Forum for Health Economics & Policy

Volume 10, *Issue* 2 2007 *Article* 1

(AGING AND MEDICAL CARE COSTS)

Aging and Future Healthcare Expenditure: A Consistent Approach

Lukas Steinmann* Harry Telser†

Peter S. Zweifel‡

*Avenir Suisse, Switzerland, lukas.steinmann@avenir-suisse.ch

†University of Zurich, Switzerland, telser@soi.unizh.ch

‡University of Zurich, Switzerland, pzweifel@soi.unizh.ch

Aging and Future Healthcare Expenditure: A Consistent Approach*

Lukas Steinmann, Harry Telser, and Peter S. Zweifel

Abstract

The impact of aging on healthcare expenditure (HCE) has been at the center of a prolonged debate. This paper purports to shed light on several issues of this debate by presenting new evidence on the "red herring" hypothesis advanced by Zweifel, Felder and Meier (1999). This hypothesis amounts to distinguishing a mortality from a morbidity component in healthcare expenditure (HCE) and claiming that failure to make this distinction results in excessive estimates of future growth of HCE. A re-estimation based on a much larger data set is performed, using the refined econometric methodology. The main contribution is consistency, however. Rather than treating the mortality component as a residual in forecasting, its dynamics are analyzed in the same detail as that of the morbidity component when predicting the impact of population aging on the future growth of HCE. For the case of Switzerland, it finds this impact to be relatively small regardless of whether or not the mortality component is accounted for, thus qualifying the "red herring" hypothesis.

KEYWORDS: health econometrics, aging, cost of dying, healthcare expenditure

---

*This paper is part of the Avenir Suisse project "Aging and Healthcare Costs". The authors gratefully acknowledge perceptive criticism by participants of the 2005 meeting of health economists in D-Mainz, in particular Walter Kraemer (D-Dortmund). We thank Rainer Winkelmann and Lorenz Goette (University of Zurich) and three anonymous referees for helpful comments and SWICA Gesundheitsorganisation for providing the data.

1 Introduction

Economic progress and advances of medical technology have enhanced both standards of living and effectiveness of health care in the industrialized world, resulting in an enormous increase of life expectancy. In addition, fertility rates have been dropping since the end of the so-called baby boom in the late 1960s. These two trends in combination cause a marked aging of population, i.e. the share of the elderly will increase dramatically in the two decades to come.

Since there is a strong cross-section relationship between age and per capita healthcare expenditure (henceforth HCE) at the individual level, aging populations are often predicted to have increasing costs of health at the population level. However, the econometric evidence is ambiguous. Only two cross-national studies find the age structure of developed countries to be a consistently significant predictor of HCE [cf. Hitiris and Posnett (1992) and Gerdtham, Søgaard, Jönsson and Andersson (1992b)]. Using a variety of econometric specifications, many other studies conclude that age is not significantly related to per capita HCE [cf. Gerdtham, Søgaard, Andersson and Jönsson (1992a); Gerdtham, Jönsson, MacFarlan and Oxley (1998); Getzen (1992); Leu (1986); O'Connell (1996); OECD (1987); Zweifel, Steinmann and Eugster (2005)].

One explanation that has been advanced is that the cross-section relationship between age and HCE fails to control for an important variable that has varied in the past and may continue to vary in the future, viz. the share of the population approaching death. Individuals approaching death usually suffer deterioration of their health status, which is known to trigger HCE [cf. Guralnik, LaCroix, Branch, Kasl and Wallace (1991)]. Unfortunately, health status is hardly ever measured by health insurers. In view of the strong correlation between health and time to death, Lee and Miller (2002) conclude that time to death (TTD) is a reasonable indicator of health status, especially when the objective is to project future HCE.

Much of the increase of average HCE with age evidenced by health insurers may be due to the fact that the share of the deathbound rises substantially at higher ages, resulting in a changing composition of the insured population. According to several studies, TTD goes along with a marked surge in HCE [cf. Lubitz and Prihoda (1984), Lubitz, Beebe and Baker (1995)]. This means that average HCE is increasingly influenced by a group whose HCE is multiples of that of survivors. Lubitz and Prihoda (1984) estimate

that the 5.1 percent of Medicare recipients dying in a given year account for no less than 29 percent of total HCE. This finding seems to be quite stable over time, as evidenced by Lubitz and Riley (1993) and Barnato, Garber, Kagay and McClellan (2006).

In the course of aging over time, two effects need to be distinguished. First, individuals will simply enter the costly final year of their lives at a later age. Aging thus amounts to a 'red herring' when it comes to predicting future HCE (Zweifel et al. (1999), ZFM henceforth). Second, however, the composition of the population will change in favor of a larger share of individuals close to death due to the baby-boomers. As a result, average HCE may increase in spite of the 'red herring' argument.

More generally, HCE needs to be split into what may be called a morbidity component and a mortality component. While the morbidity component is spent on the survivors, the mortality component is spent on the death-bound only. The role of age in this distinction cannot be determined at a given point in time, since higher age and TTD are the same. In the course of time, however, this needs not to be true. In order to test for the impact of aging on these two components of HCE, three concepts of time have to be distinguished. Historical time (the year of observation) reflects the state of medical technology; time from birth (the age of the individual) stands for the effect of age on the morbidity component (uncorrelated with mortality); and TTD determines the mortality component of HCE, which can only be identified with independent information about remaining life expectancy.

Using panel data on deceased members of two Swiss social health insurers covering the years 1983 to 1994, ZFM were the first to be able to separate these three concepts of time. Relating quarterly HCE to sex, type of insurance, age, proximity to death, and year of observation, they found age not to be a significant determinant of HCE during the last two years of life, at least for the population aged 65 and above. On the other hand, TTD was highly significant, with its estimated effect on HCE consistently increasing with TTD. Extending TTD to cover as much as the last five years of life, ZFM found these findings to be confirmed, leading them to argue that aging will not affect HCE growth when controlling for proximity to death (and hence the mortality component of HCE). For them, aging is a 'red herring' in the debate about the future development of HCE, which crucially depends on changes in medical technology over time. Their conclusion stands in stark contrast to naïve predictions using raw HCE profiles from descriptive statistics.

DOI: 10.2202/1558-9544.1041

The 'red herring' hypothesis has given rise to several conceptual and methodological criticisms [cf. Salas and Raftery (2001), Dow and Norton (2002), Seshamani and Gray (2004a), Seshamani and Gray (2004b), and Stearns and Norton (2004)]. Zweifel, Felder and Werblow (2004) addressed these issues with a more sophisticated econometric analysis that included not only deceased individuals but also survivors. They conclude that the 'red herring' hypothesis argument still stands.

This paper seeks to contribute to the debate in two ways. First, it comprehensively addresses the estimation issues that have been raised since the publication of the ZFM article. These issues are important because econometric specification determines the split between the morbidity and mortality component of HCE. The second contribution is consistency. In fact, ZFM were able to purge HCE from its mortality component by a associating HCE wit TTD. This amounts to treating the mortality component of HCE as a residual that does not deserve further analysis when it comes to forecasting aggregate HCE. This contribution seeks redress the balance by studying the dynamics governing the development of the aggregate mortality component. The following four points will be addressed.

1. Estimation based on a sample of dying and surviving individuals runs the risk of underestimating the influence of TTD on HCE. This is because information about TTD is only available for those individuals who died within the observation period. For survivors, the variable TTD and hence the dummy variable 'one year to death' ($death_1$, see section 2) is unknown. Therefore, $death_1$ takes on a 'missing value' for survivors, causing them to be excluded from the estimation. Replacing missing values of $death_1$ with zeroes, one implicitly assumes that these individuals will not die one year later. But obviously, some survivors will die in the following year. Thus, replacing missing values by zeroes shifts too much HCE to the morbidity component, resulting in an upward (downward) bias of the $death_1$ coefficient pertaining to the morbidity (mortality) component of HCE.[1]
   In case TTD is not measured using dummy variables but rather a single variable measuring distance to death, one needs to replace undefined values by large values implying that survivors will not die in

[1] An empirical comparison of the specification proposed here with replacing missing values by zeroes shows that the bias is small in the present dataset. However, this need not to be the case in general.

the foreseeable future. Assigning a large arbitrary value to TTD, as in Zweifel et al. (2004) and Breyer and Felder (2004), causes TTD to be measured with considerable error (survivors constituting the great majority of observations), which again imparts a downward bias to the coefficient of TTD in the estimation equation.

There are at least three ways to deal with this problem. The first, espoused by ZFM as well as by Seshamani and Gray (2004b), is to limit the sample to deceased individuals only. While this approach may be appropriate when the focus of research is on HCE of the elderly, it runs into problems when the objective is to forecast HCE of an entire population, which requires full age-expenditure profiles. Since the present paper revolves around forecasting, its data base must contain both survivors and deceased individuals.

The second solution, proposed by Stearns and Norton (2004), is to replace all undefined values by zeroes and to introduce an additional variable, 'time to censoring'. The considerable merit of this approach lies in the fact that it uses an indicator of the measurement error contained in TTD for inclusion in the regression, thus purging the error term in the equation of that measurement error. However, it suffers from two weaknesses. First, 'time to censoring', being just an indicator of the measurement error contained in TTD, may fail to effectively perform its purging function. Second, it is highly collinear with measured TTD, causing a severe loss of precision in the estimation of all regression coefficients.

A third alternative, advocated here, is to delete all observations with an undefined value for $death_t$ and hence TTD, where t defines the number of years for which survivors and deceased are distinguished in the estimation. This amounts to excluding the last t annual observations of all survivors in the sample. This is the only way to deal with undefined values that avoids the dubious construction of a proxy variable. The concomitant loss of sample size does not have much relevance in view of the huge size of the data base used here (cf. section 2). However, since TTD may well depend e.g. on technological change in medicine during the years deleted, there is some risk of bias, which would result in an underestimation of the influence of TTD.

2. An important econometric issue is the fact that a substantial number of individuals have zero HCE. This calls either for sample selection

DOI: 10.2202/1558-9544.1041

modeling [typically of the Heckit variety as in ZFM, with the risk of incurring identification problems, as argued by Salas and Raftery (2001)]. The alternative, pursued here, is to specify a two-part model. In the first part, the probability of nonzero HCE is estimated. In the second part, the amount of HCE is estimated for all individuals with positive HCE.

3. HCE values usually are heavily skewed. Transforming the data is a common procedure for handling this problem. However, this entails the difficulty of retransformation after estimation [cf. Manning and Mullahy (2001)]. The solution proposed in this paper is to estimate a generalized linear model.

4. Turning to the forecasting of HCE, consistency is introduced in the following way. With increasing life expectancy, the mortality component of HCE (also called 'cost of dying') will accrue at a higher age. Thus, naïve forecasts of HCE that fail to account for this shift of the mortality component have an upward bias, simply because the share of deathbound at a given age will be smaller than today. This is the ZFM argument. However, ZFM neglected the future dynamics of the aggregate mortality component. Indeed, the cost of dying will increase once the baby boomers reach dying ages. Thus, while ZFM are correct in their distinction between morbidity and mortality components of HCE, they fail to carry it over to the forecasting of future aggregate HCE, committing an inconsistency that is redressed below.

The remainder of this paper is organized as follows. In section 2, the data and econometric methods applied are presented. Section 3 contains the estimation results and presentations of corresponding age profiles of HCE. In section 4, both the morbidity and mortality components of HCE are combined to derive forecasts of future HCE that are fully consistent with estimation results. Section 5 shifts the focus to aggregate HCE and some dynamic issues and section 6 contains concluding remarks.

2 Data and econometric model

The present study is based on a data set comprising about 450,000 members of a major Swiss sickness fund. Spanning the years 1997 through 2004, it

results in a sample of more than 3.7 million observations. Throughout, HCE is defined as per capita healthcare expenditure that is reimbursed by social insurance.[2]

In Switzerland health insurance is mandatory but individually contracted (contrary to most other countries including the United States, there is no employer involvement). Premiums must not be risk-based but uniform for all adults of a given insurer with residence in a given region. There is a risk adjustment scheme designed to prevent risk selection by competing health insurers. The minimum annual deductible is CHF 230 (starting 2004: CHF 300; 1 Swiss franc CHF=0.8 US$ at 2004 exchange rates), combined with a copayment of 10 percent of HCE (capped at CHF 700 annually). Thus, total out-of-pocket payments by an individual with a standard contract is limited to a maximum of CHF 930 per year (effective 2004: CHF 1,000), or some 1.5 percent of average income. Additionally, the insured may choose deductibles of up to CHF 1,500 (starting 2005: CHF 2,500). Managed-care (MC) contracts, viz. health maintenance organizations (HMO), preferred provider organizations (PPO), and physician networks are also available.

There is a vivid debate about the effect of such non-conventional contracts. It revolves around the issue of whether the substantial cost savings achieved by MC are due to risk selection or due to innovations in the guise of changed incentives. A recent study concludes that between one-third (physician networks) and two-thirds (HMO) of cost savings cannot be traced to risk selection effects and therefore may reflect true innovation [Lehmann and Zweifel (2004)]. Still, the choice of MC contracts or policies with higher deductibles needs to be modeled in principle. In order to avoid the (difficult) estimation of the pertinent selection mechanisms, all observations having deductibles in excess of CHF 230 or a MC option were deleted from the sample.

As argued in the preceding section (item 1), since the indicators 'year of death' ($death_0$), 'one year to death' ($death_1$) and 'two years to death' ($death_2$) of 'time to death' (TTD) are to be used, the last three observations of survivors are dropped in order to limit the amount of bias that can be imparted by the measurement error contained in these indicators of TTD.

All these deletions leave 1,273,908 observations for estimation. On average, an individual has 4.5 observations, causing the data set to be of the panel

[2] HCE thus includes all ambulatory care outlays, about 20 percent of long-term care expenditure (the rest being paid out of pocket and by public assistance), and no more than 50 percent of hospital expenditure (since the cantons finance at least one-half of these costs).

DOI: 10.2202/1558-9544.1041

type. This calls for a random-effects specification for two reasons. First, it is important to control for unobserved heterogeneity caused by unobserved health, which may be poorly approximated by the socioeconomic variables available. Second, cohort and age effects can be disentangled in this way. The fixed-effects alternative does not seem to be appropriate because explanatory variables such as sex are time-invariant.

There is a significant number of observations where HCE does not exceed the deductible.[3] These individuals paid out of pocket or were not sick at all. Accordingly, the first step of the two-part model is estimated using the following random-effects probit model[4]

$$Pr(HCE_{i,t} > Deductible) = \Phi\{\alpha + \beta X_{i,t} + \upsilon_i + \epsilon_{i,t}\}, \tag{1}$$

where $\Phi\{\cdot\}$ denotes the standard normal distribution, X_i, t contains TTD (represented by three variables *death*, $death_1$, $death_2$ indicating whether individual is 0, 1, 2 years away from death), *age* (up to cubic), a dummy variable *sexf* for women, interaction terms involving *age* and *sexf*, plus a set of dummy variables controlling for regional effects and calendar years. This was necessary because morbidity and the level of HCE differ considerably between regions, while yearly dummies reflect the state of the medical technology. Moreover, all TTD variables are interacted with *age*, age^2, *sexf*, *sexfage*, and $sexfage^2$ to allow for more flexibility in modeling the mortality component of HCE. Finally, υ_i is the random effect specific to insured individual i, while $\epsilon_{i,t}$ denotes the *iid* error term pertaining individual i in year t.

In the second step of the two-part model, the amount of $HCE|HCE > Deductible$ is estimated using the same set of regressors as in the first part. Since HCE is heavily skewed, estimated parameters may not be robust; moreover, significance tests based on the normality assumption are inadequate. A Box-Cox test indicated that a log transformation results in the best approximation to normality. However, standard estimation procedures predict $E(log(y)|x)$ rather than $log(E(y)|x)$, calling for a difficult retransformation after estimation.[5] Alternatively, a generalized linear model with a

[3] HCE below the deductible is not consistently observed since consumers have no incentive to submit the bills.

[4] Choice of the appropriate model in this context is discussed in Seshamani and Gray (2004a); Dow and Norton (2002); or Jones (2000).

[5] Duan's smearing procedure to retransform the predicted values may be a solution, but has desirable statistical properties only in absence of heteroscedasticity [cf. Duan (1983)].

log-link may be estimated. The relationship between the conditional means $E(y|x)$ and variances $V(y|x)$ of HCE and the regressors suggests using the gamma distribution for the error term $\varepsilon_{i,t}$ [making $V(y|x)$ proportional to $[E(y|x)]^2$, cf. Manning and Mullahy (2001)],

$$HCE_{i,t}|HCE_{i,t} > Deductible = \kappa + \lambda X_{i,t} + \varepsilon_{i,t}. \tag{2}$$

Note that equation 2 does not include an individual random effect; however, the correlation of an individual's observations over time is taken into account by computing robust standard errors. As stated in Hardin and Hilbe (2001), this leads to unbiased and consistent estimates.

Recent research into the so-called Sisyphus syndrome [cf. Zweifel and Ferrari (1992); Zweifel et al. (2005); see also Frech and Miller (1999)] has found that HCE does positively affect remaining life expectancy at the population level, implying that TTD is endogenous to HCE. To avoid possible simultaneity bias, an instrumental variables regression was envisaged. The instruments should ideally be correlated with death, but not with HCE. However, there are no instruments at hand that could satisfy this requirements. Hospital days (contemporary or lagged), diagnostic information, and state of health [derived from previous observations of HCE as in Lehmann and Zweifel (2004)] do not qualify since they are heavily correlated with HCE. Education is known to affect remaining life expectancy at a given age, but unfortunately is not included in the data.

Endogeneity of TTD certainly is a problem in a time series context since it likely reflects advances in medical technology (which is captured by a series of dummy variables, not shown in table 1 below). Due to a comprehensive uniform list of benefits in mandatory health insurance, all Swiss citizens have access to the same medical technology both at a given point of time and over time. This means that differences in individual HCE are not related to differences in technology applied (and hence TTD). Finally, even if HCE and medical technology do prolong life, in many cases the effect is months rather than years, causing the increase in life expectancy not to be reflected in the yearly observations analyzed here. In sum, one has reason to conclude that the endogeneity of TTD is greatly mitigated in the present context and that failure to perform an instrumental variable estimation should not cause much bias.

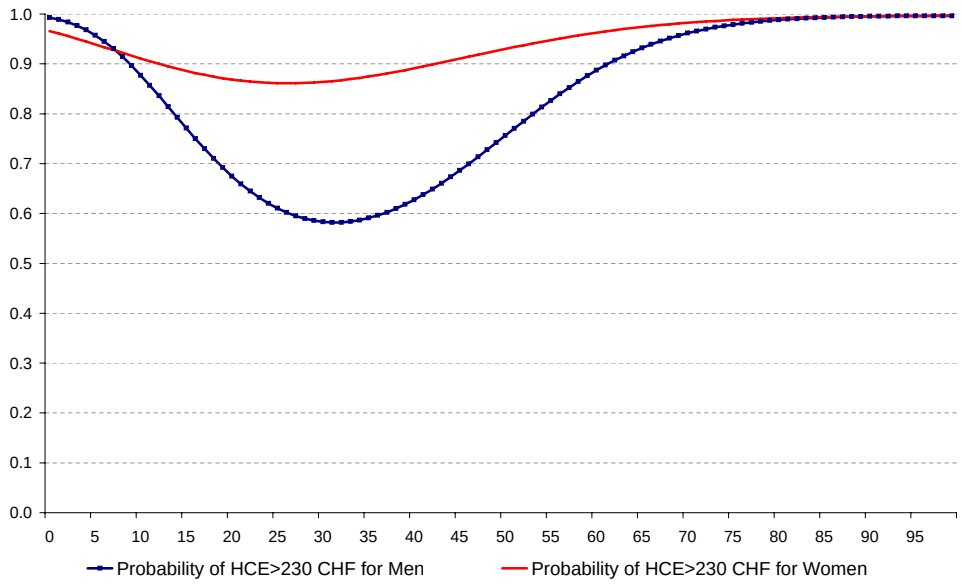
DOI: 10.2202/1558-9544.1041

3 Estimation results

The probit estimates of equation 1 (shown in the first three columns of table 1) exhibit extremely large z values of age, age^2, age^3, and their interaction terms with *sexf*. On the other hand, $death_0$, $death_1$, $death_2$ and their interaction terms often fall short of standard significance levels, likely because of multicollinearity with age-related regressors. However, they are jointly significant, as indicated by a series of Wald tests. In figure 1, the estimated relationship between age and the probability of having $HCE > Deductible$ is plotted for both genders.

As expected, this probability declines with age, reaches a minimum (at age 25 for women and 30 for men), and then increases again. At ages beyond 75, both women and men are very likely to have $HCE > Deductible$.

Figure 1: Probability of having $HCE > Deductible$



The estimation results for the second part of the model (equation 2, cf. the last three columns of table 1), again show most regressors involving age to have high z values. However, several regressors fall short of conventional significance levels, likely due to two different reasons. First, there indeed may not be a relationship between the regressor and HCE. For example, this seems to be the case for age^3 (but not for $sexfage^3$). A regression omitting this particular variable indicated that the parameters of age and age^2 remained stable, suggesting age^3 does not influence HCE of men. A second reason for lack of significance might be multicollinearity. While multicollinearity does not result in biased parameters, it causes estimated standard errors to strongly increase. Thus, in the presence of multicollinearity, the probability of type II errors looms large. For example, omitting age^3 results in the parameter of age^2 having a much higher z value than shown in table 1.

This finding is confirmed by a Wald test, which indicates that the coefficients pertaining to three of the age-related regressors jointly are not significantly different from the first two [$H_0 : \beta_{age} + \beta_{age^2} + \beta_{age^3} = \beta_{age} + \beta_{age^2}$ need not be rejected, $\chi^2(1) = 0.09$, $Prob > \chi^2 = 0.7634$]. Since the coefficients of both the first two and also of all three age-related terms are jointly different from zero, a quadratic specification for estimating the relationship between age and HCE for men is sufficient.

For women, however, the cubic specification seems to be more adequate since $H_0 : \beta_{sexfage} + \beta_{sexfage^2} + \beta_{sexfage^3} = \beta_{sexfage} + \beta_{sexfage^2}$ is rejected with very low error probability [$\chi^2(1) = 248.89$, $Prob > \chi^2 = 0.0000$].

With respect to the occurrence of death among men, most regressors, including interactions with age-related variables, are significantly different from zero. Moreover, the significance of these regressors differs between men and women [i.e. $H_0 : \beta_{death_t sexf} + \beta_{death_t sexfage} + \beta_{death_t sexfage^2} = 0$ may be rejected with error probability $Pr < 0.008$], justifying the specification presented in table 1 which allows for gender-specific HCE patterns.

The estimation results of table 1 may now be used to compute the expected age-expenditure profiles plotted in figure 2. The expected value of the morbidity component (C_{MORB}) of HCE (as a function of age) is obtained by weighting estimated HCE [$E(HCE|HCE > 230)$] from equation 2 with the estimated age-specific probability of having HCE exceeding CHF 230. Dropping i for simplicity [cf. equation 1], one has

$$C_{MORB} = E(HCE|HCE > 230|death_t = 0) * Pr(HCE > 230). \quad (3)$$

DOI: 10.2202/1558-9544.1041

Table 1: Estimation results

	First part, equation 1			Second part[a], equation 2		
Variable	Coeff	SE	z	Coeff	SE[b]	z
age	-0.1598	0.0017	-92.20	0.0416	0.0020	20.44
age^2	0.0034	0.0000	71.02	-0.0002	0.0001	-4.10
age^3	0.0000	0.0000	-47.19	0.0000	0.0000	0.63
$sexf$	-0.6183	0.0233	-26.54	-0.2597	0.0210	-12.36
$sexfage$	0.0982	0.0024	41.69	0.0311	0.0025	12.53
$sexfage^2$	-0.0019	0.0001	-29.56	-0.0009	0.0001	-12.74
$sexfage^3$	0.0000	0.0000	19.75	0.0000	0.0000	13.28
$death_0$	0.3893	0.5393	0.72	2.3152	0.3853	6.01
$death_0age$	0.0179	0.0181	0.99	-0.0044	0.0109	-0.40
$death_0age^2$	-0.0001	0.0001	-0.73	-0.0001	0.0001	-1.44
$death_0sexf$	1.6638	1.1414	1.46	0.0847	0.5610	0.15
$death_0sexfage$	-0.0553	0.0352	-1.57	0.0147	0.0155	0.95
$death_0sexfage^2$	0.0004	0.0003	1.60	-0.0002	0.0001	-1.89
$death_1$	-0.8932	0.4396	-2.03	3.2513	0.5056	6.43
$death_1age$	0.0501	0.0145	3.46	-0.0484	0.0145	-3.33
$death_1age^2$	-0.0004	0.0001	-3.57	0.0002	0.0001	2.20
$death_1sexf$	1.6041	0.8365	1.92	-1.3660	0.7519	-1.82
$death_1sexfage$	-0.0597	0.0259	-2.30	0.0556	0.0209	2.66
$death_1sexfage^2$	0.0005	0.0002	2.51	-0.0005	0.0001	-3.22
$death_2$	-1.0802	0.4973	-2.17	3.5398	0.8664	4.09
$death_2age$	0.0506	0.0163	3.10	-0.0680	0.0244	-2.79
$death_2age^2$	-0.0004	0.0001	-3.19	0.0004	0.0002	2.29
$death_2sexf$	1.0765	0.8432	1.28	-1.0765	1.0186	-1.06
$death_2sexfage$	-0.0496	0.0265	-1.87	0.0490	0.0284	1.72
$death_2sexfage^2$	0.0004	0.0002	2.19	-0.0004	0.0002	-2.17
$constant$	2.3208	0.0229	101.41	6.1659	0.0252	244.96
Observations	1,273,908			1,009,286		
Groups	285,168			225,952		
	Log-Likelihood: -508,993.24			AIC: 17.34		
Wald $Pr(\chi^2 > 0)$	0.0000					

[a] Estimation using the Newton-Raphson algorithm.

[b] Estimation using the Huber-White sandwich estimator.

All dummy variables reflecting death and TTD are set equal to zero when computing $E(HCE|HCE > 230|death_t = 0)$ in equation 3, ensuring that C_{MORB} is uncorrelated with the mortality component of HCE. On the whole, C_{MORB} (solid lines in figure 2, starting at values close to zero) increases with age for both sexes. For women, C_{MORB} generally runs higher and increases more steadily, while for men up to age 30, there seems to be no relationship between age and C_{MORB}. At ages beyond 80, women (but not men) exhibit exponentially increasing C_{MORB}. This important difference can be detected thanks to the cubic specification, which allows for additional flexibility.

The mortality component C_{MORT} of expected HCE is calculated as a function of age as follows. First, the cost of dying COD is defined as the excess of predicted HCE given that the individual dies, summed over the last three years of life. Thus, COD is given by

$$COD = \sum_{t=0}^{2} [E(HCE|HCE > 230|death_t = 1) - E(HCE|HCE > 230|death_t = 0)]. \quad (4)$$

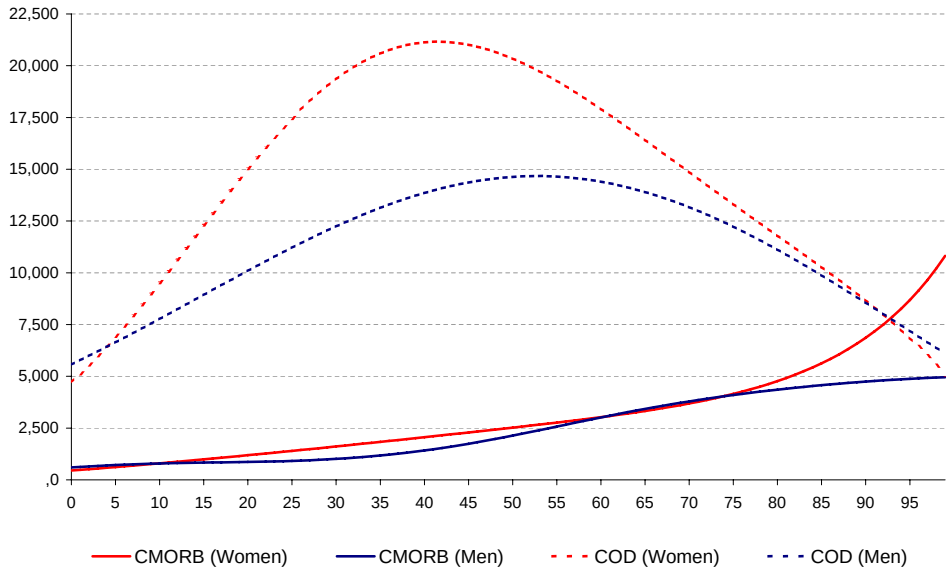
Next, COD is weighted by age-specific mortality rates $Pr(MORT)$ whose values are observed for the year 2000 and forecast for 2001 to 2030. The age-dependent mortality component of expected HCE is thus given by

$$C_{MORT} = COD * Pr(MORT) \quad (5)$$

In figure 2 only COD as defined in equation 4 is shown for the last year of life. Interestingly, the graphs tend to be inversely U-shaped, with a peak at age 54 for men and at age 41 for women, HCE amounting to approximately CHF 21,000 for women and CHF 14,500 for men, respectively. The conditional mortality component of HCE of the deceased (COD) can now be compared to the morbidity component pertaining to the survivors. In keeping with previous findings [Lubitz and Riley (1993); Zweifel et al. (1999)], COD is several multiples of C_{MORB} at most ages. However, since COD declines past ages 54 (men) and 41 (women), the factor of proportionality is not a constant but varies with age, an observation also made by Zweifel et al. (2004).

DOI: 10.2202/1558-9544.1041

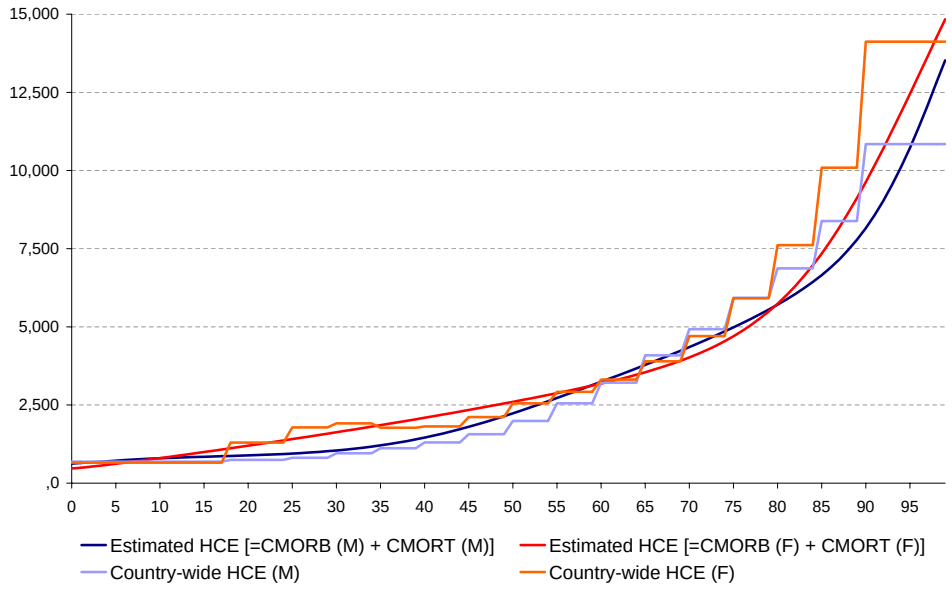
Figure 2: The morbidity (CMORB) component of HCE and the cost of dying (COD) as a function of age, in CHF (2000)



These findings are consistent with economic theory, which predicts that the relationship between the value of a statistical life and age is inversely U-shaped, with a maximum value around the age of forty [Schelling (1968), Shepard and Zeckhauser (1982)]. Therefore, willingness to pay for survival (which is reflected in COD if physicians act as reasonably good agents of their patients) should exhibit the same pattern. This prediction has been empirically confirmed by several authors [cf. Carthy, Chilton and Cookson (1999), Jones-Lee, Hammerton and Philips (1985), Mount, Weng and Schulze (2000)].

Estimated C_{MORB} and C_{MORT} sum up to predicted healthcare expenditure $\widehat{HCE}$. Comparison of $\widehat{HCE}$ with country-wide age-specific HCE published by the Swiss risk adjustment fund may serve as a check on the ex-

Figure 3: Country-wide and estimated age profiles of HCE, using observed deaths (in CHF, 2000)



ternal validity of the two-part-model estimated. As can be seen from figure 3, the profiles match quite well.[6]

[6]Using the quadratic rather than the cubic specification of equation 2, one obtains $\widehat{HCE}$ values that do not fit as well as in figure 3, especially for women at ages around 30 and at high ages, thus confirming the need for a cubic specification.

DOI: 10.2202/1558-9544.1041

4 What impact of aging on future health care expenditure?

The claim of ZFM is that neglecting the distinction between the morbidity and mortality component of HCE results in an overestimation of the impact of aging of the population on future HCE. Since then, a consensus has evolved that the mortality component must be controlled for when estimating HCE profiles designed to serve as a basis for prediction. Thus, forecasting future HCE according to ZFM accounts to simply extrapolating equations 3 and 5, with $Pr(MORT)$ adjusting for the expected increase in longevity. However, this procedure ignores that consistency requires the not only the morbidity but also mortality component to be controlled for changes in the composition of population. Failure to do so may be consequential when the number of deaths will increase in the future (at present, because the baby boomers are coming of age). This phenomenon is of relevance not only for the United States but for other industrial countries as well, among them Switzerland. While crude Swiss mortality rates decreased (or at least were stable) until the year 2000, they are predicted to increase from a current value of below 900 per 100,000 to more than 1,200 by 2050.

Incorporating the mortality component into forecasts of HCE therefore must cause a surge in future HCE that was not accounted for by ZFM, who implicitly assumed the mortality rate to remain constant. In the consistent approach proposed here, this failure is corrected as follows. Predicted $\widehat{HCE_t}$ $(t = 2000, ..., 2030)$ is adjusted for changes in the future cohort structure of the population by calculating

$$\widehat{HCE_t} = \frac{1}{P_t}\sum_{c=1}^{n}[p_{c,t} * C_{MORBc,t}] + \frac{1}{P_t}\sum_{c=1}^{n}[d_{c,t} * COD_{c,t}] \qquad (6)$$

$$P_t = \sum_{i=1}^{n} p_{c,t}.$$

Here, P_t is the number of individuals alive in future period t, composed of the n cohorts comprising $p_{c,t}$ people, of which $d_{c,t}$ die,[7] C_{MORB} denotes

[7] These numbers are available from the demographic forecasts provided by Münz and Ulrich (2001).

the predicted per capita cost of morbidity, and COD stands for the cost associated with dying as defined in equation 4. By definition of the aging effect, age-specific C_{MORB} and COD are held constant over time ($C_{MORBc,t} = C_{MORBc,2000}$ and $COD_{c,t} = COD_{c,2000}$). Thus, any future change in preferences and medical technology that might influence C_{MORB} and COD are assumed away in order to filter out the pure aging effect. In relative terms, the pure aging effect ($\mathcal{A}$) can now be defined as

$$\mathcal{A} = \frac{\widehat{HCE_t}}{HCE_{2000}}, t = 2000, ..., 2030. \tag{7}$$

The extent to which $\widehat{HCE}$ differs from a naïve forecast that does not distinguish between the morbidity and mortality component of HCE crucially depends on how the cost of dying varies with age. If COD had a flat profile, the naïve forecast, the ZFM approach, and the consistent approach advocated here would converge since the mortality component would be largely irrelevant. If however COD were to increase with age (which it does to some extent, cf. the age profile of COD in figure 2), then the mortality component of aggregate HCE will increase temporarily when the baby boomers cause a rise in the number of deaths. In this case, the consistent forecasting approach defined by equation 6 predicts an even more marked aging effect than either the naïve forecast or the ZFM approach.

The three forecasts are juxtaposed in figure 4. The naïve forecast (which fails to distinguish the morbidity and mortality components of HCE) results in a pure aging effect that increases from 0.5 percentage points (pp) in 2000 to 0.8 pp by 2010 and approaches 0.7 pp by 2030 (0.72 pp on average p.a.). The ZFM alternative (which distinguishes the two components but neglects the impact of baby boomers on the mortality component of future HCE) would predict no impact of aging on HCE, at least when using the flat, age-invariant profiles estimated in ZFM. However, the consistent approach adopted here results in a pure aging effect of 0.5 pp annually over the period 2000 to 2030.

The alternative proposed in this paper (which consistently distinguishes between the morbidity and the mortality components not only in estimation but also in forecasting) results in a pure aging effect of 0.43 pp in 2000 that reaches a maximum of 0.62 pp around 2010 and converges to 0.48 pp by 2030 (0.54 pp on average p.a.). This difference needs to be put in proper perspective. Compared to the 5 percent growth rate observed in Swiss HCE

DOI: 10.2202/1558-9544.1041

between 1996 and 2003, the question whether the aging of the population accounts for 0.5 or 0.7 pp is not very relevant. The absolute differences are not too impressive either. According to the naïve forecast, aging would cause HCE to increase from about CHF 2,200 in 2000 to CHF 2,700 by 2030. The ZFM alternative predicts CHF 2,600. The consistent approach of this paper leads to predicted HCE reaching CHF 2,500 by 2030. Therefore, the effect of aging on HCE is relatively small in all scenarios, especially if compared with an extrapolation based on a 5 percent growth rate which would have HCE increase to CHF 9,500.

On the whole, these findings serve to qualify the conclusion reached by ZFM and reiterated by Zweifel et al. (2004), who in their re-examination of ZFM state that the "naïve estimation that does not control for proximity to death will grossly overestimate the effect of population ageing on aggregate health care expenditure". This statement must be seen in relation to the relatively small impact of aging on future HCE. Only if future HCE growth were to fall markedly below the high values experienced in the recent past would the distinction between the morbidity and mortality components of HCE make a significant difference.

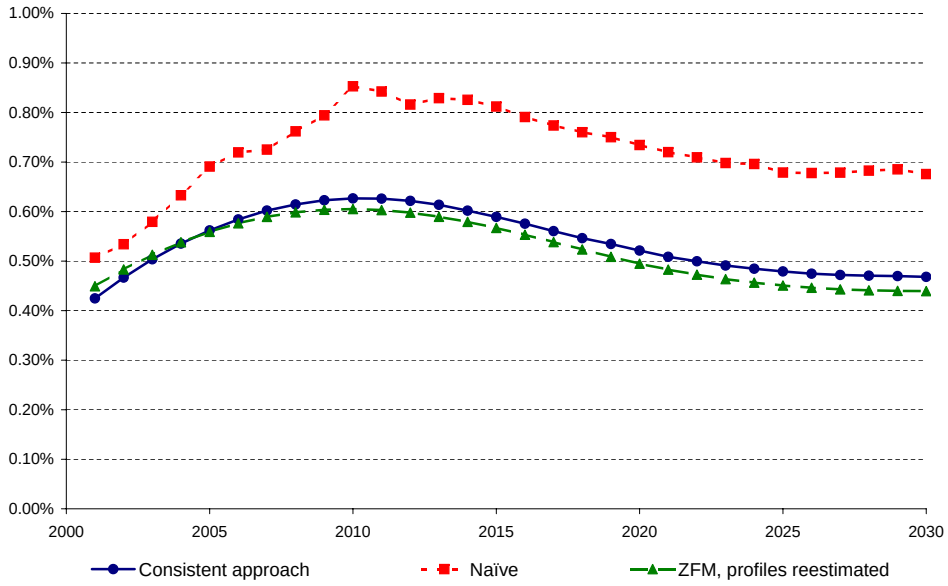
On the aggregate level, the picture looks quite different. According to the consistent approach, aggregate HCE in mandatory health insurance will increase by a third, from CHF 15.5 in 2000 to 20 billion in 2030 solely due to demographic changes. Since in Switzerland HCE is financed by flat rate premiums, the younger will have to bear a substantial part of the costs induced by aging. While the under 60 year old were supporting the elderly with CHF 4 billion in 2000, in 2030 it will be CHF 6 billion, an increase by 50 percent (in prices of 2000). Thus, even if the aging effect is quite small in terms of per capita HCE, the distributional effect is significant.

5 Discussion

One has to keep in mind that the preceding analysis is based on a static concept. Age profiles are held constant when computing the aging effect. This is not a simplification but rather—by definition—a necessity in order to isolate the consequences of aging. Only the demographic structure of the population is allowed to change.

However, age profiles of HCE have been changing in the past and will most certainly also change in the future. There are two competing hypothe-

Figure 4: Contribution of aging to future growth of HCE (2000-2030)

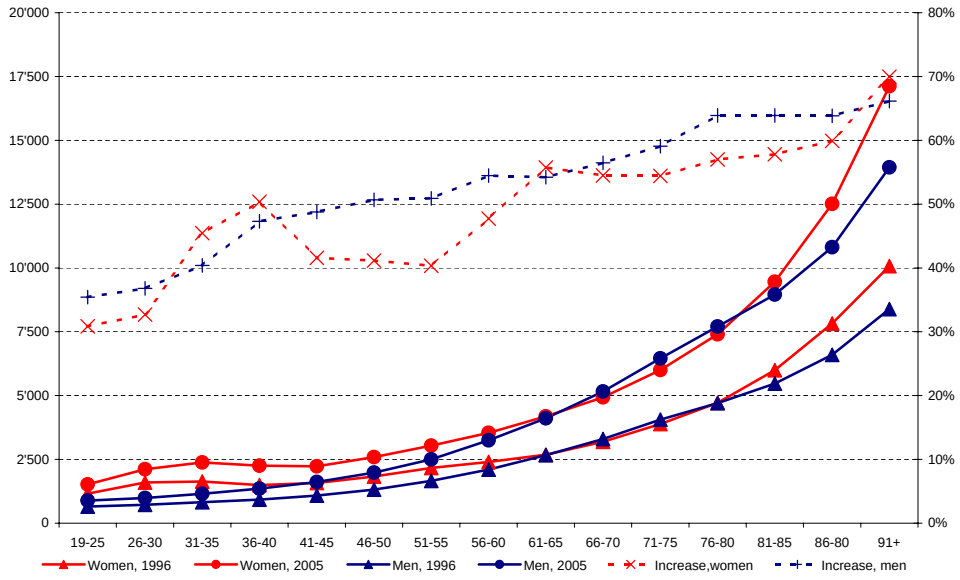


ses. The first—*compression of morbidity*—goes back to Fries (1980) and postulates a compression of individual HCE towards the end of life. People will become healthier over time, thus reducing their HCE until they enter the terminal phase of life. Olshansky, Rudberg, Carnes, Cassel and Brady (1991) on the other hand claim that people at all ages will get more medical treatment, resulting in an upward shift of the age profiles, possibly accompanied by an increase of the age gradient. This second hypothesis is known as the *expansion of morbidity*. If true, a steepening of age profiles will indirectly increase the aging effect and HCE growth in an aging society.

In Switzerland, a significant steepening of age profiles has been observed since the new health insurance law was put in force in 1996. While growth of HCE to 2005 was below 50 percent for those under 50 years, it was up to 70 percent for the oldest age categories (cf. figure 5). This early age where steepening begins contributes evidence in favor of the expansion of

DOI: 10.2202/1558-9544.1041

Figure 5: Steepening of age profiles of HCE, Switzerland (1996-2005)



morbidity hypothesis advanced by Olshansky et al. (1991). There is also a theoretical argument in favor of the expansion hypothesis. HCE certainly are inputs in the production of health and health has to be viewed an output. Fries' hypothesis is based on an improvement of the output which in turn also requires more inputs and therefore more HCE. For many other countries, a steepening of age profiles has been observed as well.[8] Cutler and Meara (1999), analyzing Medicare data for the period 1985 to 1995, find that this steepening is due to an increase of post-acute services such as home healthcare and skilled nursing care. Based again on Medicare data,

[8]For Belgium, cf. Van Tielen, Peys and Genaert (1998); for France, Australia and other OECD-countries, cf. Jacobzone (2002); for Germany, cf. Buchner (2001); for the United Staates of America, cf. Fuchs (1998), Cutler and Meara (1999); for Switzerland, cf. Steinmann and Telser (2005).

Fuchs (1998) concludes that the increasing age gradient of HCE is mainly due to technological progress. Therefore, it might become even steeper in the wake of technological change in medicine, reinforcing this indirect aging effect. In their international comparison, Seshamani and Gray (2002) report mixed evidence, suggesting that age-specific growth of HCE is linked to the type of healthcare system. Since an increasing age gradient of HCE was not observed in England and Wales and Canada with their National Health Services, this constancy may be interpreted as the reflection of age-based rationing (which is typically not imposed by insurance-based systems). This leads to the conclusion that institutional characteristics cause differences in age-specific growth of HCE.

If the steepening of age profiles were to continue at the pace observed between 1996 an d 2005, forecast per capita HCE in Switzerland would be more than three times as large as the 0.55 pp. estimated based on constant age profiles, resulting in yearly growth rates of 1.7 pp. between 2000 and 2030. Aggregate HCE will almost double, reaching CHF 28 billion in 2030 (compared to 20 billion with constant age profiles). This figure still excludes HCE growth that is independent of age. While both per capita and aggregate HCE would be higher still, the distribution of costs and benefits of HCE between young and old would not be affected.

6 Conclusions

Aging will not contribute much to future growth of per capita healthcare expenditure (HCE). When computing the aging effect in a naïve way, i.e. neglecting the distinction between the morbidity and mortality component of HCE as proposed by Zweifel et al. (1999), HCE growth in Switzerland is predicted to hover around 0.7 percentage points per annum over the next 25 years. If proximity to death is controlled for when estimating the age profile of the morbidity component while neglecting the likely dynamics of the mortality component [as Zweifel et al. (1999) implicitly do], the predicted contribution of aging to HCE growth shrinks to 0.5 points. Finally, the fully consistent approach, proposed in this paper, is to use the same decomposition developed in estimation also for forecasting. This means taking into account the fact that the cost of dying will be weighted by an increasing mortality rate due to the baby boomers' coming of age. This causes the predicted contribution of aging to HCE growth to increase back to 0.55 percentage

DOI: 10.2202/1558-9544.1041

points per annum. Thus, a fully consistent decomposition approach leads to a result that does not substantially modify a naïve extrapolation.

There is, however, concern that these projections underestimate the consequences of of aging. First, even if HCE growth due to aging will be quite small in per capita terms, on the aggregate level HCE will increase from CHF 15 billion to 20 billion in 2030 (1 CHF=0.8$). Second, a steepening of age profiles has been observed in the past. If this process were to continue, an even larger share of HCE will be devoted to the elderly, causing the financial burden on the younger generations to increase. Third, age profiles for long-term care (LTC) are known to be even steeper than for the other components of HCE [Werblow, Felder and Zweifel (2006)]. Therefore, the aging effect is expected to be larger for LTC. Since, in the case of Switzerland, only 20 percent of LTC is financed by mandatory health insurance, the full extent of aging will be underestimated in this study.

There is an additional channel of influence that may reinforce the impact of aging on future HCE. It is the so-called Sisyphus syndrome which claims that aging shifts voting power to the elderly who in turn seek—through the democratic process—to allocate more and more resources to public health care [Zweifel and Ferrari (1992), Zweifel et al. (2005)]. This means that the young members of society will increasingly have to carry the burden of aging. Since they also have to finance old-age pensions for retired beneficiaries whose remaining life expectancy continues to increase, aging may well become a major challenge.

References

Barnato, A.E., A.M. Garber, Ch.R. Kagay and M.C. McClellan (2006). Trends in the Use of Intensive Procedures at the End of Life. *Forum for Health Economics & Policy, Forum: Frontiers in Health Policy Research*, 4(5).

Breyer, F. and S. Felder (2004). Life Expectancy and Health Care Expenditures: A New Calculation for Germany Using the Costs of Dying. Working Paper 452, DIW Berlin.

Buchner, F. (2001). Versteilerung von Ausgabenprofilen in der Krankenversicherung. In G. Gäfgen and P. Oberender, eds., *Gesundheitsökonomische Beiträge*, volume 37. Nomos, Baden-Baden.

Carthy, T., S. Chilton and R. Cookson (1999). *The Valuation of Benefits of Health and Safety Control: Final Report to the Health and Safety Executive*. Health and Safety Executive, London.

Cutler, D. and E. Meara (1999). The Concentration of Medical Spending: An Update. Working Paper 7279, NBER.

Dow, W. H. and E. C. Norton (2002). The Red Herring that Eats Cake: Heckit versus Two-part Model Redux. *Triangle Health Economics Working Paper Series*, 1.

Duan, N. (1983). Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*, 78(383):605–610.

Frech, III., H. E. and Jr. Miller, R. D. (1999). *The Productivity of Health Care and Pharmaceuticals: An International Comparison*. The AEI Press, Washington D.C.

Fries, J. F. (1980). Aging, Natural Death, and the Compression of Morbidity. *New England Journal of Mededicine*, 303(3):130–135.

Fuchs, V. R. (1998). Health Care for the Elderly: How Much, Who will Pay for It? Working Paper 6755, NBER.

Gerdtham, U. G., B. Jönsson, M. MacFarlan and H. Oxley (1998). The Determinants of Health Expenditure in the OECD Countries: A Pooled Data Analysis. In P. Zweifel, ed., *Health, the Medical Profession, and Regulation. Developments in Health Economics and Public Policy, vol. 6*, 113–34. Kluwer Academic, Boston and Dordrecht.

Gerdtham, U. G., J. Søgaard, F. Andersson and B. Jönsson (1992a). An Econometric Analysis of Health Care Expenditure: A Cross-Section Study of the OECD Countries. *Journal of Health Economics*, 11(1):63–84.

Gerdtham, U. G., J. Søgaard, B. Jönsson and F. Andersson (1992b). A Pooled Cross-Section Analysis of the Health Care Expenditures of the OECD Countries. In P. Zweifel and III. Frech, H. E., eds., *Health Economics Worldwide. Developments in Health Economics and Public Policy Series, vol. 1, Norwell, Mass*, 287–310. Kluwer Academic, Boston and Dordrecht.

DOI: 10.2202/1558-9544.1041

Getzen, T.E. (1992). Population Aging and the Growth of Health Expenditures. *Journal of Gerontology*, 47:98–104.

Guralnik, J.M., A.Z. LaCroix, L.G. Branch, S.V. Kasl and R.B. Wallace (1991). Increasing Longevity and Medicare Expendituresm. *American Journal of Public Health*, 81(4):443–447.

Hardin, J. and J. Hilbe (2001). *Generalized Linear Models and Extensions*. Stata Press, College Station.

Hitiris, T. and J. Posnett (1992). The Determinants and Effects of Health Expenditure in Developed Countries. *Journal of Health Economics*, 11:173–181.

Jacobzone, S. (2002). Healthy Ageing and the Challenges of New Technologies: Can OECD Social and Health-Care Systems Provide for the Future? In *Healthy Ageing and Biotechnology, Policy Implications of new Research*, 37–53. OECD.

Jones, A. (2000). Health Econometrics. In Anthony J. Culyer and Joseph P. Newhouse, eds., *Handbook of Health Economics*, volume 1A, 265–343. Elsevier, Amsterdam, London, New York, Oxford, Paris, Shannon, Tokyo.

Jones-Lee, M. W., M. Hammerton and P. R. Philips (1985). The Value of Safety: Results of a National Survey. *Economic Journal*, 95:49–72.

Lee, R. and T. Miller (2002). An Approach to Forecasting Health Expenditures, with Application to the U.S. Medicare System. *Health Services Research*, 37(5):1365–1386.

Lehmann, H. and P. Zweifel (2004). Innovation and Risk Selection in Deregulated Social Health Insurance. *Journal of Health Economics*, 23(5):997–1012.

Leu, R. E. (1986). The Public-Private Mix and International Health Care Costs. In A. J. Culyer and B. Jönsson, eds., *Public and Private Health Services: Complementarities and Conflicts*, 41–63. Blackwell, Oxford and New York.

Lubitz, J. D. and G. F. Riley (1993). Trends in Medicare Payments in the last Year of Life. *New England Journal of Medicine*, 328:1092–1096.

Lubitz, J.D., J. Beebe and C. Baker (1995). Longevity and Medicare Expenditures. *The New England Journal of Medicine*, 332(15):999–1003.

Lubitz, J.D. and R.T. Prihoda (1984). he Use and Costs of Medicare Services in the Last Two Years of Life. *Health Care Financing Review*, 5(3):117–131.

Manning, W. G. and J. Mullahy (2001). Estimating Log Models: To Transform or not to Transform? *Journal of Health Economics*, 20:461–497.

Mount, T., W. Weng and W. Schulze (2000). Automobile Safety and the Value of Statistical Life in the Family: Valuing Reduced Risk for Children, Adults and the Elderly. Working paper, United States Environmental Protection Agency.

Münz, R. and R. Ulrich (2001). *Alterung und Migration: Alternative demographische Szenarien für die Schweiz*. Avenir Suisse.

O'Connell, J. (1996). The Relationship between Health Expenditures and the Age Structure of the Population in OECD Countries. *Health Economics*, 5:573–578.

OECD (1987). *Financing and Delivering Health Care: A Comparative Analysis of OECD Countries*. Paris.

Olshansky, S. J., M. A. Rudberg, B. A. Carnes, B. A. Cassel and J. A. Brady (1991). Trading off Longer Life for Worsening Health: The Expansion of Morbidity Hypothesis. *Journal of Aging and Health*, 3(2):194–216.

Salas, C. and J. P. Raftery (2001). Econometric Issues in Testing the Age Neutrality of Health Care Expenditure. *Health Economics Letters*, 10:669–671.

Schelling, T. C. (1968). The Life You Save May be Your Own. In S. B. Chase, ed., *Problems in Public Expenditure Analysis*. Brookings Institution, Washington, DC.

Seshamani, M. and A. Gray (2002). The Impact of Ageing on Expenditures in the National Health Service. *Age and Aging*, 31:287–294.

DOI: 10.2202/1558-9544.1041

— (2004a). Ageing and Health-care Expenditure: The Red Herring Argument Revisited. *Health Economics*, 13:303–314.

— (2004b). A Longitudinal Study of the Effects of Age and Time to Death on Hospital Costs. *Journal of Health Economics*, 23:217–235.

Shepard, D. S. and R. J. Zeckhauser (1982). Life-Cycle Consumption and Willingness to Pay for Increased Survival. In M. W. Jones-Lee, ed., *The Value of Life and Safety*, 95–141. North-Holland Publishing Company, Amsterdam, New York, Oxford.

Stearns, S. C. and E. C. Norton (2004). Time to Include Time to Death? The Future of Health Care Expenditure Predictions. *Health Economics*, 13:315–327.

Steinmann, L. and H. Telser (2005). *Gesundheitskosten in der alternden Gesellschaft - Wieso nicht die Zunahme älterer Menschen, sondern falsche Anreize das Gesundheitswesen unter Druck setzen. (Health Costs in the Aging Society)*. Verlag NZZ, Zürich.

Van Tielen, R., F. Peys and J. Genaert (1998). The Demographic Impact on Ambulatory Pharmaceutical Expenditure in Belgium. *Health Policy*, 45:1–14.

Werblow, A., S. Felder and P. Zweifel (2006). Population Ageing and Health Care Expenditure: A School of 'Red Herrings'? *Revision submitted to Health Economics*.

Zweifel, P., S. Felder and M. Meier (1999). Ageing of Population and Health Care Expenditure: A Red Herring? *Health Economics*, 8(6):485–496.

Zweifel, P., S. Felder and A. Werblow (2004). Population Ageing and Health Care Expenditure: New Evidence on the "Red Herring". *Geneva Papers on Risk and Insurance - Issues and Practice*, 29(4):652–667.

Zweifel, P. and M. Ferrari (1992). Is there a Sisyphus Syndrome in Health Care? In P. Zweifel and H. E. Frech III, eds., *Health Economics Worldwide. Developments in Health Economics and Public Policy Series, Vol. 1, Norwell, Mass*, 311–30. Kluwer Academic, Boston and Dordrecht.

Zweifel, P., L. Steinmann and P. Eugster (2005). The Sisyphus Syndrome in Health Revisited. *International Journal of Health Care Finance and Economics*, 5(2):127–145.

DOI: 10.2202/1558-9544.1041