Estimating and forecasting Costs of Illness in the Netherlands

Maria Gheorghe*, Parida Obulqasim* & Pieter van Baal**

*Institute for Medical Technology Assessment

Institute of Health Policy & Management

Erasmus University Rotterdam

** corresponding author: vanbaal@bmg.eur.nl

Summary

In the Netherlands, there has been a tradition of conducting so-called Costs-of-Illness (COI) studies. In the Dutch COI studies total healthcare expenditures (HCE) in different healthcare settings in the Netherlands are uniquely attributed to disease categories, healthcare provider categories, specified by sex and age. COI studies in the Netherlands have been conducted for years 1994, 1999, 2003, 2005, 2007 and 2010. This report described the results of a study with the aim to a) predict HCE for the years 1995-1998, 2000-2002, 2004, 2006, 2008-2009 and b) forecast HCE for the years 2011-2015.

Contents

1. General approach

In this report we will make a distinction between *prediction* and *forecasting.* With prediction we mean the prediction of cost values for the years between different COI studies. These are the years 1995-1998, 2000-2002, 2004, 2006, 2008 and 2009. In fact, we will predict all missing year's costs for the years in the period 1994-2010 using data for years 1994, 1999, 2003, 2005, 2007 and 2010. Predictions are based on a regression model that besides including age, gender and calendar year also includes variables indicating health care utilization and/or health care need.

With forecasting we mean the forecasts of cost values for the years after the most recent COI study. These are the years 2011-2015. Forecasts are purely based on extrapolating trends and do not require any information on health care utilization and/or health care need.

1.1 Prediction model

Given that the age patterns differ strongly per disease and health care provider we have decided to develop regression models for each disease separately. The general regression model we used for predictions health care expenditures for each disease and health care expenditures is:

$$\log\left(c_{x,t,g}\right) = \beta_t X_t + \gamma_{x,g} Y_{x,g} + \alpha_i Z(i)_{x,t,g} + \varepsilon_{x,t,g} \qquad (1)$$

Where c stands for health care expenditures for a particular disease and health care provider available from the COI studies, x for age, t year and g gender. X_t indicates a vector of dummy variables indicating the different years; $Y_{x,g}$ indicates a vector of dummy variables indicating the different age and gender groups; β_t a vector of parameters to be estimated capturing the average health care expenditures over all age and gender categories; $\gamma_{x,g}$ a vector of parameters to be estimated capturing the age and gender effects. Crucial parameters are α_i. These parameters explain the deviations from the year, gender and age specific effects on the level of costs. $Z(i)_{x,t,g}$ indicates a vector of variables indicating health care utilization and/or health care need. These may differ per health care provider depending on data availability. Data on $Z(i)_{x,t,g}$ does not come from the COI studies but from other studies. In case of hospital expenditures these variables consist of hospital admissions and length of stay. $\varepsilon_{x,g,t}$, reflect particular age- and gender specific historical influences not captured by the model, is normally distributed with mean 0.

We choose to use levels instead of first differences as the distance in terms of years between different COI studies differed. We could not model the year specific influences using splines as the number of years in the data was too low (6 years). As a result we could not directly predict values for the year specific influences in the years we needed to predict. Therefore, coefficients for these years were interpolated from the estimated coefficients of the years that we had. For example, the coefficients for year 2009 were derived from the estimated coefficients of the year 2007 and 2010 as follows:

$$\widehat{\beta}_t^{2009} = \frac{1}{3}\widehat{\beta}_t^{2007} + \frac{2}{3}\widehat{\beta}_t^{2010}, \qquad t = 4,5 \tag{2}$$

For the estimated years' costs, the model was calibrated using the total health care costs per health care costs from National health care accounts data. For each year, for all disease and age groups and for both genders, the total for the estimated costs had to match the one published by CBS (This was the initial idea, however, the total costs from cbs does not match the total for the years that we have. Therefore, this step was not performed).

1.2 Forecasting model

To predict HCE we used the so-called Lee-Carter model which has been used often to forecast age specific mortality and life expectancy (Lee and Carter 1992; Lee 2000; Koissi et al. 2006). The Lee-Carter model as applied to health care expenditures postulates that the per capita health care expenditures for a given age *x* and calendar year *t* could be modelled as a function of three sets of parameters: age-specific constants, a time-varying index and interaction terms between time and age (Lee and Carter 1992; Lee 2000). If we stratify by gender and treat gender in a similar manner as age we can denote the Lee-Carter model in the following manner:

$$\log[c_{x,g,t}] = a_{x,g} + k_t \times b_{x,g} + \varepsilon_{x,g,t}, \tag{3}$$

where $c_{x,g,t}$ denote per capita health care expenditures *c* by age *x*, gender *g* and year *t*. Equation (3) differs from a standard regression model as there are no observable variables on the right hand side. Given the data we have equation (1) states that 78 coefficients need to be estimated (21 x 2 =42 *a* coefficients; 30 *k* coefficients and 42 *b* coefficients). Lee and Carter showed how the Singular Value Decomposition (SVD) method can be used to find the least square solution. Since the model is underdetermined, the following constraints have to be put on the parameters so that a unique solution can be determined: $\sum_x \sum_g b_{x,g} = 0, \sum_t k_t = 1$. Normalising the parameters this way does not alter the model fit, in addition these constraints make the parameters interpretable: $a_{x,g}$ indicate the time-average log-per capita health care expenditures by age and gender; k_t refers to an age-independent latent HCE index indicating a general level of health care expenditures at time *t*, that is k_t quantifies the evolution of health care expenditures over time; $b_{x,g}$ can be interpreted as the interaction between each age/gender category with the general time trend. In other words, it tells us at which ages/gender HCE decline rapidly and which ages/gender HCE decline slowly in response to changes in k_t. The white noise error term $\varepsilon_{x,g,t}$, which reflects particular age- and gender specific historical influences not captured by the model, is normally distributed with mean 0. SVD is applied to the logarithm of the per capita HCE after the age-specific averages of the log HCE have been subtracted. After fitting the model, one retains a series of time-dependent k_t values for each calendar year, which can be treated as a time series and forecasts of HCE can be made by forecasting k_t and substituting values of these forecasts of k_t into equation (3). Lee and Carter proposed that specifying a time series model of a random walk with drift parameter describes k_t best. If there is a stable linear declining trend in k_t then such a model that describes the developments of k_t is sensible:

$$k_t = d + k_{t-1} + e_t \tag{4}$$

where d is the drift term, and e_t is the random part that is normally distributed with zero mean. Then predicting the future log per capita HCE can be done by substituting predicted values of k_t into the systematic part of the Lee-Carter model. Although Lee & Carter have suggested alternate specifications time series, one of the main insights of applying the Lee & Carter method was that the time parameter k_t in many data sets could be modelled simply as a random walk with drift. Plugging equation (4) into equation (3) and taking first differences we can obtain the following:

$$\Delta\log\left(c_{x,g,t}\right) = (d + e_t) \times b_{x,g} + \varepsilon_{x,g,t} \tag{5}$$

Equation (5) states that age and gender specific growth in health expenditures is the product of the random walk and the b coefficients. From equation (5) we can also see that there is a strong positive correlation of health care expenditures between age and gender categories in a particular year. While in the context of mortality modeling, this has been considered as a drawback as it assumes strong period effects, for HCE this is not a problem as strong period effects are likely as GDP is an important determinant of HCE which is determined at the macro level.

2. Hospital expenditures

2.1 Data

For predicting the hospital costs for the years 1995-1998, 2000-2002, 2004, 2006, 2008-2009, we used two data sources. First, data from the COI studies for the years 1994, 1999, 2003, 2005, 2007 and 2010 was employed. These data includes total hospital costs (expressed in millions of euros) by age, gender and disease group where the variable age contains 21 levels and the variable disease includes 18 diseases groups. Health care costs, in general, differ by disease, by age and by gender. Figure 1 (below) displays the complicated pattern by disease group, gender and age category for the hospital provider. Some diseases are only or more specific to women than to men (for example, the pregnancy related diseases). Also the pattern by age is different for each disease category. Some diseases costs are larger in the first year of life (for example, the Aandoeningen perinatale period) while others are larger at the older ages or at the adult age. Moreover, in general the total hospital costs are larger for women than for men. Figure 2 displays the pattern for two diseases, i.e. Aandoeningen perinatale period and

Zwangerschap, bevalling en kraambed, for which the hospital costs will be estimated separately from the other diseases groups. Costs for the Aandoeningen perinatale period are only made in the first year of life whereas costs for the Zwangerschap, bevalling en kraambed are made before the age of 45 years.

The second data source was the LMR data. From these data, we used information on day admission, clinical admission and average length of stay for the same age, gender and diseases group as used in the COI studies. Day admission denotes the number of patients out of 100000 patients accepted in hospital for one day, clinical admission denotes the number of patients from 100000 patients accepted in hospital for more than one day. For this last group of patients, one can calculate the average length of stay in the hospital that is the average number of hospitalized days. Figure 3 displays the patter of total day admission, total clinical admission and average length of stay in the period 1994-2010. By ‘total’ we mean the total for all diseases, all ages and both genders. We notice that the number of day admissions increased in time whereas the average length of stay decreased in time. The number of clinical admissions decreased in the period 1994-2001 and increased in 2001-2010.

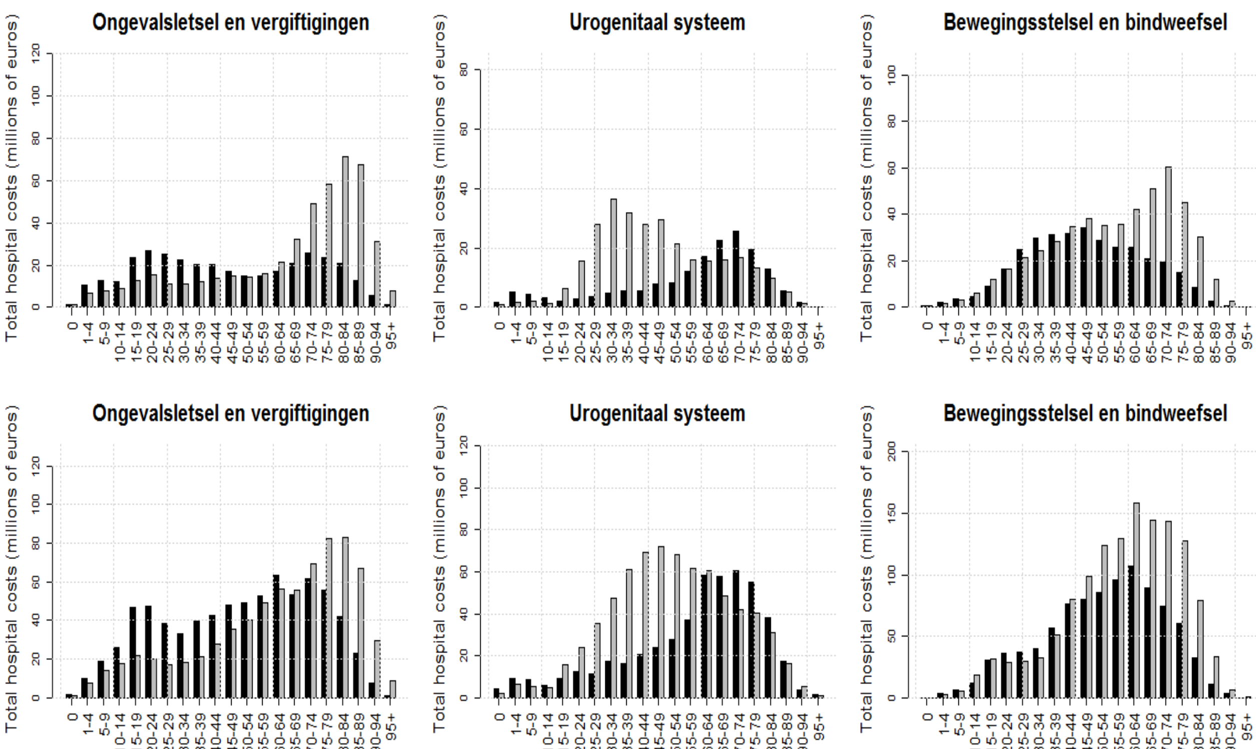

Figure 1: *Some diseases by age and gender for year 1994 (top) and for year 2010 (bottom) for men (black bars) and women (grey bars)*

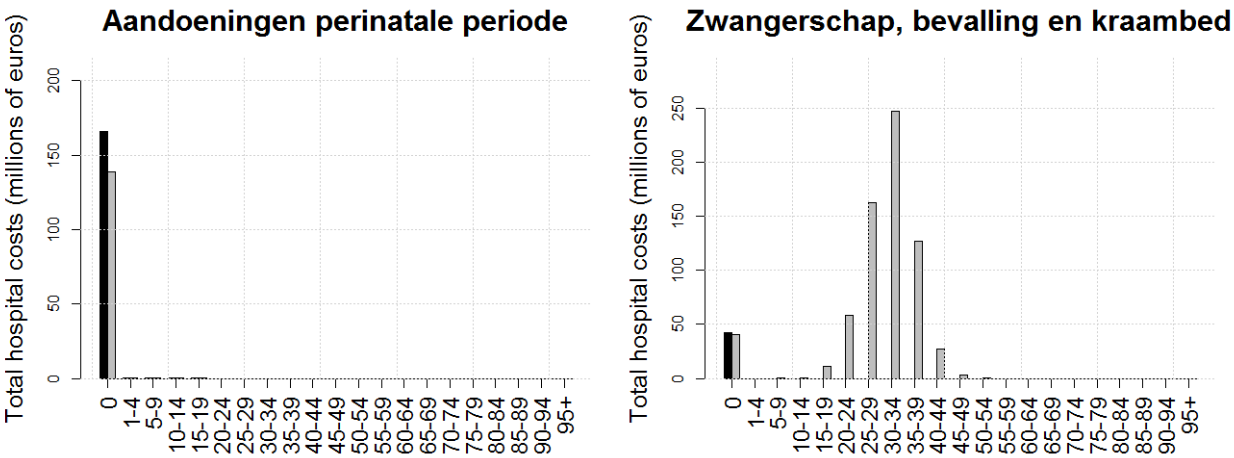

Figure 2: *Diseases by age and gender for year 1994*

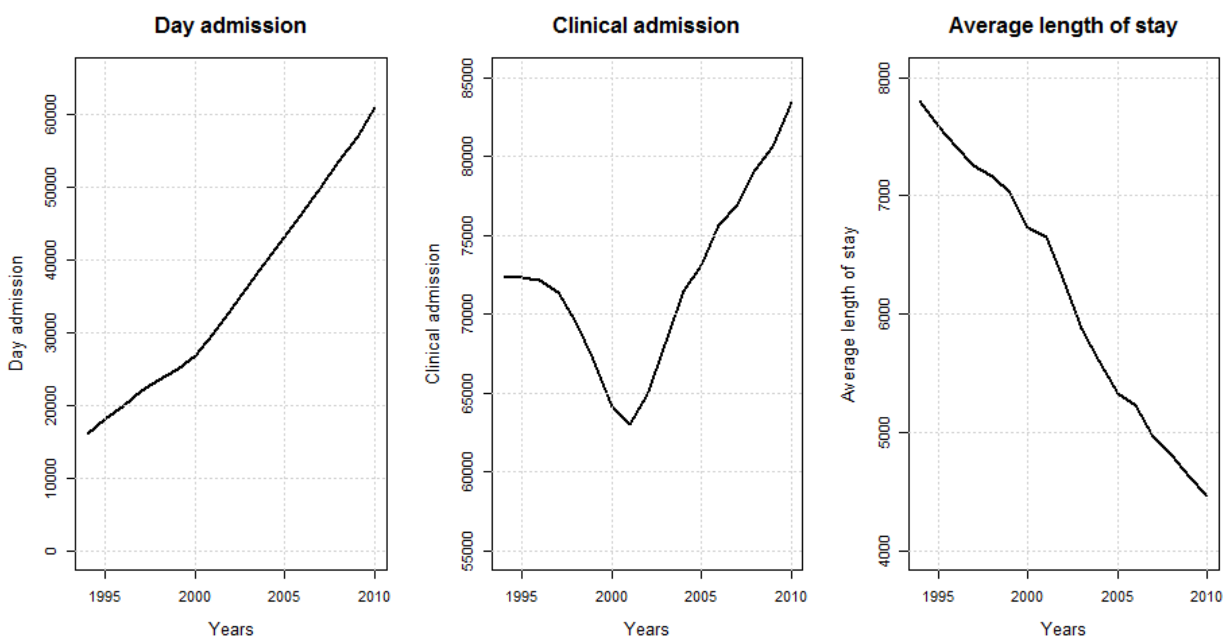

Figure 3*:Day admission, clinical admission and average length of stay for the period 1994-2010*

2.2 The prediction model

We modeled the hospital costs per each disease group as a function of covariates available from the cost of illness data and covariates available from the LMR data: age, gender, year, day admission (dayadm), clinical admission (clinadm), and average length of stay (alos). The model for each of the 16 classes of diseases, excluding the diseases Aandoeningen perinatale period and Zwangerschap, bevalling en kraambed, is:

$$\log\left(c_{x,t,g}\right) = \beta_t X_t + \gamma_{x,g} Y_{x,g} + \alpha_1 dayadm + \alpha_2 clinadm \times alos + \varepsilon_{x,t,g} \, , \qquad (6)$$

where $c_{x,t,g}$ denotes the hospital cost for age x, year t and gender g. The log-transformation of the costs improved the fit of the model to the data. In order to apply this transformation to all costs we adjusted the costs that were equal to 0 euros to 1 euro. This model was developed for the years 1994, 1999, 2003, 2005, 2007, 2010.

For the diseases Aandoeningen perinatale period and Zwangerschap, bevalling en kraambed we used interpolation of the hospital costs for the missing years.

2.3 Model fit

The fit of the model to the observed data was investigated. Figures 4 and 5 show the observed versus the fitted hospital costs by age groups and by gender for various diseases (in figure 4) and for the years available in the period 1994-2010. The missing costs' for the remaining years in 1994-2010 were estimated using the interpolation method described above. Both figures 4 and 5 indicate that models were providing a good fit to the data. Furthermore, the R-squared values for the developed models are between 0.8 and 0.95 suggesting that the explanatory variables in the model explain well the variance in the hospital costs (see figure 6).

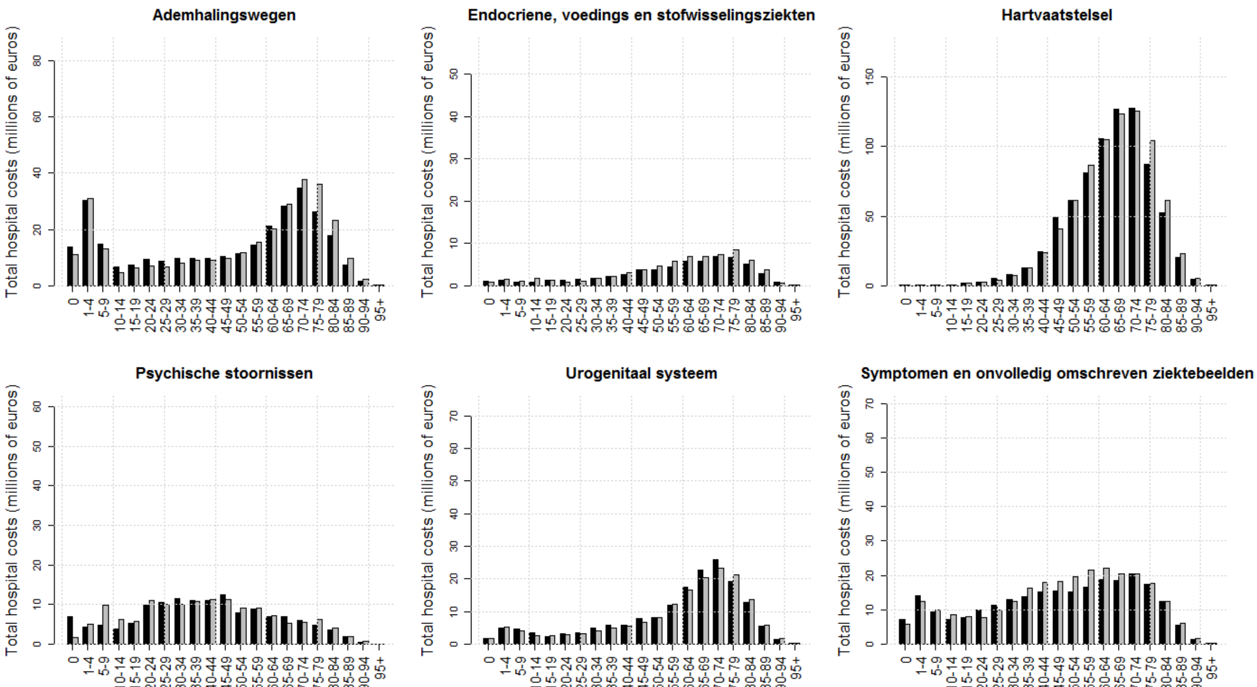

Figure 4: *Observed (black bars) versus estimated (grey bars) costs per age for various diseases. Results for men in year 2003*

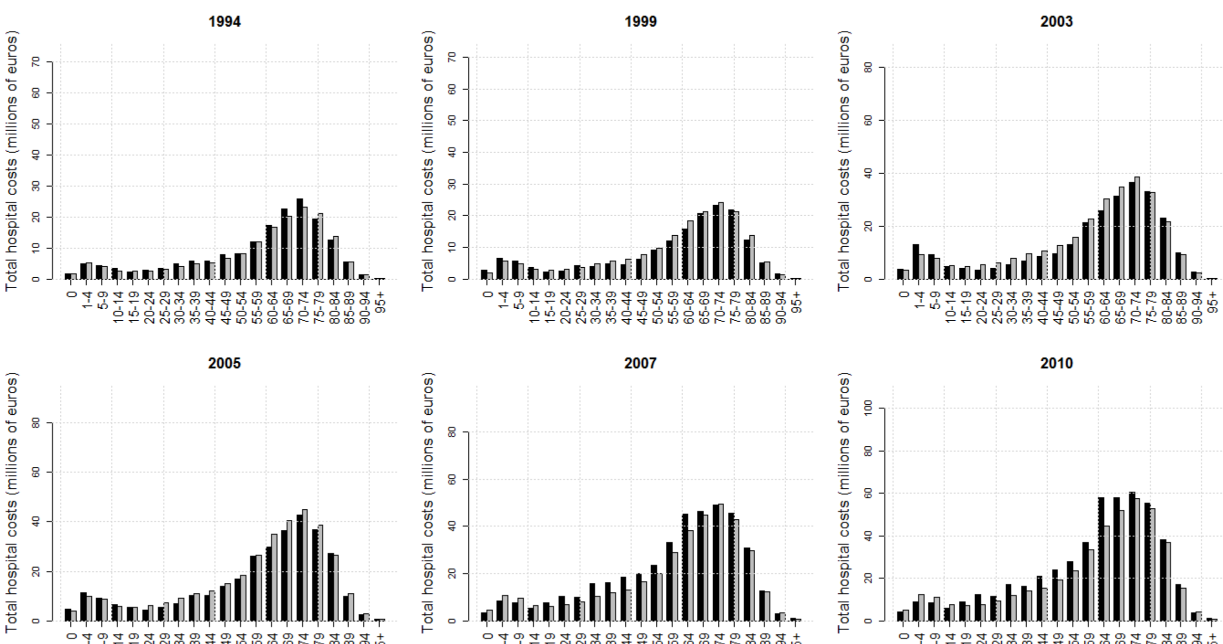

Figure 5: *Observed (black bars) versus estimated (grey bars) costs for Hartvaatstelsel for some years in the period 1994-2010. Results for men*

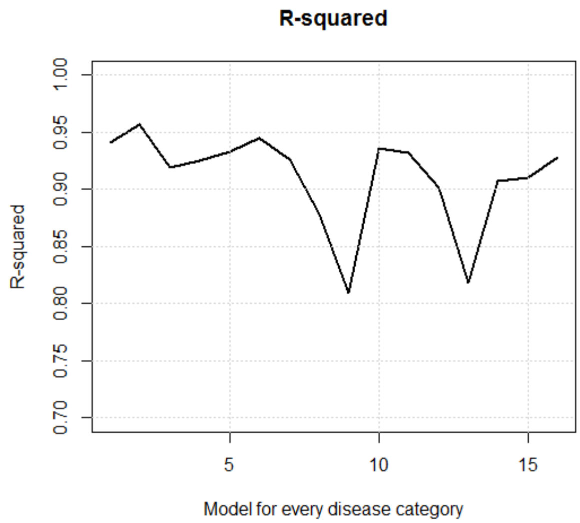

Figure 6: *The R-squared for each disease' model*

2.4 Results prediction

Figure 7 illustrates the total hospital costs available from COI data togheter with those estimated using the approach described above versus the total hospital costs published by cbs. Note that by 'total' here we mean total for all diseases, ages and both genders. CBS has published total costs only for years 1998 onwards. Figure 7 indicates that both the available data from COI and our estimates differ from the published CBS values. In the period 1998-2003 and in 2008-2010 the total hopsital costs seem underestimated while in the period 2004-2007, the total hopsital costs seem overestimated. These changes range in between 1%-5%. Therefore, we did not calibrate our models with the available totals from CBS since the COI data did not match these CBS values. With calibration we could have identical total hopsital costs with CBS, however, since the COI data shows simillar but not identical total hospital costs we did not performed this step in our analysis yet.

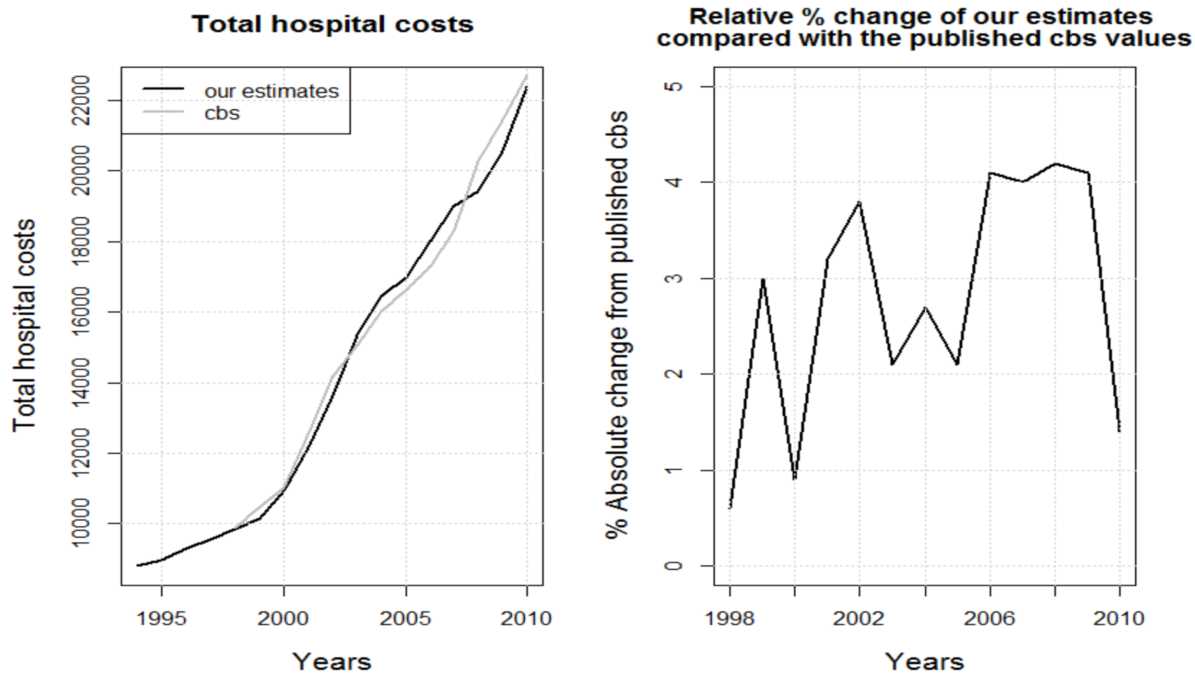

Figure 7: *Estimated total hospital costs versus cbs published costs*

Note that the results presented in this report are not adjusted with the inflation rates. Figure 8 (bellow) describes the estimated total hospital costs by age group and gender for some of the years in the period 1994-2010. We observe that total hospital costs increased over time and that the general trend by age and gender appears relatively similar for various years in 1994-2010. However, there are some important aspects of this trend which are worth noting. First, there are quite some differences between various age groups with total costs being larger in the first age group (the first year of life) and then dropping and increasing with advancing age until at the very older ages when the total hospital costs tend to decrease. Second, there are also some gender differences with women spending more the men especially between the ages 25 and 55 for all years and at a very old age, older than 80 years. A more detailed description of the total hospital costs for all estimated years in the period 1994-2010 is illustrated in table 1.

Figure 9 (bellow) presents total hospital costs and per capita hospital costs in a chosen year from the estimated period, here year 2002, for both men and women. Per capita hospital costs increased more or less exponentially with advancing age. This patter was relatively similar for both genders. Indeed, as for total costs, we observe that also per capita costs for newborn children were high. During childhood and adulthood per capita costs were low and stable; however, from age 60 onwards the per capita hospital costs were rising to almost 4000 Euro per inhabitant

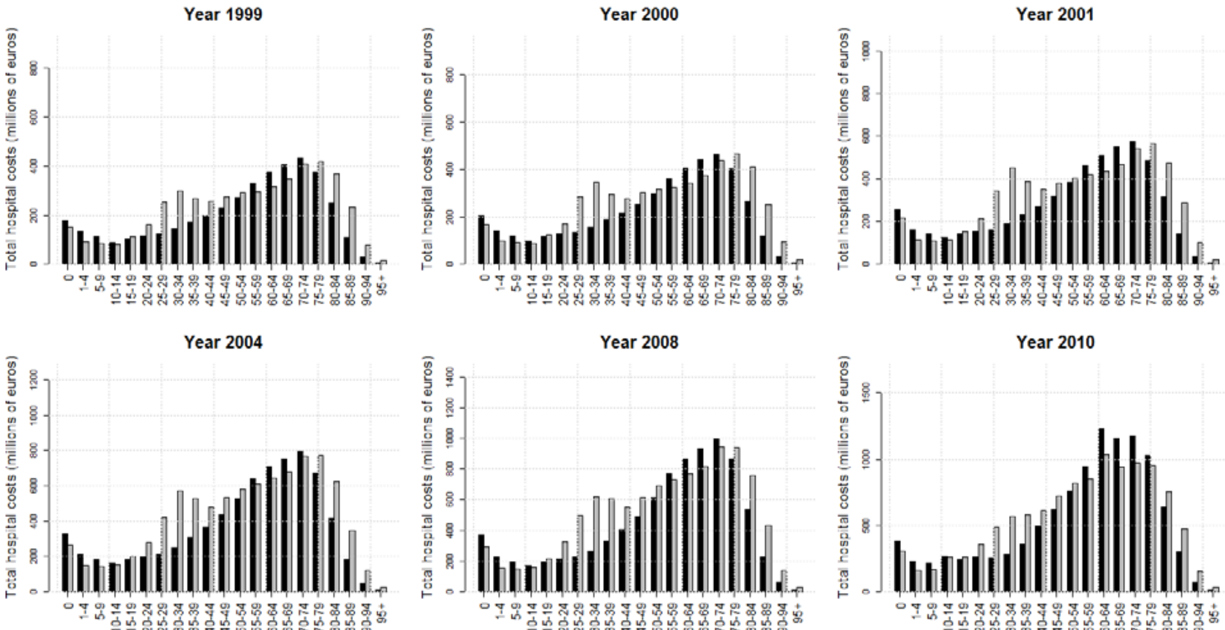

Figure 8: *Total hospital costs(for all 18 disease groups) in the Netherlands by age and gender (black bars – men, grey bar-women) for years 1999, 2000, 2001, 2004, 2004, 2008 and 2010*

Table 1: *Estimated total hospital expenditures costs by age group in millions of euros*

Age group	Year 1995	Year 1996	Year 1997	Year 1998	Year 2000	Year 2001	Year 2002	Year 2004	Year 2006	Year 2008	Year 2009
0	326	344	358	372	423	472	518	589	625	666	700
1-4	227	238	226	240	247	277	310	362	387	381	416
5-9	198	197	202	210	225	248	276	324	342	340	366
10-14	170	175	181	187	214	237	265	312	319	330	347
15-19	214	221	228	236	267	293	325	381	392	409	430
20-24	277	284	289	297	334	367	408	475	505	541	568
25-29	377	390	403	417	467	507	556	634	670	721	757
30-34	440	461	482	501	577	639	713	818	841	885	927
35-39	439	453	469	484	555	619	697	834	887	939	979
40-44	457	470	481	494	556	620	698	848	906	961	1012
45-49	505	525	540	555	628	700	792	969	1039	1110	1178
50-54	565	580	599	616	699	786	895	1109	1210	1305	1377
55-59	626	648	669	687	787	881	1002	1244	1375	1501	1583
60-64	692	710	731	750	843	944	1083	1353	1500	1638	1735
65-69	752	780	800	815	901	1013	1153	1431	1600	1752	1883
70-74	844	871	888	902	993	1113	1261	1561	1758	1942	2054
75-79	795	831	852	872	948	1053	1177	1444	1651	1808	1950
80-84	619	632	661	675	717	790	867	1041	1179	1292	1366
85-89	339	349	369	373	400	425	460	527	607	657	686
90-94	106	113	120	127	132	138	145	164	187	199	205
95+	20	20	22	24	24	25	26	30	34	37	37
Total	**8988**	**9292**	**9570**	**9834**	**10937**	**12147**	**13627**	**16450**	**18014**	**19414**	**20556**

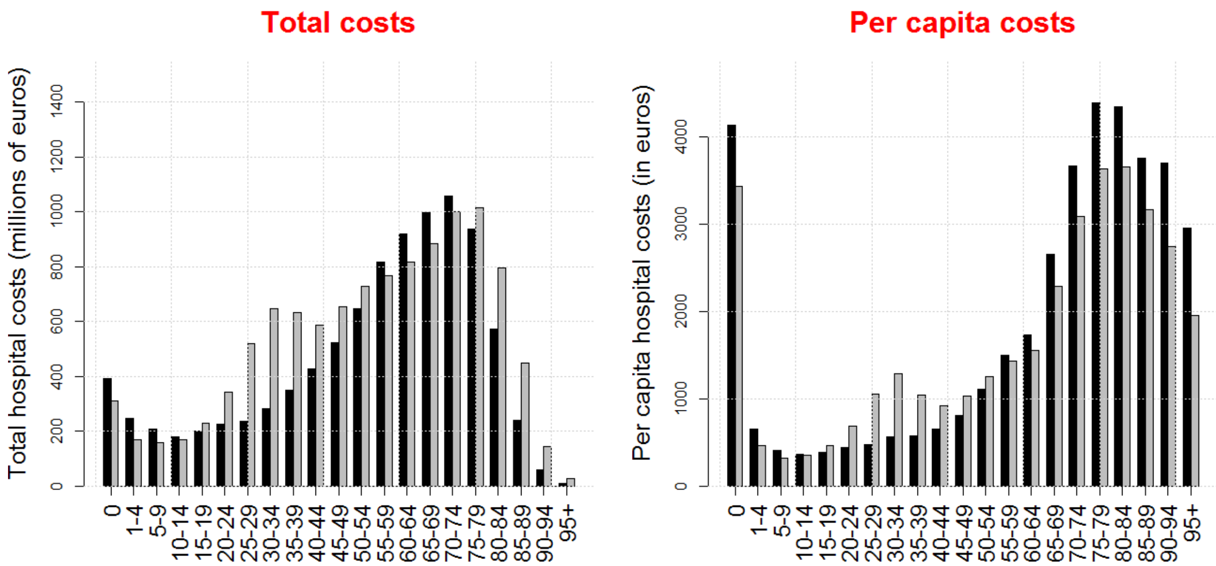

Figure 9: Total hospital costs (left) and per capita hospital costs (right) by age and gender *(black bars – men, grey bar-women) in year 2002*

Diseases and year

It is important to disentangle costs per disease group. For all diseases the total costs per year increased in the estimated years in 1994-2010 (see figure 10). Because the results were not adjusted with the inflation rates, the changes in absolute value from 1995 to 2009 should be smaller than shown in figure 6; however, changes from one disease relative to another can be observed. For example, largest increases were observed for *Hartvaatstelsel* and *Nieuwvormingen* while small changes were noticed for *Bloed en bloedvormende organen*, *Congenitale afwijkingen* and *Huid en subcutis* categories. Table 2 presents in more detail the total hospital costs by disease estimated for the period 1994-2010.

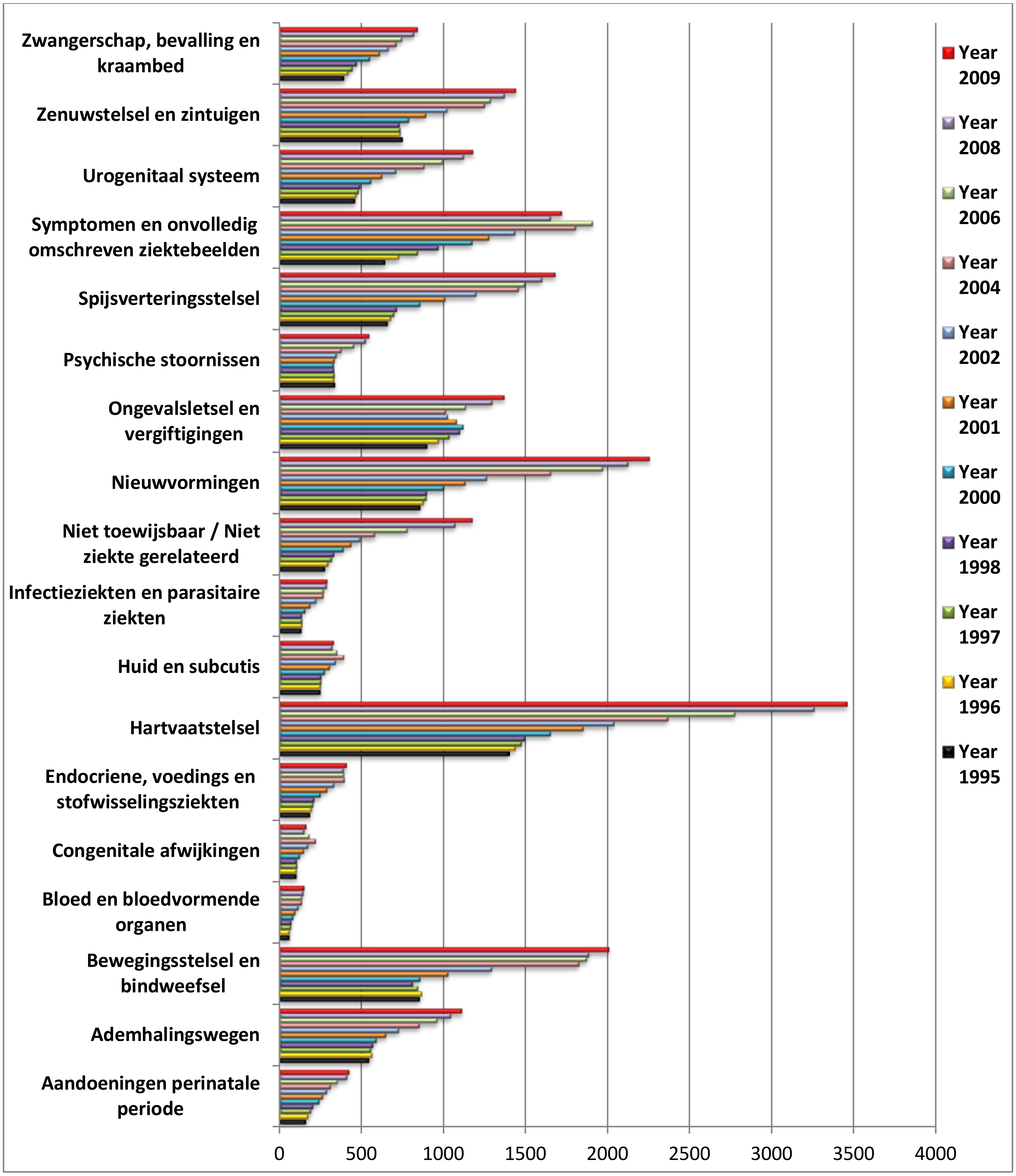

Figure 10: *Trends in total costs for each disease category (in millions of euros)*

Table 2: *Total costs per disease group by year (costs for both men and women)*

Disease category	Year 1995	Year 1996	Year 1997	Year 1998	Year 2000	Year 2001	Year 2002	Year 2004	Year 2006	Year 2008	Year 2009
Aandoeningen perinatale periode	164	177	189	202	237	260	283	308	351	408	426
Ademhalingswegen	546	565	553	566	589	645	725	849	959	1040	1112
Bewegingsstelsel en bindweefsel	853	867	844	810	854	1024	1288	1822	1869	1883	2007
Bloed en bloedvormende organen	62	64	67	69	79	92	109	133	136	142	151
Congenitale afwijkingen	107	106	105	102	119	145	171	216	181	148	165
Endocriene, voedings en stofwisselingsziekten	189	197	202	208	247	287	328	393	390	387	409
Hartvaatstelsel	1394	1431	1473	1494	1647	1850	2037	2364	2773	3255	3456
Huid en subcutis	253	252	251	250	272	302	337	390	348	317	332
Infectieziekten en parasitaire ziekten	135	138	135	132	154	182	221	264	264	283	291
Niet toewijsbaar / Niet ziekte gerelateerd	280	297	314	329	387	433	488	579	777	1068	1176
Nieuwvormingen	856	874	891	895	998	1128	1259	1651	1967	2123	2253
Ongevalsletsel en vergiftigingen	900	966	1031	1094	1116	1076	1023	1011	1130	1294	1370
Psychische stoornissen	339	335	331	329	327	331	343	376	451	521	546
Spijsverteringsstelsel	659	677	695	711	856	1006	1196	1455	1493	1595	1680
Symptomen en onvolledig omschreven ziektebeelden	645	726	839	963	1170	1274	1431	1804	1906	1647	1720
Urogenitaal systeem	460	467	477	487	553	622	707	879	992	1119	1179
Zenuwstelsel en zintuigen	750	735	732	727	785	887	1019	1246	1283	1367	1441
Zwangerschap, bevalling en kraambed	395	418	442	465	547	605	662	709	743	815	842
Total	**8987**	**9292**	**9571**	**9833**	**10937**	**12149**	**13627**	**16449**	**18013**	**19412**	**20556**

Diseases and gender

Figure 11 displays the total hospital costs in one estimated year (here we chose year 2004) by gender. We notice that for the majority of diseases women spent more hospital costs than men, exception seems to be the *Hartvaatstelsel* disease category for which men spent more in 2004. For both genders, high costs are noticed for: *Bewegingsstelsel en bindweefsel*, *Hartvaatstelsel*, *Nieuwvormingen, Symptomen en onvolledig omschreven ziektebeelden* and *Spijsverteringsstelsel* while low costs were observed for: *Bloed en bloedvormende organen, Congenitale afwijkingen, Infectieziekten en parasitaire ziekten*. Similar patterns of the estimated total hospital costs by gender were observed for other years.

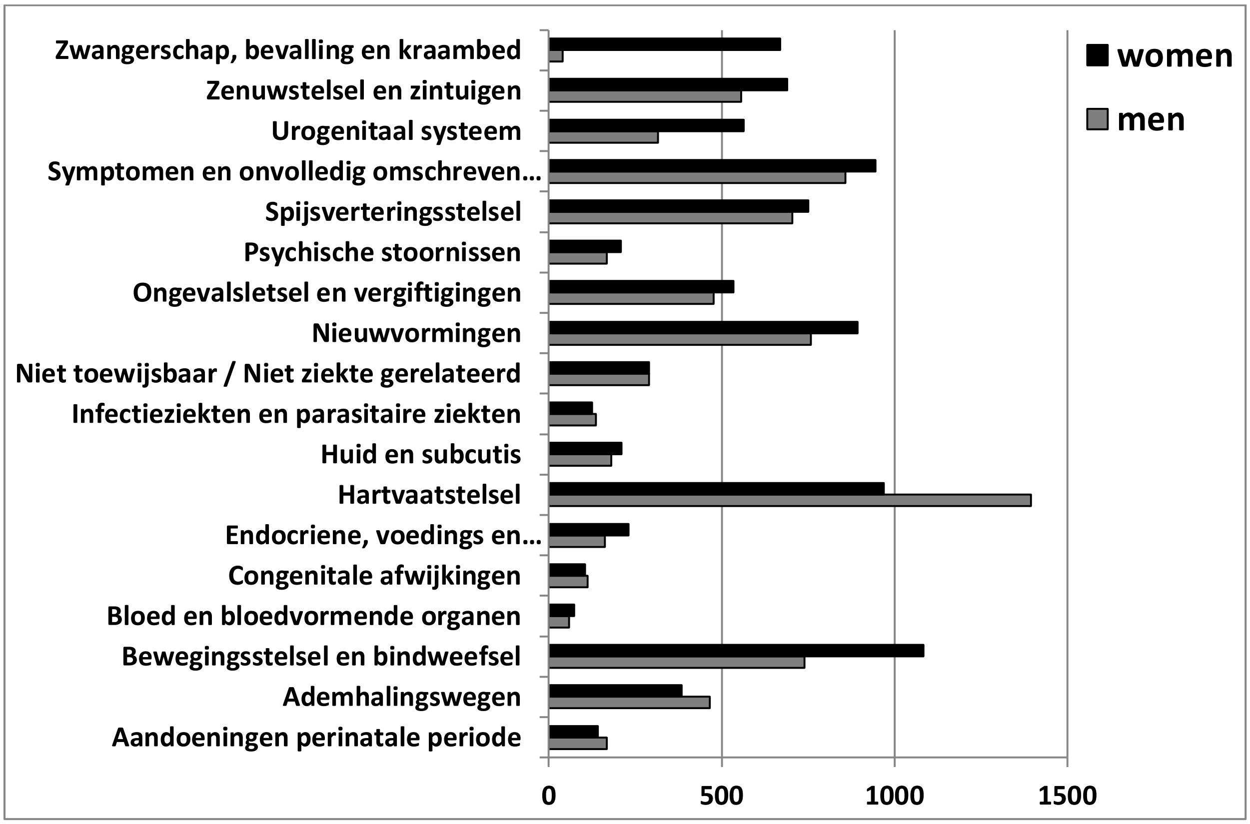

Figure 11: *Total hospital costs in 2004 by disease group and by gender (in million euro)*

Diseases and age

The hospital costs of diseases are strongly depending on age as shown in figure 12. For the purposes of this plot we collapsed the 21 age groups in just 8 main categories. Within all disease groups most care was given to people older than 45 years old (see figure 8). Exception from this are the pregnancy related disease group for women *Zwangerschap, bevalling en kraambed* and the costs associated with the first year of life *Aandoeningen perinatale periode*.

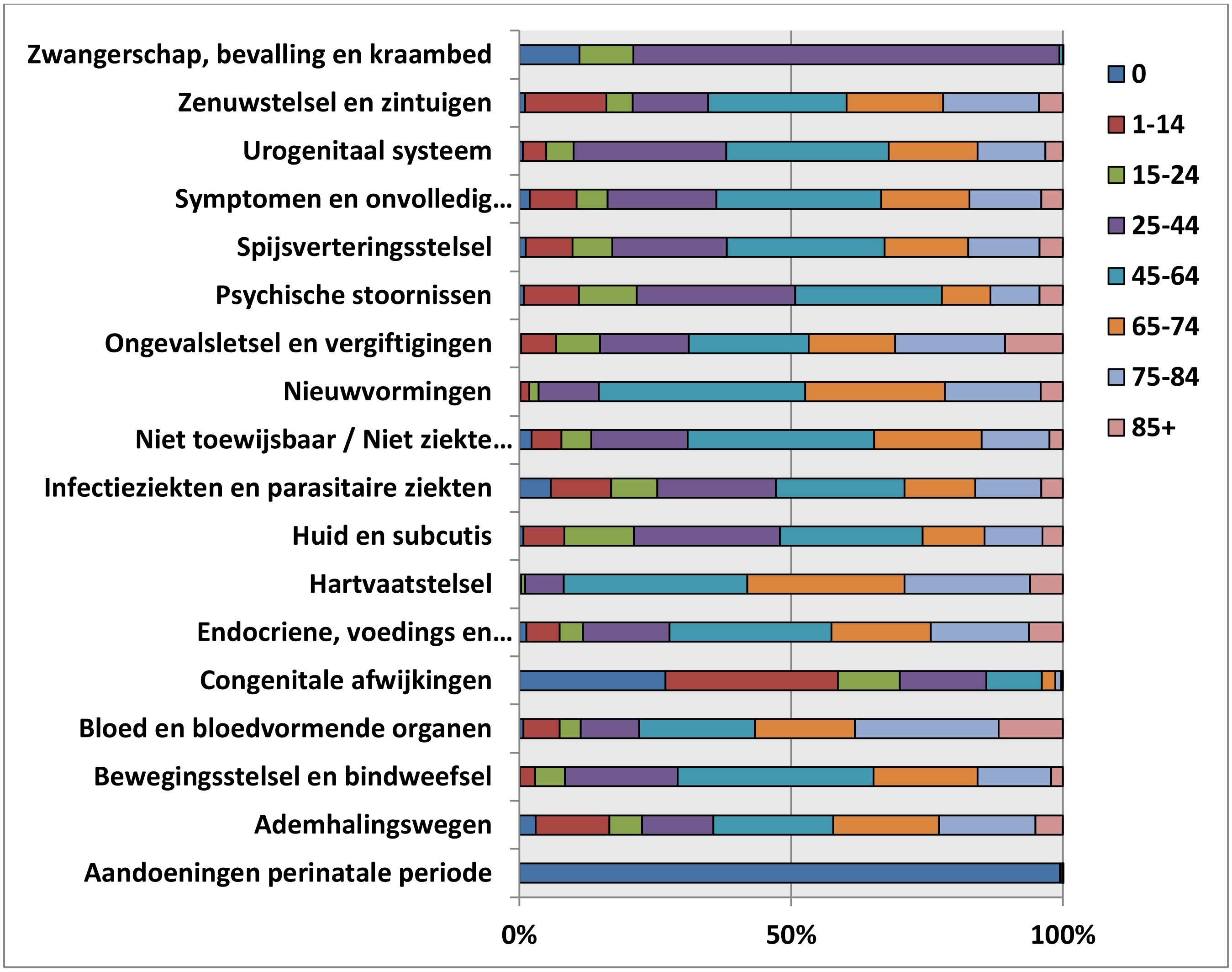

Figure 12: *Percentage of the total hospital costs in 2004 by disease group and by age (for both genders)*

2.5 Results forecasting

Figure 13 displays parameter estimates of the Lee-Carter for cardiovascular disease. The top panel displays estimates of the age- and gender-specific coefficients which simply are the time averaged log of per capita HCE for cardiovascular disease in hospital. The second panel displays the general time trend in per capita HCE, which have been increasing over time. The lowest panel indicates the age- and gender-specific interactions with the general time trend. Figure 14 shows the forecasted costs for the *hartvaatstelsel* disease costs for men in 2010-2015. The first panel presents the per capita *hartvaatstelsel* hospital costs for men at age 50. As expected, these costs will increase in 2010-2015. The second panel compares per capita hospital costs associated with *hartvaatstelsel* in 2010 and in 2015 for men for all ages. Figure 14 illustrates that, in 2010-2015, hospital costs of *hartvaatstelsel* disease will increase for all ages but especially for men older than 60 years.

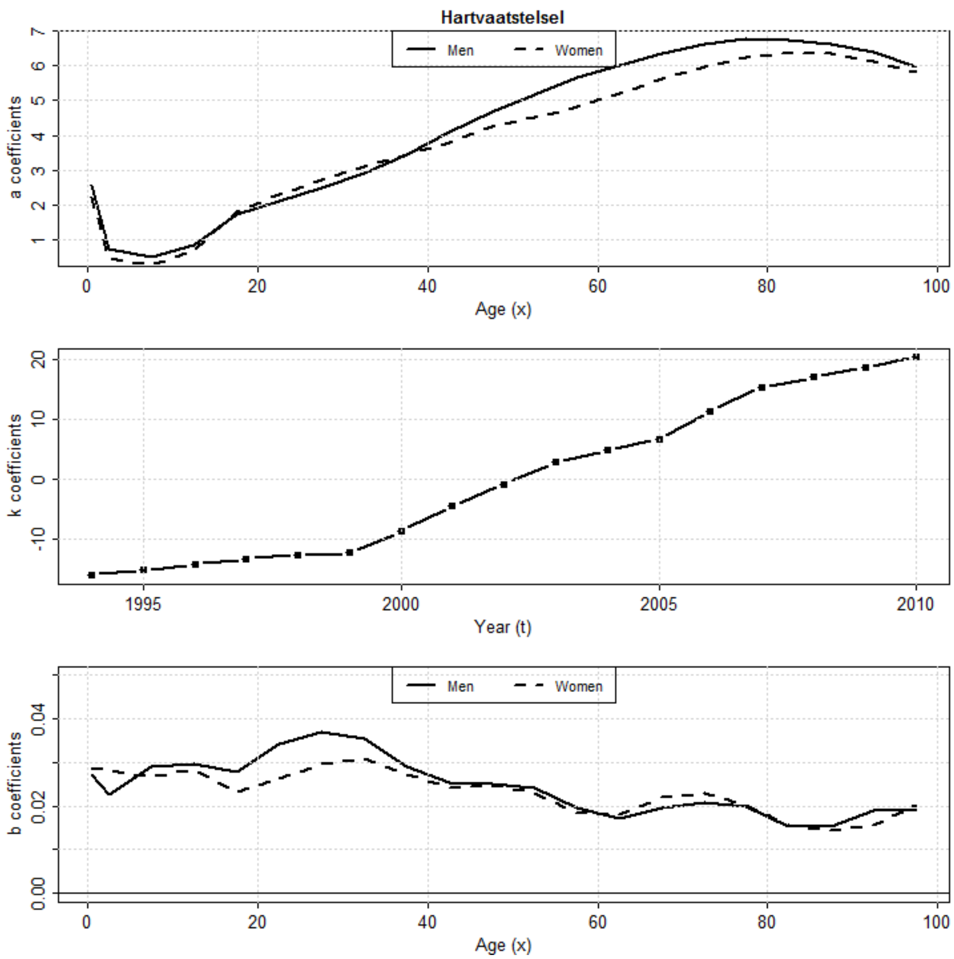

Figure 13: *Parameter estimates of the Lee-Carter for cardiovascular disease*

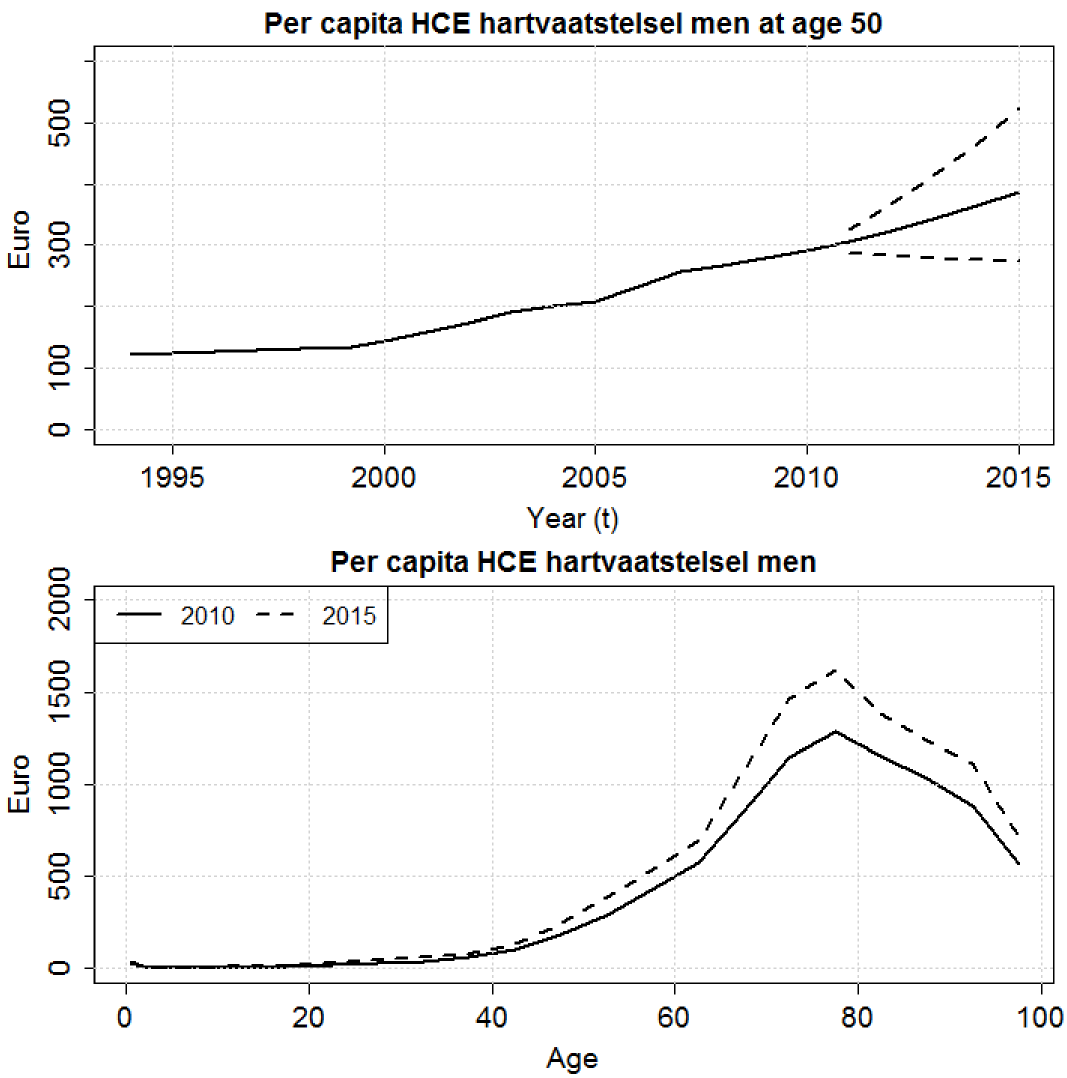

Figure 14: *Prediction of costs for the hartvaatstelsel disease for men*