy indicating that an individual health insurance premium is subsidised by the government (= 0, otherwise), Ins denotes a dummy indicating an insurance policy providing for privileged treatment in the hospital (= 0, otherwise), Quar\( q \) a dummy taking on the value of one in quarter \( q \) before death (= 0, otherwise, with \( q = 8 \) serving as the benchmark), and, finally, Year\( t \), is a dummy taking on the value of one in year \( t \) (= 0, otherwise, with \( t = 1985 \) as the benchmark year), and \( \epsilon \) is the error term.

4. Estimation results

Table 2 presents the estimated coefficients and their standard errors (in parentheses) as well as general statistics of the regressions for both the total sample and the elderly subsample, which excludes individuals aged 64 and less at their date of death. The first set of results refers to the regressions where quarterly HCE in the last 2 years of life is the dependent variable. The second regression refers to the logarithms of HCE and implements a two-stage approach (Heckman, 1989) to capture the effect of including only positive observations.
### Table 2

OLS regression results for real quarterly HCE \( a \).

<table>
<thead>
<tr>
<th>Dep. Var.</th>
<th>HCE ( b )</th>
<th>log(HCE) ( b )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All individuals</td>
<td>Individuals 65+</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>353.7 (1343.1)</td>
<td>12657 (2094.4)</td>
</tr>
<tr>
<td>( A )</td>
<td>-9.21 (38.35)</td>
<td>-323.4* (0.060)</td>
</tr>
<tr>
<td>( A^2 )</td>
<td>0.11 (0.26)</td>
<td>2.07* (1.00)</td>
</tr>
<tr>
<td>SexF</td>
<td>1169.1** (642.2)</td>
<td>1849.9 (1481.7)</td>
</tr>
<tr>
<td>ASexF</td>
<td>-11.81 (8.54)</td>
<td>-20.0 (17.9)</td>
</tr>
<tr>
<td>Subs</td>
<td>-292.2* (31.03)</td>
<td>-325.2* (141.1)</td>
</tr>
<tr>
<td>Ins</td>
<td>438.3* (0.061)</td>
<td>547.3* (149.4)</td>
</tr>
<tr>
<td>( D_{65} )</td>
<td>622.10* (281.48)</td>
<td>-</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Quar1</td>
<td>2554.5* (226.4)</td>
<td>2721.9* (225.3)</td>
</tr>
<tr>
<td>Quar2</td>
<td>784.5* (222.7)</td>
<td>898.5* (251.0)</td>
</tr>
<tr>
<td>Quar3</td>
<td>456.6* (220.9)</td>
<td>432.7* (248.9)</td>
</tr>
<tr>
<td>Quar4</td>
<td>237.7* (219.7)</td>
<td>194.5 (247.4)</td>
</tr>
<tr>
<td>Quar5</td>
<td>451.5* (218.1)</td>
<td>460.7* (245.4)</td>
</tr>
<tr>
<td>Quar6</td>
<td>475.3* (215.9)</td>
<td>391.4* (243.2)</td>
</tr>
<tr>
<td>Quar7</td>
<td>123.0 (214.9)</td>
<td>96.6 (241.8)</td>
</tr>
<tr>
<td>Year1986</td>
<td>48.59 (370.8)</td>
<td>-49.36 (420.5)</td>
</tr>
<tr>
<td>Year1987</td>
<td>707.9* (360.4)</td>
<td>640.6* (407.1)</td>
</tr>
<tr>
<td>Year1988</td>
<td>199.0 (354.6)</td>
<td>158.5 (397.7)</td>
</tr>
<tr>
<td>Year1989</td>
<td>430.1 (354.0)</td>
<td>422.2 (396.0)</td>
</tr>
<tr>
<td>Year1990</td>
<td>629.5* (350.8)</td>
<td>681.8* (392.9)</td>
</tr>
<tr>
<td>Year1991</td>
<td>1089.9* (367.3)</td>
<td>1208.3* (412.7)</td>
</tr>
<tr>
<td>Year1992</td>
<td>1386.9* (445.4)</td>
<td>1597.5* (506.7)</td>
</tr>
<tr>
<td>Adj. ( R^2 )</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td>( F )-value</td>
<td>16.71</td>
<td>16.09</td>
</tr>
<tr>
<td>( N )</td>
<td>3320</td>
<td>2768</td>
</tr>
</tbody>
</table>

\( a \): Deflated by the Swiss consumer price index (1981 = 100).

\( b \): Standard errors in parentheses.

* Significantly different from zero at the 99% confidence level.

** Significantly different from zero at the 95% confidence level.

Closeness to death (Quar\(_1\) to Quar\(_r\)) has a decisive effect on the level of HCE. In the sample with all individuals, expenditure in the last quarter is 2550 Swiss francs above expenditure in the benchmark quarter. In the two stage estimation, the difference in the HCE between the 1st and the 8th quarter is 578% (\( e^{1.939} - 0.50(221)^2 - 1 \)). Moreover, the pattern of coefficients conforms very much with expectation, showing a clear increase with closeness to death. The largest expenditure surge occurs from the second last to the last quarter. Since one expects

\( \text{for a correct interpretation of a dummy variable in semilogarithmic equations, see Kennedy (1981).} \)
clear signs indicating near death in the last quarter of life at the latest, those results support Hypothesis 1, which is based on the reasoning that willingness to pay for life-saving health care services increases with the risk of death.

Hypotheses 2 and 3 postulate lower expenditure for the elderly and a negative correlation between age \( A \) and HCE in the subsample of individuals aged 65+ . Our sample shows higher expenditure for the elderly. The coefficient is significant, contradicting Hypothesis 2. Given that we have a very small number of young decedents in the sample, it is not clear how representative this result is.\(^5\) There is evidence from other studies that the cost of dying is higher for young than for elderly persons (Emanuel and Emanuel, 1994; Lamers and Van Vliet, 1998).

The coefficients for age show the expected signs. Age has a negative, although decreasing effect on HCE. These findings confirm Hypothesis 3. Table 2 also shows a significant effect of \( \text{SexF} \). In the subsample, a 65-year-old woman has an HCE of 1850 francs in excess of that of a man. This differential declines to 1450 at age 85, in view of the negative coefficient of \( A\text{SexF} \). It is interesting to connect the sign of the age coefficients with the life expectancy difference between men and women. As it is well known, life expectancy of women is higher at every age than that of men, but the difference decreases at old age. The effect of an increasing age on the willingness to pay for life is determined by the remaining life expectancy. Hence, we would expect that (i) women c.p. have a higher demand for life-saving health care services than men and (ii) the gender difference shrinks with increasing age; a pattern that is confirmed by the regression.

The parameter of the variable \( \text{Subs} \) has a negative sign, indicating that patients with subsidised health insurance premiums have lower HCE in the last 2 years of life. Since premium subsidies are paid to low-income households, this result is in line with Hypothesis 4, which postulates that willingness to pay for life increases with wealth. However, the difference is only significant in the regression, which includes zero observations, where the difference amounts to 290 and 325 Swiss francs, depending on the sample.

Patients who have a supplementary hospital insurance policy (\( \text{Ins} = 1 \)) incur significantly higher HCE \( (e^{0.806 - 0.500.205} - 1 = 119\%) \). Given the fact that most people die in a hospital (the hospital share in total costs during the last year of life amounts to about 80% in Switzerland, according to the present sample), this is not surprising. Even if hospitals were to apply the same technology to patients who have differing insurance coverage, they still have the right to charge higher fees for the treatment of insured patients who have supplementary coverage. This result clearly vindicates Hypothesis 5.

Finally, the year dummies in the regression indicate that cost increases in the health care sector exceed economy-wide inflation. In 1992, real HCE of decedents

---

5 The lack of HCE data for accident victims may bias the result since younger persons are more prone to accidental death.
was 2 times \(e^{0.724 - 0.5(0.239)^2} = 2.00\) higher than in the benchmark year 1985, probably reflecting technological changes in medicine.

5. Conclusion

HCE of persons in the last months of life shows a clear increase towards death. Moreover, there is evidence that cost of dying is higher for young than for elderly persons (Emanuel and Emanuel, 1994; Lamers and Van Vliet, 1998). Finally, Baker et al. (1995) found, while studying Medicare payments, that HCE in the last 2 years of life decreases with age. All three observations can be explained by the theory of life saving (Schelling, 1968). Willingness to pay for survival over the life cycle is hump shaped, increasing at young ages, peaking in middle age and declining thereafter. From that one derives a willingness to pay for life-saving health care services to be an increasing (decreasing) function of age at young (old) ages. An increasing demand for health care services as death approaches concurs with the hypothesis that willingness to pay increases with the risk of death.

The present paper studies HCE in the last eight quarters of life of 415 members of a Swiss health insurance company who died in the period 1987–1992. The estimations show closeness to death to be a decisive factor for HCE. Regarding the difference in the cost of dying between young and retired persons, the Swiss sample shows higher cost for the elderly, contradicting the hypothesis. In the subsamples of individuals aged 65+, there is a significant decrease in cost as age increases. Furthermore, income and the extent of insurance coverage have a significant impact on HCE in the last 2 years of life. Patients with supplementary insurance for hospital treatment incur higher cost of dying than patients with average insurance coverage. Finally, low-income households spend less on health care services in the last 2 years of life than high-income households.

The high cost of dying is one reason why critics of high-tech medicine in the industrialised world ask for a rationing of health care services according to the age of patients (cf. Callahan, 1987). The result of the present study, namely that HCE in the last 2 years of life depends on the health insurance contract, hints at alternatives to rationing. For instance, from a coinsurance rate that increases with age according to the risk of death, a dampening effect on the ever rising health care cost might be expected (cf. Felder, 1997).

The present paper may be criticised in focusing too much on the demand side of the health care sector. One could argue that the demand for health care services, in particular, if it stems from terminally ill persons, are to a large extent determined by the preference of the providers of treatment. However, an adequate theory of supply induced demand for life-saving health care services would have not only to explain that physician exploit the possible lack of consumer sovereignty but also the effect of the patients’ characteristics such as age, gender and survival
probability on the physicians' decisions — a task that does not seem promising. We propose an alternative way to incorporate the supply side and argue that physicians have paternalistic preferences, i.e., that they follow rules that are socially desirable. The theory from which optimal physicians' decisions under the paternalistic preference assumption can be derived is, of course, the value of life concept we have presented in Section 2.

The data used in the present study could serve as a base for further investigations. From a comparison of the amount of life-saving health care services demanded at different ages, willingness to pay for survival could be derived. It would be interesting to compare the corresponding findings with existing results in literature, for instance derived from the risk premium paid in jobs with different death risks, respectively (see Rosen and Thaler, 1975 and Viscusi, 1992).

Acknowledgements

The paper was presented at the inaugural conference of the International Health Economics Association, Vancouver, Canada, May 19–23, 1996, and the Annual Meeting of the Verein of Socialpolitik, Kassel, September 25–27, 1996. We thank the participants as well as two anonymous referees for their helpful comments, and the Swiss National Science Foundation for financial support under grant 4032-035660.

Appendix A

In order to complete Eqs. (16) and (17), respectively, we must prove the following two statements.

Lemma 1. Let \( l'(t) \leq 0 \) for all \( t \geq a \) and \( w_a \geq 0 \). Then

\[
\frac{\delta \text{WE}(a)}{\delta q(t)} < 0.
\] (19)

Cf. Detsky et al. (1981), who studied the correlation between the predicted short-term survival probabilities of patients admitted to an intensive care unit and the expenditure for these patients. The spending was largest for patients whom physicians expected to live but did not, and for patients whom physicians expected to die but did not. Newhouse (1992) noticed that these findings are consistent with rational sequential choice under uncertainty.
Proof. For the Gompertz’s model, i.e., \( q(t) = ae^{\beta t} \), the function \( p_a(t) \), given by Eq. (2), satisfies the differential equation \( \dot{p}_a(t) = -\beta p_a(t) \int p_a(s)ds \). Therefore, we find by integrations by parts

\[
\frac{\delta \mu_a}{\delta q(t)} = \int_a^\infty \frac{\delta p_a(t)}{\delta q(t)} e^{-r(t-a)} dt = \frac{1}{\beta} \left( \mu_a(q(a) + r) - 1 \right) < 0. \tag{20}
\]

It follows that

\[
\frac{\delta \mu_a}{\delta q(t)^2} = \frac{q(a) + r}{\beta} \frac{\delta \mu_a}{\delta q(t)} < 0. \tag{21}
\]

Similarly

\[
\frac{\delta \lambda_a}{\delta q(t)} = \int_a^\infty \frac{\delta p_a(t)}{\delta q(t)} l(t)e^{-r(t-a)} dt
\]

\[
= \frac{1}{\beta} \left( \lambda_a(q(a) + r) - l(a) \right) - \frac{1}{\beta} \lambda_a^r \leq 0, \tag{22}
\]

where

\[
\lambda_a^r := \int_a^\infty p_a(t) l'(t)e^{-r(t-a)} dt,
\]

which implies

\[
\frac{\delta \lambda_a}{\delta q(t)^2} = \frac{q(a) + r}{\beta} \frac{\delta \lambda_a}{\delta q(t)} - \frac{1}{\beta} \int_a^\infty \frac{\delta p_a(t)}{\delta q(t)} l'(t) e^{-r(t-a)} dt \leq 0. \tag{23}
\]

In the next step, we show that \( \delta c/\delta q(t) \) is positive. Since \( \delta c/\delta q(t) \) is given by

\[
\frac{1}{\mu_a} \int_a^\infty \frac{\delta p_a(t)}{\delta q(t)} (l(t) - c) e^{-r(t-a)} dt
\]

it is sufficient to show that

\[
\int_a^\infty \frac{\delta p_a(t)}{\delta q(t)} (l(t) - c) e^{-r(t-a)} dt > 0
\]

holds. Note that the function \( v_1(t) := -\int_a t \delta q(s)ds \) is negative and strictly decreasing. Furthermore, we denote \( v_2(t) := p_a(t)(l(t) - c)e^{-r(t-a)} \). Since \( \int_a^\infty v_2(t) dt = -w_a \), we have

\[
\int_a^\infty v_2(t) dt < 0.
\]
If \( v_2(t) \) is negative for all \( t \in (a, \infty) \), we are done. In the opposite case, there exists an \( s \in (a, \infty) \) such that \( v_2(s) = 0 \), as well as \( v_2(t) \geq 0 \) for all \( t \in (a, s) \) and \( v_2(t) \leq 0 \) for all \( t \in (s, \infty) \). Hence,

\[
\int_a^\infty \frac{\delta p_a(t)}{\delta q(t)} (l(t) - c) e^{-r(t-a)} dt = \int_a^s v_1(t) v_2(t) dt \\
= \int_a^s v_1(t) v_2(t) dt + \int_s^\infty v_1(t) v_2(t) dt \\
\geq \int_a^s v_1(t) v_2(t) dt + \int_s^\infty v_1(s) v_2(t) dt \\
= v_1(s) \int_s^\infty v_2(t) dt > 0.
\]

Finally, using Eq. (14), it follows

\[
\frac{\delta \text{WE}(a)}{\delta q(t)} < - \frac{\mu \gamma}{u^2} \frac{\delta c}{\delta q(t)} \frac{\delta \mu_a}{\delta q(t)} < 0,
\]

which proves the lemma. \( \square \)

**Lemma 2.** Let \( l'(t) \leq 0 \) for all \( t \geq a \) and \( (\partial / \partial a)(\delta \mu_a / \delta q(t)) > 0 \). Then

\[
\frac{\partial}{\partial a} \frac{\delta \lambda_a}{\delta q(t)} \geq 0.
\]

**Proof.** We calculate

\[
\frac{\partial}{\partial a} \frac{\delta \lambda_a}{\delta q(t)} = \frac{1}{\beta} \left( q(a) + r \right) \left( -l(a) + (q(a) + r) \lambda_a - \lambda'_a \right) + \lambda_a q(a) \\
= (q(a) + r) \frac{\delta \lambda_a}{\delta q(t)} + \lambda_a q(a) \\
= \int_a^\infty \left( (q(a) + r) \frac{q(a) - q(t)}{\beta} \right) + q(a) \right) l(t) p_a(t) e^{-r(t-a)} dt
\]
and
\[
\frac{\partial}{\partial a} \frac{\delta \mu_a}{\delta q(t)} = \frac{1}{\beta} \left( q(a) + r \right) \left( -1 + \left( q(a) + r \right) \mu_a + \mu_a q(a) \right) \\
= \frac{\delta \mu_a}{\delta q(t)} \left( q(a) + r \right) + \mu_a q(a) \\
= \int_a^\infty \left( q(a) + r \right) \left( q(a) - q(t) \right) \frac{q(a)}{\beta} + q(a) \right) p_a(t) e^{-\gamma(t-a)} dt,
\]
which shows that the representations of \((\partial/\partial a)(\delta \mu_a/\delta q(t))\) and \((\partial/\partial a)(\delta \lambda_a/\delta q(t))\) are similar. Following now the proof of Lemma 1 with
\[
v_1(t) := l(t) \quad \text{and} \quad v_2(t) := \left( q(a) + r \right) \left( q(a) - q(t) \right) \frac{q(a)}{\beta} + q(a) \right) p_a(t) e^{-\gamma(t-a)}
\]
gives the desired result. □

References


