Evolution of outpatient healthcare expenditure due to ageing in 2030, a dynamic micro-simulation model for France

Charlotte Geay, Grégoire de Lagasnerie, Makram Larguem

Laboratoire interdisciplinaire d'évaluation des politiques publiques

LIEPP Working Paper

« Health Policy » Research Group
May 2014, nº28

Evolution of outpatient healthcare expenditure due to ageing in 2030, a dynamic micro-simulation model for France

Charlotte Geay
INSEE
charlotte.geay@gmail.com

Grégoire de Lagasnerie
LIEPP
gregoire.delagasnerie@gmail.com

Makram Larguem
Université Panthéon-Assas, Paris 2
maklarguem@yahoo.fr

www.sciencespo.fr/liepp

EVOLUTION OF OUTPATIENT HEALTHCARE EXPENDITURE DUE TO AGEING IN 2030, A DYNAMIC MICRO-SIMULATION MODEL FOR FRANCE

Charlotte Geay

Grégoire de Lagasnerie (LIEPP)

Makram Larguem (Université Panthéon-Assas, Paris 2)

Abstract

Population ageing will be a major challenge in Europe in the coming decades. This phenomenon will raise the question of the sustainability of public spending with increasing healthcare provision costs. This paper presents a dynamic micro-simulation model for outpatient healthcare expenditure in France, which projects healthcare costs in the long run. Like all the dynamic micro-simulation models, the model projects the population structure over time. The projections are run using a transition process between three states: two non-absorbing (good-health or ill-health) and one absorbing state (death). The outpatient healthcare expenditure is estimated on data between 2002 and 2008 through a two-part model. While healthcare spending of 25 years old and more represent 3.9% of GDP in 2008, they would reach 4.6% in the baseline scenario in 2032 (+0.7 percentage point of GDP or +17.5%). A difference in the share of expenditure in GDP appears between scenarios with different evolutions of health status during the projection period. Outpatient healthcare spending represents 4.6% of GDP in the central scenario in 2032, against 4.4 % in the most optimistic scenario and 4.7% in a pessimistic scenario.

INTRODUCTION

Population ageing will be a major challenge in Europe in the coming decades due to the retirement of the baby boom cohorts and the increased life expectancy (Blanpain, *et al.*, 2010). This phenomenon will raise the question of the sustainability of public spending with increasing costs related to healthcare provision. Furthermore, health expenditure continues to outstrip economic growth in the majority of OECD countries. Due to public finance constraints, France should try to maintain the same or a better health level, without spending more. In this context, it becomes a priority to be able to evaluate how healthcare spending would spontaneously evolve and how it would evolve in function of different health status evolutions' scenarios. This paper presents a dynamic micro-simulation model which predicts the evolution of outpatient healthcare expenditure in France due to ageing and different evolution of health status of the population.

Dynamic micro-simulation models have attracted growing attention as shown by the very precise and complete overview of Zucchelli et *al.* (2012). Numerous dynamic models dealing with several issues related to healthcare have been developed. In their study Zucchelli et *al.* (2012) focused on two advanced models: Population Health Model (POHEM) and the Future Elderly Model (FEM). The first model simulates the lifecycle pathways of the Canadian Population. This model simulates risk factors associated to each individual and their links with specific diseases and the level of expenditure. The second model is based on three modules; a module for health care costs, a module for health status transition and a module which simulates the new recipients of Medicare. For a description of other models, Spielauer (2007) made a survey and review of models used in several countries.

In the model, the population structure is projected over time. Dynamic ageing through the estimation of health status transition probabilities allows updating demographic and epidemiologic characteristics of the population at an individual level over time. This dynamic ageing is implemented in discrete time because of data availability. Indeed, the data used are only available every four years and consequently do not allow implementing a continuous-time micro-simulation model. The dynamic ageing affects the epidemiologic change between different health statuses, and consequently the outpatient healthcare expenditure.

In order to compute the health status of each individual, we use an approach developed in studies focused on the link between health status and retirement decision (Bound et *al.*, 1999, Hagan et *al.*, 2006; Disney et *al.*, 2006 ; Blanchet et *al.*, 2007, Jones et *al.*, 2010) and a study about the construction of a continuous health score (Rochaix et *al.*, 2006). This method allows constructing an individual health stock from the subjective self-assessed health, a more objective measure of health conditions (disease declarations) and the social status in order to control for the declarative bias (Etilé et *al.*, 2006, Deveaux et *al.*, 2008 for notably a literature review and also Bago d'Uva et *al.*, 2008). From the results of this method, we estimate a health score using the technique developed by Rochaix et *al.* (2006): it consists in estimating the influence of the degree of severity of the different diseases from which the individual suffers on the self-assessed health declared in the survey after having controlled for a declarative bias.

Then, to run the projections, we consider a first order Markow process between three states: two non-absorbing ones (good-health or ill-health) and an absorbing one (death). To populate the transition matrix, we estimate transition probabilities thanks to a new panel dataset built with the merger of four datasets of the Social Protection and Health survey between 2002 and 2008. In this specific survey, people are surveyed every four years. Consequently, we are able to follow change in health status for individual in the panel every four years. We use this original database for France in order to estimate transition probabilities depending on age, gender and health status. For the Markow process, the occurrence of a transition is established through a Monte-Carlo simulation with 25 iterations in order to take into account the uncertainty linked to the draw of a random variable.

The starting population for the projections comes from two different sources of data. The social and health characteristics of each individual are included in the Social Protection and Health Survey from the Institute for Research and Information in Health Economics. These data are paired with a dataset from the National Health Insurance Fund which furnishes the level of outpatient healthcare expenditure for each individual during the year. The baseline dataset is representative of the metropolitan population of France in 2008 in age, gender, social affiliation and household size.

Finally, we decide the causal relationship between the different events simulated in the dynamic micro-simulation model. We first simulate the health status of each individual thanks to the transition probabilities. Then, taking account of the health status of the individual, we associate an outpatient healthcare expenditure which is estimated on the data between 2002 and 2008. After having reviewed the different estimation techniques, we use a two-part model where the probabilities to consume is estimated through a probit model. Then we run a GLM estimation on consumers' outpatient healthcare expenditure. Using the results of this estimation, we calculate the outpatient healthcare expenditure for each individual at each period. We validate our model thanks to the official population projections.

In 2011, spending on outpatient care (ambulatory care and medical goods) reached €-96 billion in France, that is 53% of total healthcare consumption (outpatient and inpatient care) and 4.8% of GDP. 67% of the ambulatory care is paid by the mandatory public health insurance, 21% by the private health insurance companies and 12% by households as out-of-pocket. Medical goods are supported for 61% by public insurance, 21% private insurers and 18% by households. Like in the Future Elderly Model (FEM, CMS/RAND) (Zucchelli et *al.*, 2012), we only focus in the projections on a particular population, older than 25 years old[1]. In 2008, the starting point of our projections, insured persons aged 25 and over represent almost 70% of the population and more than 86% of total outpatient expenditure[2] *i.e.* 77 billion euros in 2008 or 3,9% of GDP.

The use of a dynamic micro-simulation model will improve the analysis of the evolution of outpatient healthcare expenditure due to ageing as it will allow identifying expenditure trajectories at an individual level and thus focus some further studies on redistributive effect of healthcare insurance system. In order to reach this objective the model builds in this article will be adapted to fit in the DESTINIE 2 model, the micro-simulation model of the National Institute of Statistics and Economic Studies (Blanchet et *al.*, 2011). Thanks to this merge, we should complete the demographic module by simulating birth and professional pathways thanks to the implementation of the model built in DESTINIE 2. The implementation of this module in DESTINIE 2 will also allow taking into account more variables in the simulation of healthcare expenditure but also

[1] Indeed, our model does not integrate a birth simulator until the matching with DESTINIE 2.
[2] The 50 years old and older expenditure represents 61.4% of total outpatient healthcare expenditure.

in the calculation of the transition probabilities. A study on this merge should be released before the end of 2013 (Geay, Koubi, Lagasnerie, 2013, see annex 4 for the first results).

Our paper is structured as follows. In Section 1, we describe the data. In section 2, we define the health status variable. In Section 3, we describe our conceptual framework and the estimation of healthcare demand. In Section 4, we discuss the results of the projections notably due to the ageing population and we conclude in Section 5.

1. DATA DESCRIPTION

We use three different databases in this paper which all stem from the Social Protection and Health survey. This survey details social, health conditions of each individual. The health conditions are measured by both subjective and more objective variables. Thus, we have in the data the self-assessed health but also disease declarations during the year of the survey. The survey gives also the level of healthcare expenditure for a part of the sample (around 50% in function of the survey). The expenditure is divided by type: general practitioners, specialists, drugs. For each aggregate, we know the total expenditure, the reimbursement by the National Health Insurance Fund and the out-of pocket.

First, to compute the health status, we use the data on the four years between 2002 and 2008. It contains 87 714 observations (20 831 for 2002, 22 460 for 2004, 22 150 for 2006 and 22 273 for 2008). Among the health characteristics of the individual surveyed are the vital risk and the incapacity degree of all the diseases declared by each individual[3]. The diseases' database has 233 737 observations (50 408 for 2002, 60 113 for 2004, 61 769 for 2006 and 61 447 for 2008). The vital risk associated to each disease is classified in five classes whereas the degree of incapacity is sorted in six levels (Table 1). As a health indicator, we also have information on self-assessed health in each survey. This variable is coded in five categories Very good, Good, Fair, Poor, Very Poor (Table 2). The large majority of the population evaluates his health as Excellent or Good (more than 70 % each year). We notice that for individuals present in 2002 and 2006 or in 2004 and 2008, the share of people with an Very good or Good self-assessed health diminishes between the two samples.

[3] Given the difficulty to achieve a synthesis of the health status of individuals from many diseases or the detailed nature of these diseases, researchers in the Institute for research and information in health economics, doctors and statisticians have developed two synthetic indicators of disease: the vital risk and the incapacity degree (Mizrahi and Mizrahi An Ar, 1985). Vital risk corresponds to the death likelihood level and is built from illness and individual risk factors. It is rated on a 6-point scale from "no vital risk" to "high risk" which reflects an 80% probability of death within five years. Incapacity degree is rated on an 8-point scale from "no discomfort" to "No autonomy".

Table 1: Definition of vital risk and degree of incapacity associated with each disease

Vital Risk	
0	No vital risk
1	Very slight life risk
2	Slight life risk
3	Vital Risk Possibility
4	Likely High Risk
5	High risk

Incapacity degree	
0	No discomfort
1	Very small discomfort
2	Small discomfort
3	Discomfort but normal life
4	Diminution of the daily life activities
5	Reduced activity
6	No autonomy

Source: Mizzrahi et *al.* (1997)

Table 2: Self-assessed health, relative to year (%)

	Percentage of whole sample reporting relative health status					Percentage of sample present in 2002 and 2006 or 2004 and 2008 reporting relative health status			
	2002	2004	2006	2008		2002	2004	2006	2008
Excellent	25.4	28.4	31.3	28.5		26.3	23.5	18.7	21.6
Good	47.5	51.1	48.5	49.5		49.2	54.9	55.2	52.3
Fair	21.9	17.6	16.8	18.6		20.8	18.9	21.9	21.7
Poor	4.0	2.5	2.7	2.9		2.9	2.5	3.3	3.9
Very Poor	1.2	0.4	0.7	0.5		0.8	0.2	1.0	0.5

Source: ESPS 2002, ESPS 2004, ESPS 2006 and ESPS 2008

For the estimation of transition probabilities, we use a non-balanced panel. The panel covers four different years (2002, 2004, 2006 and 2008). Each individual is surveyed every four years. Thus, for each individual present in 2002 and 2004, we have either two observations (2002 and 2006 or 2004 and 2008) or one observation (2002 or 2004) and the reason why the next year is missing (death or attrition[4]). The 2002-2006 non-balanced panel contains 11 918 observations; the 2004-2008 non-balanced panel 12 630[5]. Among people observed in 2002, we follow 6 954 observations in 2006, 259

[4] Different reasons (other than death) explain the attrition: not reached, refusal, change of social security fund, change of household composition and "out-of-field". "Out-of-field" mainly refers to the settlement in institution.

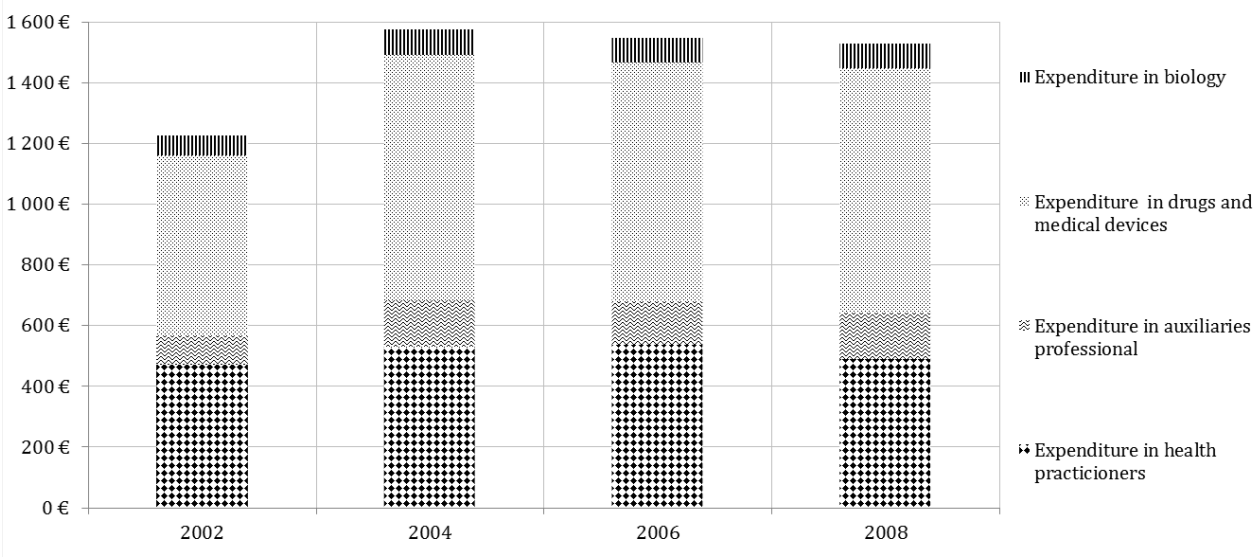
[5] The difference between the number of the observations in the baseline datasets and the non-balanced-panel is due to the presence of individual with the same identity number but with non coherent caracteristics as age and gender or non available data.

died and 4 695 disappear from the sample for another reason. Among people observed in 2004, we follow 7 627 observations in 2008, 264 died and 4 739 are considered as attrition. 51.2% of the attrition are women against 51.5% for people in the panel. Due to settlement in institution, we observe that people who don't answer the two years are older (4.2% of the attrition is 80 years old or more against 2,1% of people in the panel). Consequently, people in the attrition are also more in ill-health[6] (58.7% against 50% for people in the panel).

For the estimation of outpatient healthcare demand, we use data form 2002 to 2008 on individual for who health status and outpatient healthcare expenditure are known[7]. As for some individual the data of the National Health Insurance Fund are not paired with the Social Protection and Health Survey, the data contains 26 677 observations. The average level of outpatient expenditure reaches €-1284 per year for the whole population. The consumption for women is equal to €-1384 per year whereas it is €-1175 for men. The level of expenditure in 2013 euros is very close in the last three periods. However the level of expenditure is smaller in 2002. The first reason could be the percentage of women in the sample. Women have a higher average outpatient healthcare expenditure and they represent only 48% of the sample in 2002 against around 51% in the last three periods. For instance in 2004, the average outpatient healthcare expenditure is around 30% higher for women compared to men.

[6] Ill-health refers to the health status calculated in the next section.
[7] The data matching between the survey of the Institute for research and information in health economics and the administrative data of healthcare expenditure is only done for one individual in the household. This explains that the database uses in this part is smaller than the database used for the calculation of the health status.

Figure 1: Average outpatient healthcare expenditure per type of care (€, 2013)



Source: ESPS-EPAS 2002 ESPS-EPAS 2004 ESPS-EPAS 2006 ESPS-EPAS 2008

The baseline dataset for the projections comes from the Social Protection and Health survey 2008 from which we select only individuals for whom we know the health status. The database is then weighted to be representative of the metropolitan French population in 2008 according to four main variables gender, age, social security affiliation and household size (Table 4). The two variables age and gender are the basis for a demographic projection of the population. Consequently, it is important to have a representative population in age and gender. Then, the social security affiliation is important for two main reasons. First, it could allow tackling in the future funding issues between the different schemes. Besides the social security affiliation could be seen as a proxy for the professional occupation of the individual in the sample or a proxy for social background.

Table 4: Metropolitan French population in 2008 according to gender, age, household size and social security affiliation

	Metropolitan French population in 2008
Gender	
Women	52%
Age	
50 to 54 years old	7%
55 to 59 years old	7%
60 to 64 years old	5%
65 to 69 years old	4%
70 to 74 years old	4%
75 to 79 years old	4%
80 years old and over	5%
Household size	
1	34%
2	33%
3	15%
4	12%
5 and more	7%
Social security affiliation	
General Scheme	87%
Agricultural Scheme	6%
Self-employed workers Scheme	5%
Others	3%

Source: National Institute of Statistics and National Health Insurance Fund

2. HEALTH STATUS AND TRANSITION PROBABILITIES

In order to predict the evolution of the health status of the population, a binary health variable (good health, bad health) is created by dichotomizing a continuous health stock. Following the seminal work of Bound et *al.* (1999) and the literature about the link between health status and retirement decision (Hagan et *al.*, 2006 ; Disney et *al.*, 2006 ; Roberts et *al.*, 2006, Blanchet et *al.*, 2007), we first determine an health stock for each individual by analysing the relationship between self-assessed health, exogenous personal characteristics (such as age, gender, job status), a more objective health measure (disease declaration) and unobservable variable in order to predict a health stock free of any justification bias. The purpose is to create an unbiased and synthetic health stock for each individual summarizing the disease declaration and the information on the self-assessed by the individuals.

The Social Protection and Health Survey data contain a variable giving information of the health perceived by the individuals surveyed: it can take 5 values (Very good, Good, Fair, Poor, Very poor) and represents the answer to the question 'In general would you say that your health is…very good, good, fair, poor or very poor?'. We also have data on the illnesses of the individuals. They are recorded according to the ICD nomenclature (International Classification of Diseases). We use Com-Ruelle et *al.*'s (1997) methodology to attribute to each disease a minimal vital risk and a minimal incapacity degree. These variables are defined as the vital risk, incapacity degree of an individual suffering only from this disease: without any other information (that could only aggravate the severity of the disease), this minimal variable could be attributed to the individual himself. Then, following Rochaix et *al.* (2006), we aggregate these two variables into a unique variable of disease's severity. Hence, to each illness is attributed one severity level, ranging from 0 to 9 whose values are not equidistant.

Table 5: Definition of the severity level of each disease in function of their respective vital risk and incapacity degree

		Minimum incapacity degree						Total/line
		0	1	2	3	4	5 or 6	
Minimum Vital Risk	0	26336	60882	54784	15290	2565	88	159945
		11,73%	27,12%	24,40%	6,81%	1,14%	0,04%	71,24%
		16,47%	38,06%	34,25%	9,56%	1,60%	0,06%	
		76,58%	98,68%	61,18%	51,51%	34,16%	5,25%	
	1	8003	721	10393	1544	588	3	21252
		3,56%	0,32%	4,63%	0,69%	0,26%	0%	9,47%
		37,66%	3,39%	48,90%	7,27%	2,77%	0,01%	
		23,27%	1,17%	11,61%	5,20%	7,83%	0,18%	
	2	26	27	23994	8594	559	192	33392
		0,01%	0,01%	10,69%	3,83%	0,25%	0,09%	14,87%
		0,08%	0,08%	71,86%	25,74%	1,67%	0,57%	
		0,08%	0,04%	26,80%	28,95%	7,45%	11,46%	
	3	23	67	375	3977	2831	773	8046
		0,01%	0,03%	0,17%	1,77%	1,26%	0,34%	3,58%
		0,29%	0,83%	4,66%	49,43%	35,19%	9,61%	
		0,07%	0,11%	0,42%	13,40%	37,71%	46,12%	
	4 or 5	0	0	0	281	965	620	1866
		0%	0%	0%	0,13%	0,43%	0,28%	0,83%
		0%	0%	0%	15,06%	51,71%	33,23%	
		0%	0%	0%	0,95%	12,85%	36,99%	
Total/ column		34388	61697	89546	29686	7508	1676	224501
		15,32%	27,48%	39,89%	13,22%	3,34%	0,75%	100%

Source: Rochaix et *al.* (2006) and ESPS 2002 to 2008
Reading: First line, first column, 16.47% of disease with a minimum vital risk of 0 have a minimum incapacity degree of 0. The diseases with a minimum vital risk of 0 and a minimum incapacity degree of 0 or 2 have an aggregate severity level of 1.

Then, for all the individuals for which we know the diseases and their severity' level, the following model could be estimated, using an ordered logistic regression:

$$n_i^* = \sum_j \beta_j X_{ij} + \sum_j \gamma_j m_{ij} + \sum_{t=1}^{4} \delta_t a_k + u_i$$

where :

- n_i n_i^* is the latent (continuous) variable of n_i, the health status assessed by the individuals themselves ;
- X_{ij} are individual socio-economic characteristics of i : age, gender, occupation, additional health measure notably inpatient consumption;
- m_{ij} is the number of diseases of severity j from which suffers i (j ranging from 1 to 9) ;
- a_i is a dummy variable relative to the year when i is being surveyed.

Table 6: Results of ordered logit estimating self-assessed health status as a function of objective health measures and individual characteristics

Parameter	Coefficient		Standard error	$Pr > \chi^2$
Intercept 1	-8,7752	(***)	0,0826	<0,0001
Intercept 2	-6,5283	(***)	0,0578	<0,0001
Intercept 3	-3,4993	(***)	0,0461	<0,0001
Intercept 4	-0,3252	(***)	0,0422	<0,0001
Nb. maladies sans indication de gravité	0,2530	(***)	0,0206	<0,0001
Number of disease with severity level equal to 1	0,0593	(***)	0,00861	<0,0001
Number of disease with severity level equal to 2	0,2883	(***)	0,00682	<0,0001
Number of disease with severity level equal to 3	0,1417	(***)	0,0314	<0,0001
Number of disease with severity level equal to 4	0,1975	(***)	0,0211	<0,0001
Number of disease with severity level equal to 5	0,4620	(***)	0,0115	<0,0001
Number of disease with severity level equal to 6	0,6161	(***)	0,0396	<0,0001
Number of disease with severity level equal to 7	0,7873	(***)	0,0230	<0,0001
Number of disease with severity level equal to 8	0,6579	(***)	0,0766	<0,0001
Number of disease with severity level equal to 9	0,9610	(***)	0,0519	<0,0001
Age	0,0244	(***)	0,000961	<0,0001
Gender (ref. = Woman)				
Man	-0,00453		0,0198	0,8187
Employment status (ref. = pensioner)				
Working person	-0,1133	(***)	0,0176	<0,0001
Unemployed	0,4753	(***)	0,0349	<0,0001
Student	-0,2273	(***)	0,0341	<0,0001
Other inactive	0,5410	(***)	0,0340	<0,0001
At least one inpatient care				
Inpatient expenditure different from zero	0,3943	(***)	0,0341	<0,0001
Year (ref.= 2008)				
2002	-0,1691	(***)	0,0359	<0,0001
2004	0,0807	(***)	0,0289	0,0052
2006	-0,0529	(**)	0,0265	0,0464
				Number of observations : 46669
Testing Global Null Hypothesis: BETA=0				
Test	χ^2			$Pr > \chi^2$
Likelihood Ratio	21387,2070			<0,0001
Score	16717,7057			<0,0001
Wald	16408,1312			<0,0001

Note: (***) = coefficient significantly non-zero at the 1% (**) = coefficient significantly non-zero at the 5% (*) = coefficient significantly non-zero at the 10%.
Reading note: The ordered logit model is estimated by modeling the different values of health declared in ascending order. Modality 1 is a very bad condition and modality 5 very good health. Thus, an estimated positive coefficient represents a positive relationship with being in poor health. The number of diseases of varying severity is indeed positively correlated with the probability of reporting poor health. For the intercept 1, the log odds ratio corresponds to the declaration of a very poor health compared to other health conditions when the explanatory variables take the value 0.

Using Rochaix et *al.'s* (2006) methodology, the variable of health status is deduced from the construction of a continuous health grade. The aim is to build an aggregated variable giving information on the health status of each individual, but also to eliminate at most the individual declaration bias as well as the influence of the socio-economic status on the perceived health status. Thanks to the results of the estimation, we can compute a new health score, s_i, independent from the individual subjective declaration bias and the influence of the socio-economic characteristics by controlling in the regression by these variables and the non-observable variables through the residual.

$$s_i = \sum_j \hat{\gamma}_j m_{ij}$$

This grade is finally normalized following again Rochaix et *al.* (2006)'s methodology to obtain a health indicator ranging from 0 to 100. People whose level is lower than 90 (over 100) are considered in bad health (the threshold of 90 is close to the mean value). This dichotomization will allow calculating transition probabilities between these two states of the nature, bad health and good health.

Table 7: Mean value of the normalized health indicator by year

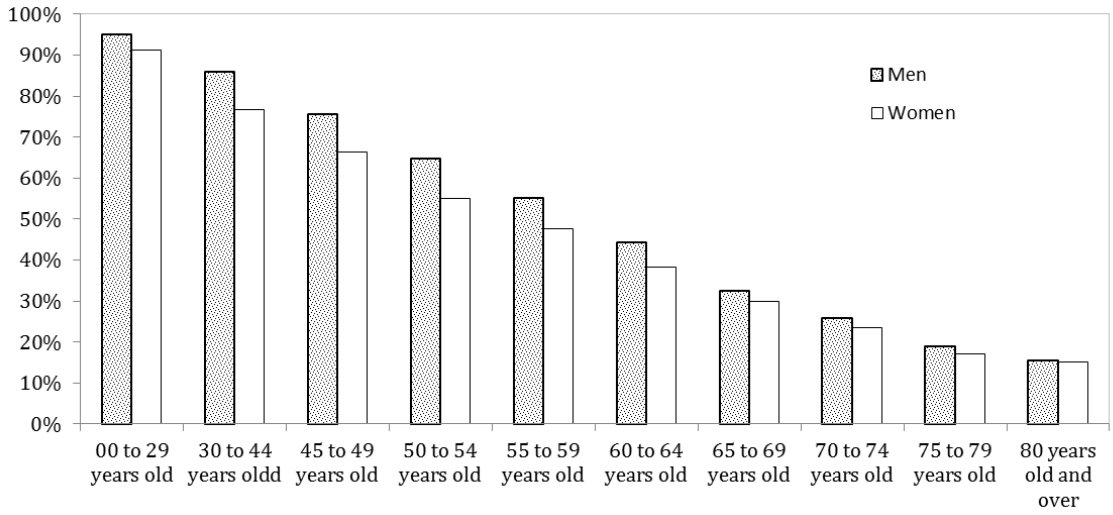
	2002	2004	2006	2008
Mean value of the normalized health indicator	93	91	91	90
Limit 1st quartile	90	88	87	86
Median	95	94	94	93
Limit 3rd quartile	98	98	97	97

Source: ESPS 2002 to 2008

Consequently, we obtain a health status variable for 54 876 individuals (we lose 32 838 observations from the initial dataset of 87 714 persons). It divides the population in good (68%) and bad (32%) health. As expected, the share of people in good health within a cohort decreases with age (Figure 2). At a given age, men are in relative better health than women, which is consistent with recent developments of the literature (CESE, 2010). In addition, there is a strong connection between suffering from a chronic disease and being in poor health. 80% of people who suffer from long-term illness are in

poor health, it is the case for only 26% of people who do not benefit from an exemption for a chronic disease. 95% of healthy people do not suffer from a chronic disease and 65% of those in poor health have a chronic disease. Lastly, 90% of people in good health according to our score declared to be in excellent health or in good health. 92% and 90% of people who declared to be respectively in very poor health or in poor health are classified in ill-health through our score.

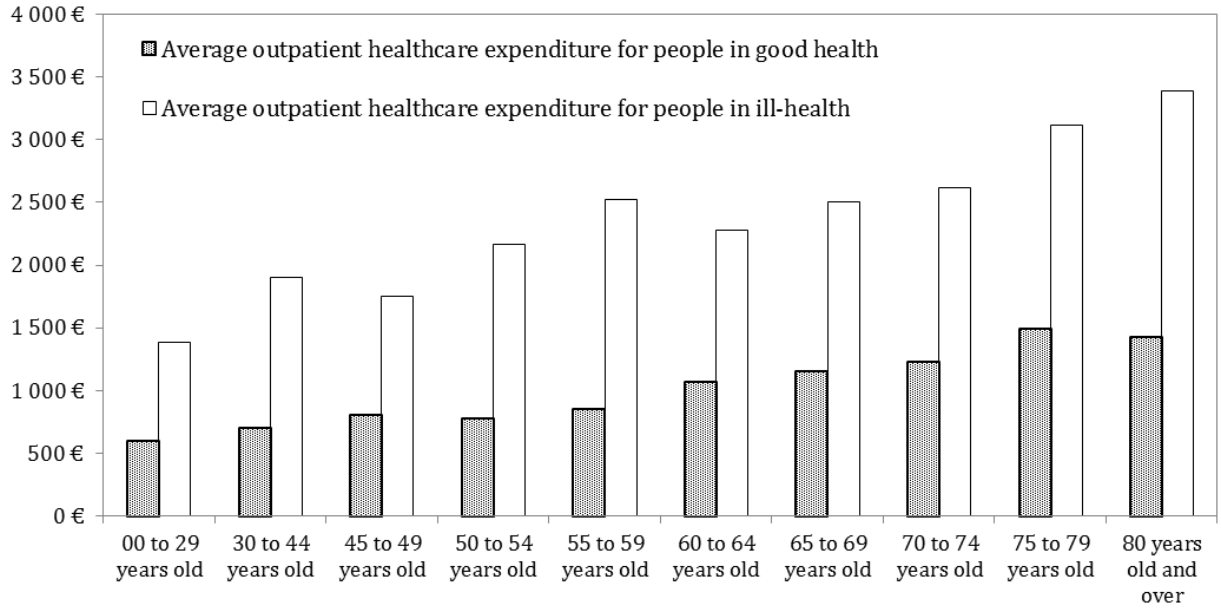
Figure 2: Share of population in good health, by gender and age group



Source: ESPS 2002 to 2008

People in good health have lower outpatient healthcare expenditure consumption than people in ill-health (Figure 3). In average the total outpatient healthcare expenditure for an individual in ill-health is more than twice the level of expenditure for somebody in good health. For instance, an individual who is between 75 to 79 years old has an average expenditure of around €-1 500 against around €-3100 if he is the same age but in ill-health.

Figure 3: Average outpatient healthcare expenditure in good health and in ill-health, by age group



Source : ESPS-EPAS 2008

To generate a population in the future, it is then necessary to simulate the future health status of the population: for this purpose, we start from a baseline population (observed in 2008) that we make randomly evolve. To assess the health trajectory of a person, we need to know the probability of occurrence of the different health statuses.

In our model, health can take 3 "values": good health (1), bad health (2) or death (3). We write p_{ij} the probability to jump from a health state i to a state j between two dates $t_1$and t_2. We seek to determine the values in the following matrix by age and gender:

$$\begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix}$$

By definition, there could be no remission from death, which implies that $p_{31} = p_{32} = 0$ and $p_{33} = 1$. Besides, each row of the matrix corresponds to a starting health status and we consider all the future health statuses that can occur: hence, $p_{11} + p_{12} + p_{13} = 1$ and $p_{21} + p_{22} + p_{23} = 1$. These matrixes are gender and age group specific: we construct 20 transition matrix to simulate the evolution of the health status of people in the baseline dataset.

Unlike paper on related subjects in the literature for France like Barnay et *al.* (2009), we allow for remissions from bad health, that is $p_{21} \neq 0$. The hypothesis $p_{21} = 0$ is often necessary to guarantee the existence of a unique solution to the system we need to solve. We try here, using the temporal dimension of ESPS data used in panel, to untighten this simplifying hypothesis. We seek to calculate 6 parameters: p_{11}p$_{11}$,p_{12} , and p_{13} on the one hand, and p$_{21}$, p$_{22}$, and p$_{23}$ on the other hand.

The baseline population (surveyed in 2002 or 2004) can be split into two groups: people in good health (N_1persons) and people in bad health (N_2persons). Among these two groups, some people will survive and stay in the survey sample ($s_1$and s_2), others will also survive but won't be asked again (a$_1$ and a$_2$), and others will die (m_1 and m_2). As we have data on the reasons of the attrition, we observe s_1, s$_2$, m_1, and m_2, and we can easily deduce a$_1$ and a_2.

Survivors (whenever they stay in the sample or not) will be either in good or bad health at the second period at which they should be theoretically asked: consequently we define a_{11}, a_{12}, a_{21}, a_{22}a$_{21}$, s_{11}, s_{12} s$_{11}$, s_{21}s$_{12}$, s$_{21}$(with $a_{ij}$the number of people who disappear without dying and s_{ij} the number of persons staying in the panel). Among these 8 variables, we only observe s_{ij}. Without additional information, we assume that the global evolution of the health of the survivors is the same, whenever they stay or not in the sample. This hypothesis implies that individual would have the same evolution of their health status if they are in the panel or in the attrition by age and gender. This can be formalized by the following equations:

$$\frac{a_{11}}{a_1} = \frac{s_{11}}{s_1}$$

and

$$\frac{a_{22}}{a_2} = \frac{s_{22}}{s_2}$$

The following equations result from the previous reasoning:

$$p_{11} = \frac{s_{11} + a_{11}}{N_1}$$

$$p_{12} = \frac{s_{12} + a_{12}}{N_1}$$

With $p_{13} = 1 - p_{11} - p_{12}$

$$p_{21} = \frac{s_{21} + a_{21}}{N_2}$$

$$p_{22} = \frac{s_{22} + a_{22}}{N_2}$$

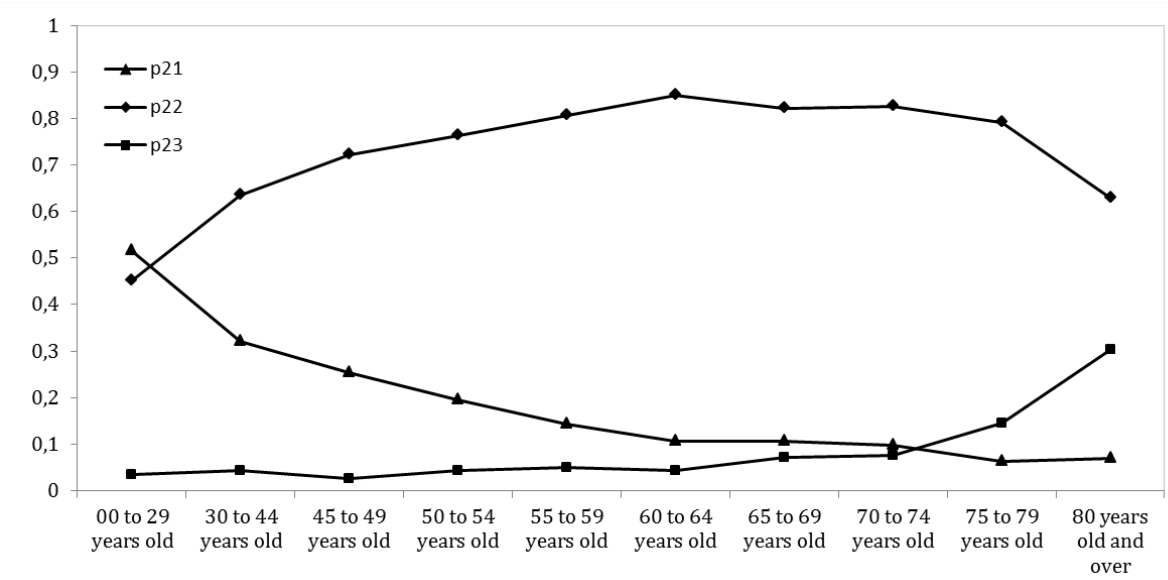
With $p_{23} = 1 - p_{21} - p_{22}$

Solving this system allows calculating transition probabilities by applying these following equations to our sample (Figure 4):

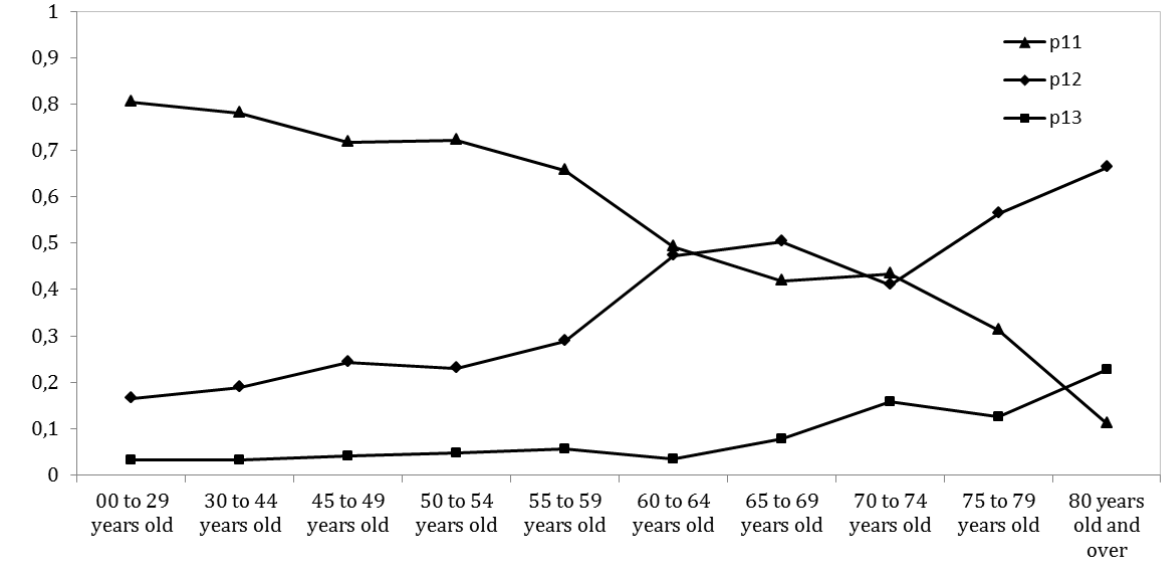
$$p_{11} = \frac{s_{11}}{N_1}\left(1 + \frac{a_1}{s_1}\right), \quad p_{12} = \frac{s_{12}}{N_1}\left(1 + \frac{a_1}{s_1}\right), \quad p_{13} = 1 - \frac{s_{11} + s_{12}}{N_1}\left(1 + \frac{a_1}{s_1}\right)$$

$$p_{21} = \frac{s_{21}}{N_1}\left(1 + \frac{a_2}{s_2}\right), \quad p_{22} = \frac{s_{22}}{N_1}\left(1 + \frac{a_2}{s_2}\right), \quad p_{23} = 1 - \frac{s_{21} + s_{22}}{N_1}\left(1 + \frac{a_2}{s_2}\right)$$

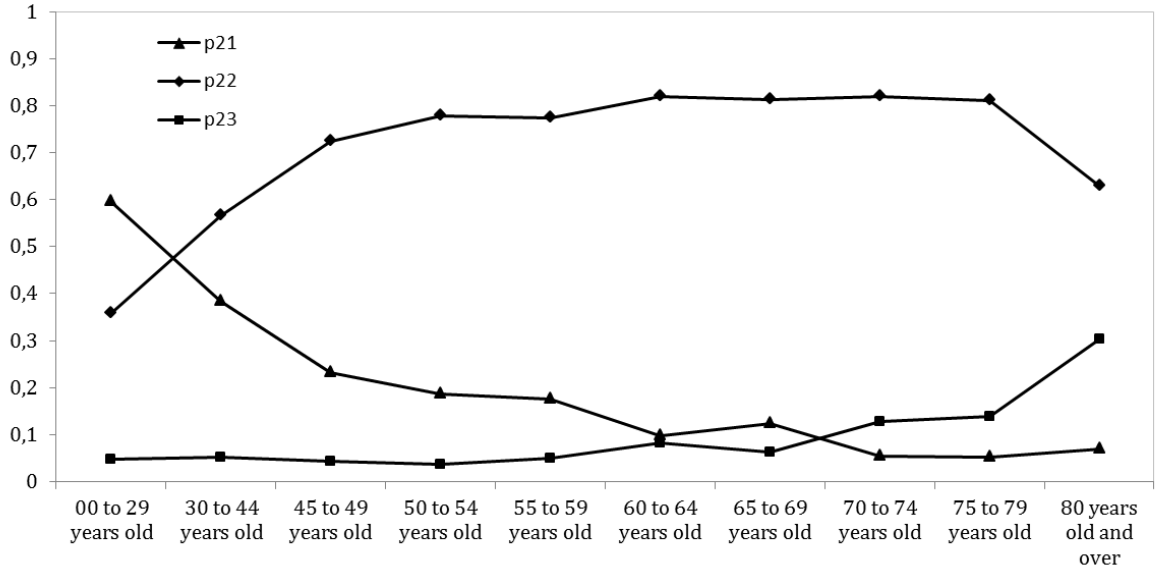
Figure 4: Transition probabilities from bad health to good health, bad health and death for women by age



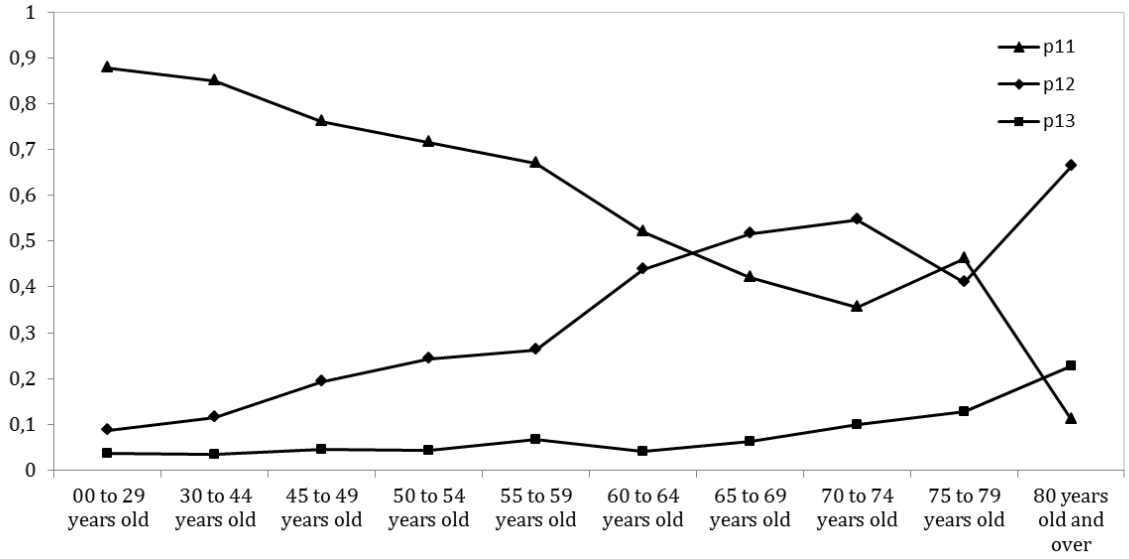
Transition probabilities from good health to good health, bad health and death for women by age



Transition probabilities from good health to good health, bad health and death for men by age



Transition probabilities from bad health to good health, bad health and death for men by age



Source: non balanced panel 2002-2006 and 2004-2008

Even if our baseline sample is representative of the French population regarding gender, age, social security affiliation and household size, nothing guarantees that it is exactly representative also regarding mortality rates. More precisely, the differences in mortality between people in good-health and in bad-health might be slightly different between ESPS-population and metropolitan French population. Hence, without adjustment, the probabilities estimated before could lead to simulate a different

population (in terms of numbers) from the demographic baseline scenario of the National Institute of Statistics. To avoid this, we adjust the death probabilities p_{13} and p_{23} before projecting the population, according to the following constraint:

$$r_{GH} * p_{13}^{adj} + r_{BH}\, p_{23}^{adj} = r_m^{INSEE}$$

where r_{GH}(respectively r_{BH}) stands for the rate of people in good (respectively bad) health in the population. We choose to do this adjustment maintaining the ratio of mortality rates between good and bad health observed in the sample, that is :

$$\frac{p_{13}{}^{adj}}{p_{23}{}^{adj}} = \frac{p_{13}{}^{obs}}{p_{23}{}^{obs}}$$

Finally, we have to make sure that :

$$p_{11}{}^{adj} + p_{12}{}^{adj} + p_{13}{}^{adj} = 1$$

and

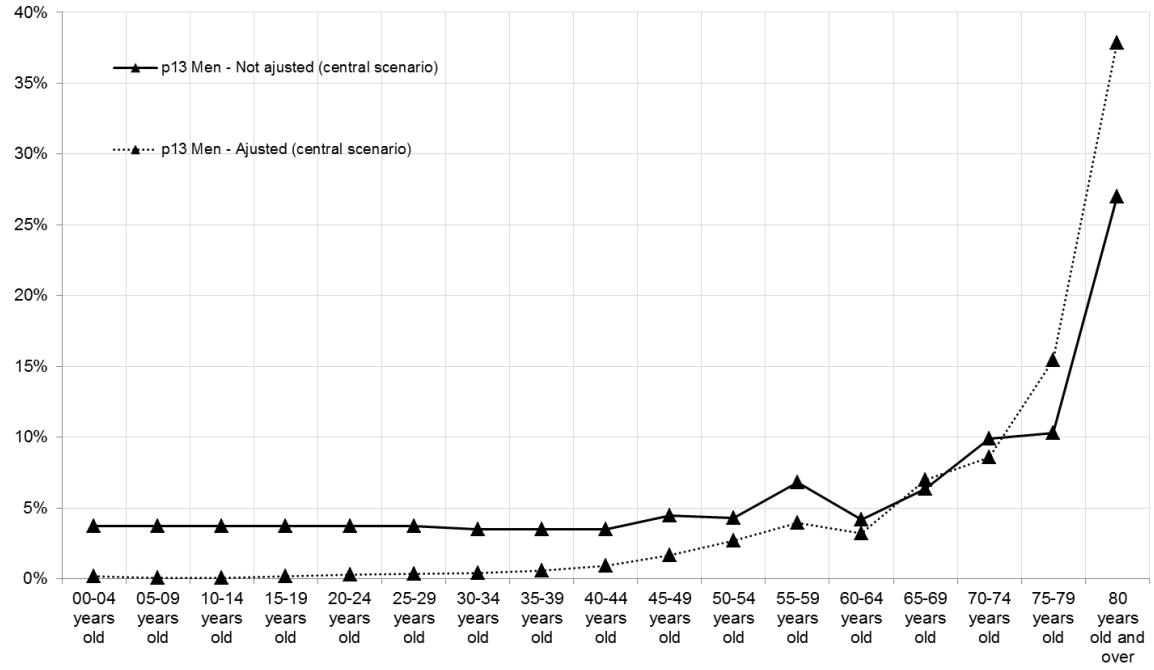
$$p_{21}{}^{adj} + p_{22}{}^{adj} + p_{23}{}^{adj} = 1$$

Thus, we adjust the probabilities without changing the odds ratios from a given health status. More precisely, that means, for $i = 1\ or\ 2$ that $p_{i1}^{adj}$and p_{i2}^{adj} are solutions to the following system of equations (Figure 5 for p_{13}):

$$p_{11}{}^{adj} + p_{12}{}^{adj} = 1 - p_{13}{}^{adj}$$

$$\frac{p_{i1}{}^{adj}}{p_{i2}{}^{adj}} = \frac{p_{i1}{}^{obs}}{p_{i2}{}^{obs}}$$

Figure 5: Adjusted and not adjusted probability of death for men in good health in function of age



Source: non balanced panel 2002-2006 and 2004-2008

3. ESTIMATION OF OUTPATIENT HEALTHCARE DEMAND

After a survey on different estimation techniques and notably through Jones (2000), we have chosen to estimate a two-part model. The results of the first stage regression (Probit) are presented in table 8 whereas those of the second stage (GLM) are presented in table 9. The aim of these regressions is rather to simulate expenditure for a projected population than to interpret and analyze the different coefficients of the demand for healthcare.

The distribution of health expenditure has particular characteristics. Besides the fact that healthcare expenditure is either positive or equal to zero, there are also a large number of null values. Indeed, a significant part of the population does not undertake healthcare spending some years. Meanwhile, the distribution of this expenditure is highly skewed because a small portion of the population has high healthcare expenditure.

As part of an econometric estimation, these features must be taken into account to estimate a model to explain the level of individual spending. Firstly, the expenditure distribution should be transformed in order to come closer the expenditure distribution to the normal distribution. This transformation should reduce the skewness (moment of order 3, the degree of asymmetry of a distribution) but also the kurtosis (moment of order 4, descriptor of the shape of a distribution). In order to reach these objectives, the log transformation is often used for estimations of healthcare spending. Moreover, this transformation allows interpreting the estimation results in terms of elasticity and estimating conventional utility functions, demand or cost functions such as the Cobb-Douglas function (Manning, 1998).

However, the necessary transformation of the results obtained from the logarithm of the expenditure in order to predict expenditure in the original scale of the observations is complex. A large literature has focused on the biases due to log-transformation of estimation (Duan, 1983, Duan et *al.*, 1983, Manning and Mullahy, 2001)[8]. Duan (1983)

[8] If $Y_i = X_i\beta + u_i$ with $Y_i = \log(y_i)$ then $y_i = \exp(Y_i) = \exp(X_i\beta + u_i) = \exp(X_i\beta)\exp(u_i)$. Once the results of the estimation are available, we predict the expenditure by the formula $\hat{E}(y_i) = E(\exp(X_i\hat{\beta})\hat{E}(\exp(u_i))$. We see directly that the exponential transformation is not straightforward.

has developed, under certain conditions, a robust method to reprocess the estimated expenditure in logarithm scale to the original data. However, this method has limitations that have led to use more often in the literature another method to estimate healthcare expenditure (Mullahy, 1998).

The Generalized Linear Model (GLM) allows taking into account certain characteristics of healthcare distribution as the asymmetry of the distribution and fat tailed distribution without having to use the transformation of the dependent variable in log (Manning et Mullahy, 2001). In the case of this model, the dependent variable can follow any law and the relationship between the dependent variable and the explanatory variables is not limited to the linear case like in the estimation by OLS. Since the relationship between the dependent variable and the explanatory variables is not specified, we have to choose a so-called "link function" that specifies this relationship. The "log link" is most often used in health economics because the healthcare expenditure is either positive or equal to zero. Furthermore, this link allows also reducing the skewness of the dependent variable.

Thus, we estimate the outpatient healthcare demand with X a set of explanatory variables including the socio-demographic characteristics and health status of individuals (age, gender, employment status, health status as dummy variable in order to have for each state of the nature, bad or good health, a link with the healthcare expenditure, professional status...) :

$$\ln[E(y|X)] = X'\beta$$

and :

$$E(y|X) = e^{(X'\beta)}$$

Estimated by :

$$\hat{\mathrm{E}}(y) = e^{(X\hat{\beta})}$$

The second problem related to the estimation of health expenditure is the fact that the distribution of healthcare spending has a large number of null values. To deal with this feature, different models have been studied in the literature on the estimation of healthcare demand (Jones, 2000). These models seek to characterize the relationship

between the decision to consume and the amount of consumption. Three possibilities can treat this aspect of the distribution of healthcare spending.

1. The choice to consume is independent of the level of consumption. In this case, we can use a model which estimates separately the decision to consume and the amount of consumption (two-part model).
2. The choice to consume and the amount of consumption relate to a single decision. In this case, the individual chooses the level of consumption including zero (Tobit model).
3. The two decisions are interdependent. The decision to use healthcare services and the choice of the amount of consumption can be explained by correlated variables. The link between these two decisions is implemented by controlling in the consumption equation by the result of the estimation of choice equation through the inclusion of the inverse Mills ratio in the regression (sample selection model).

The choice between these different models led to lots of debates in the literature (Manning et *al.*, 1987, Leung et *al.*, 1996). Manning et *al.* (1987) has shown the superiority of the two-part model compared to other models by comparing the mean-squared error of the two models (Two-part model and sample selection model). However, Leung et *al.* (1996) have refined the results of Manning et *al.* (1987) by demonstrating that the relevance of the sample selection model was based on the correlation between the inverse Mills ratio introduced in the consumption equation and the explanatory variables used. They show that in the case of a strong correlation between the inverse Mills ratio and the explanatory variables in the second step, the estimation of a two-part model is more robust than with a sample selection model[9]. In the case of our model, the explanatory variables used to estimate the outpatient healthcare expenditure are strongly correlated with the inverse Mills ratio. The correlation coefficient is 0.96. According to the criteria developed by Leung et *al.*

[9] In the case of a strong correlation between the Mills ratio and the second step variable, the mean square error is lower by estimating a model with a two part model than with a sample selection model. In fact, the estimator is not precise in the Sample Selection Model because of the colinearity of the Mills ratio. Then, the Two Part Model estimator even biased has a lower squared error.

(1996), it appears more robust to use in our estimation a two-part model. We use, in this case, the modified two-part model of Mullahy (1998), which assumes independence between the two equations of the model, and allows the use of a GLM to estimate second step. The link function is the log function and the distribution used to estimate the outpatient healthcare expenditure is a gamma function generally called up in the case of estimation of health expenditure (Dormont et *al.*, 2006, Paret, 2012).

We estimate the decision to use healthcare services by a Probit estimator with dichotomous variable coding participation. Φ is the distribution function of the standard normal distribution and $X_{1,i}$the independent variables. The decision to use is predicted by:

$$\hat{\mathrm{E}}(P_i \big| X_{1,i}) = \Phi(X'_{i,1}\hat{\alpha})$$

Then we estimate and simulate conditional expenditure using a GLM specification with D_i healthcare spending and $X_{2,i}$the explanatory variables used in the estimation. The amount of consumption is predicted by:

$$\hat{\mathrm{E}}(D_i \big| P_i = 1; X_{2,i}) = \exp(X'_{2,i}\hat{\beta})$$

We can hence observe the variables that influence the probability of spending for medical care but also the level of consumption for those whose expenditure is not zero. The coefficient related to the variables we project through the micro-simulation model will be used to evaluate the total outpatient healthcare expenditure in the future.

When controlling for health status, the age of the individual does not influence the probability of having non-zero expenditure. On the contrary, for those who consume, the older the patient is, the more he consumes. This indicates that at a given age, habits regarding healthcare consumption do not vary over time. However, the treatments necessary to cure illnesses become more and more complicated and hence costly with age, especially due to the appearance of multi-pathologies. This might explain why expenditure levels increase over the lifetime.

Being a woman increases both the probability to spend for health and the expenditure level for the consumers. This is consistent with recent evidence (Dormont, et *al.*, 2006). It can obviously be explained by higher expenses for physicians, more precisely specialists (gynecologists-obstetricians), but also for biological analyses.

We control for household income to take into account of the budget constraint of the household. However, due to the weakness of the reporting of the income in the survey, results must be taken with great caution. Besides, the sign of the coefficient varies along the income distribution, suggesting non-linear effects, quite complicated to interpret here.

Pensioners seem to be more likely to have expenditure than almost all the other individuals. The choice between leisure time and working time might be an explanation, justifying the difference with active workers in particular. Moreover, in case of consumption, retirees do not have significantly different expenditure than the other statuses, except active workers whose expenditure level is significantly lower. This seems to confirm that the main gap lies between active workers and other, all other things being equals.

Finally, being in good health (according to our health variable) logically has a negative impact on the likelihood to have non-zero expenditure, and also on the level of consumption when relevant. The health status variable represents 16.5 % of the total share of the variance of the outpatient healthcare expenditure explained by the explanatory variables.

Table 8: First stage (Probit) – Choice equation, Total outpatient expenditure

Choice equation	Coefficient		Std. Dev.	z	P>|z|	[95% Conf. Interval]	
Age (reference: 00-29)							
30-39	-.0347751		.0483633	-0.72	0.472	-.1295656	.0600153
40-49	-.110407	**	.0488393	-2.26	0.024	-.2061304	-.0146837
50-59	-.068213		.0519959	-1.31	0.190	-.1701231	.0336971
60-69	-.0652367		.0780632	-0.84	0.403	-.2182377	.0877643
70-79	-.0660253		.0921897	-0.72	0.474	-.2467138	.1146632
80 and over	.1541115		.1128582	1.37	0.172	-.0670865	.3753095
Health (reference: bad health)							
Good health	-.3637033	***	.0328998	11.05	<0.001	-.4281857	-.2992209

Gender (reference: man)							
Woman	.1874471	***	.0275426	6.81	<0.001	.1334645	.2414297
Wage (reference: refusal or unknown)							
€0-€599	-.0439834		.0969733	-0.45	0.650	-.2340476	.1460809
€600-€699	-.1848519	*	.1003902	-1.84	0.066	-.3816131	.0119093
€700-€899	.1263919		.0966129	1.31	0.191	-.0629659	.3157497
€900-€1099	.1207249		.0828063	1.46	0.145	-.0415726	.2830224
€1100-€1299	-.0928768		.070856	-1.31	0.190	-.231752	.0459984
€1300-€1499	-.077583		.0557185	-1.39	0.164	-.1867893	.0316233
€1500-€1799	-.1581286	***	.0525908	-3.01	0.003	-.2612046	-.0550526
€1800-€2199	-.0986027	*	.0535123	-1.84	0.065	-.2034849	.0062796
€2200-€2499	-.136558	**	.0533682	-2.56	0.011	-.2411577	-.0319583
€2500-€2999	.0569113		.0554112	1.03	0.304	-.0516925	.1655152
€3000-€3499	-.0136996		.0498075	-0.28	0.783	-.1113205	.0839212
€3500-€4499	.0349926		.0506094	0.69	0.489	-.0642001	.1341853
€4500 and over	-.029326		.0602973	-0.49	0.627	-.1475065	.0888546
Education (reference: primary education)							
Secondary education	-.0926047	**	.0389079	-2.38	0.017	-.1688628	-.0163466
Tertiary education	-.2781515	***	.0396807	-7.01	<0.001	-.3559243	-.2003787
No or other diploma	.113516	***	.0350699	3.24	<0.001	.0447803	.1822517
Professional status (reference: public sector employee)							
Self-employed worker	.2342228	***	.0783512	2.99	0.003	.0806573	.3877883
Other	.9632735	***	.3249845	2.96	0.003	.3263156	1.600231
Private sector employee	.258999	***	.0653636	3.96	0.000	.1308886	.3871093
Occupation (reference: pensioner)							
Active worker	-.1608264	**	.0715996	-2.25	0.025	-.3011591	-.0204937
Other inactive person	-.3839828	***	.0758768	-5.06	0.000	-.5326985	-.235267
Student	-.1053056		.1170997	-0.90	0.369	-.3348168	.1242057
Unemployed	-.2003672	**	.0840392	-2.38	0.017	-.365081	-.0356534
Other or unknown	-.1279079		.1287106	-0.99	0.320	-.380176	.1243603
Complementary insurance (reference: no coverage)							
Coverage	.0779961		.0475897	1.64	0.101	-.015278	.1712702
Professional status * Complementary insurance							
Private sector and coverage	.2929126	***	.0643796	4.55	<0.001	.1667309	.4190942

Household size (reference: 2 persons)							
1 person	-.0892456	**	.0422502	-2.11	0.035	-.1720543	-.0064368
3 persons	.0604518		.0402234	1.50	0.133	-.0183846	.1392882
4 persons	.0233754		.0372716	0.63	0.531	-.0496756	.0964264
Year (reference: 2008)							
2002	-.0323461		.03849	-0.84	0.401	-.1077852	.043093
2004	.0196163		.0376618	0.52	0.602	-.0541995	.0934321
2006	.0310534		.0374755	0.83	0.407	-.0423972	.1045041
PCS (reference: manager)							
Farmer	-.079257		.1043281	-0.76	0.447	-.2837364	.1252223
Craftsman	-.1794834	**	.0852817	-2.10	0.035	-.3466323	-.0123344
Other	-.8698046	***	.3199053	-2.72	0.007	-1.496807	-.2428018
Employee	-.0810581		.0575089	-1.41	0.159	-.1937735	.0316573
Worker	-.2649801	***	.0573054	-4.62	0.000	-.3772966	-.1526636
Intermediary	-.1375572	***	.052645	-2.61	0.009	-.2407395	-.0343749
Constant							
	1.655633	***	.1179519	14.04	<0.001	1.424451	1.886814

Note: the dependent variable is the probability to have non-zero outpatient expenditure

*** : 1% significant ; ** : 5% significant ; * : 10% significant.

Table 9: Second stage (GLM) – Consumption equation, Total outpatient expenditure

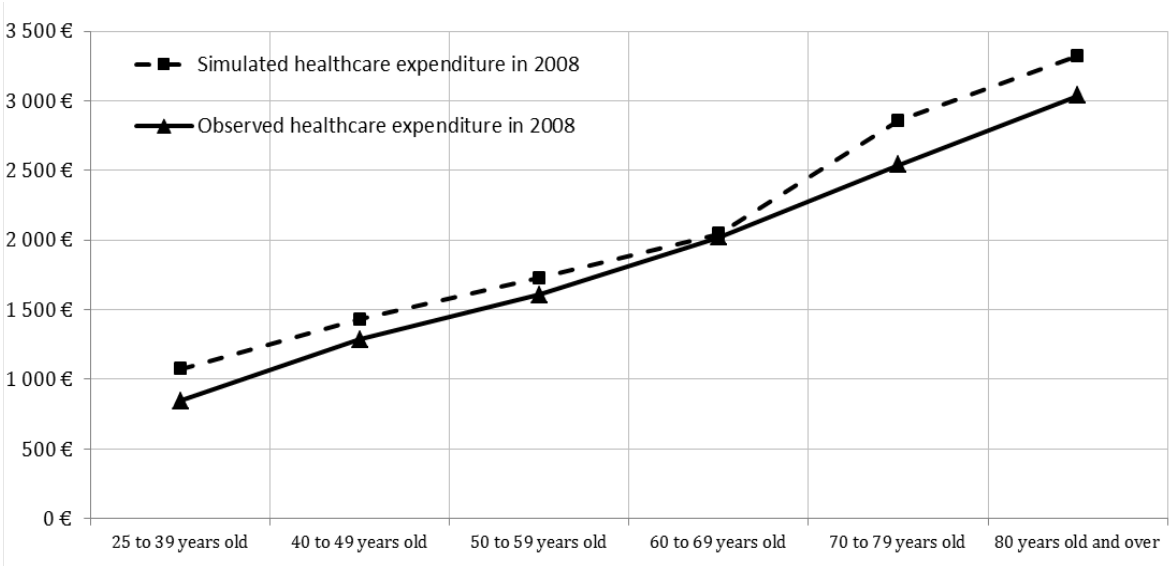
Consumption equation	Coefficient		Std. Dev.	t	>|t|	[95% Conf. Interval]	
Age (reference: 00-29)							
30-39	.0954453	***	.0423189	2.26	0.024	.0125019	.1783888
40-49	.2595395	***	.0429474	6.04	<0.001	.1753641	.3437148
50-59	.347535	***	.0447134	7.77	<0.001	.2598984	.4351717
60-69	.3440598	***	.064297	5.35	<0.001	.21804	.4700796
70-79	.540252	***	.0737813	7.32	<0.001	.3956433	.6848606
80 et plus	.7199338	***	.0795991	9.04	<0.001	.5639224	.8759451
Health (reference: bad health)							
Good health	-.7151044	***	.0249398	-28.67	<0.001	-.7639856	-.6662232
Gender (reference: man)							
Woman	.1096823	***	.0221419	4.95	<0.001	.0662849	.1530797
Wage (reference: refusal or unknown)							
€0-€599	-.0337139		.0854249	-0.39	0.693	-.2011436	.1337159
€600-€699	-.2748769	***	.090163	-3.05	0.002	-.4515931	-.0981607
€700-€899	-.0609701		.0707473	-0.86	0.389	-.1996323	.0776922
€900-€1099	-.1019647	*	.0598574	-1.70	0.088	-.2192831	.0153537
€1100-€1299	-.0916497		.0572077	-1.60	0.109	-.2037748	.0204754
€1300-€1499	-.0908277	***	.0443266	-2.05	0.040	-.1777062	-.0039493
€1500-€1799	-.1774171	***	.0434382	-4.08	<0.001	-.2625545	-.0922798
€1800-€2199	-.1795771	***	.0436273	-4.12	<0.001	-.2650851	-.0940691
€2200-€2499	-.1832092	***	.044184	-4.15	<0.001	-.2698083	-.0966101
€2500-€2999	-.1570666	***	.0427737	-3.67	<0.001	-.2409015	-.0732317
€3000-€3499	-.1424902	***	.0400617	-3.56	<0.001	-.2210096	-.0639708
€3500-€4499	-.0906337	***	.0405265	-2.24	0.025	-.1700642	-.0112032
€4500 and over	-.08534	*	.050469	-1.69	0.091	-.1842574	.0135775
Education (reference: primary education)							
Secondary education	.0191639		.0322685	0.59	0.553	-.0440813	.0824091
Tertiary education	-.0198829		.0343679	-0.58	0.563	-.0872428	.047477
No or other diploma	.02702		.0272283	0.99	0.321	-.0263466	.0803865
Professional status (reference: public sector employee)							
Self-employed worker	.039171		.0701611	0.56	0.577	-.0983422	.1766843
Other	-.1799966		.3908155	-0.46	0.645	-.9459808	.5859877
Private sector employee	.0824089		.0612046	1.35	0.178	-.0375499	.2023677
Occupation (reference: pensioner)							
Active worker	-.1771957	***	.0568373	-3.12	0.002	-.2885947	-.0657966
Other inactive person	.2590018	***	.0601249	4.31	<0.001	.1411592	.3768445
Student	-.0526899		.0980705	-0.54	0.591	-.2449046	.1395247
Unemployed	-.0546207		.0686245	-0.80	0.426	-.1891222	.0798808
Other or unknown	-.0369451	**	.1074715	-0.34	0.731	-.2475854	.1736952
Complementary insurance (reference: no coverage)							
Coverage	.1684268	***	.0431507	3.90	<0.001	.0838529	.2530007
Professional status * Complementary insurance							
Private sector and coverage	-.1030843	*	.0582497	-1.77	0.077	-.2172516	.011083

Household size (reference: 2 persons)							
1	-.0361553		.0337811	-1.07	0.284	-.102365	.0300543
2	-.0893454	***	.0320414	-2.79	0.005	-.1521454	-.0265454
4	-.1189804	***	.0307139	-3.87	0.000	-.1791786	-.0587822
Year (reference: 2008)							
2002	-.1972794	***	.0312612	-6.31	0.000	-.2585502	-.1360085
2004	.0058813		.0299298	0.20	0.844	-.05278	.0645425
2006	.0159457		.0299181	0.53	0.594	-.0426926	.074584
PCS (reference: manager))							
Farmer	-.2174149	***	.0831692	-2.61	0.009	-.3804235	-.0544064
Craftsman	.0109468		.0731683	0.15	0.881	-.1324605	.1543542
Other	.0192476		.3881667	0.05	0.960	-.7415452	.7800403
Employee	-.0432039		.0458225	-0.94	0.346	-.1330143	.0466066
Worker	-.0445687		.0464573	-0.96	0.337	-.1356233	.0464859
Intermediary	-.0491909		.0434354	-1.13	0.257	-.1343227	.0359409
Intercept							
	7.400652	***	.0984454	75.18	<0.001	7.207703	7.593601

*** : 1% significant ; ** : 5% significant ; * : 10% significant.

Thanks to the results of the estimation, we can compare the simulated outpatient healthcare expenditure in 2008 and the observed expenditure in the Social Protection and Health survey in 2008 (figure 6). The simulated expenditure reproduces quite well the shape of the observed expenditure. There is a slight overestimation of average outpatient healthcare expenditure notably for older ages (over 70 years old).

Figure 6: Average outpatient healthcare expenditure by age simulated and observed in 2008



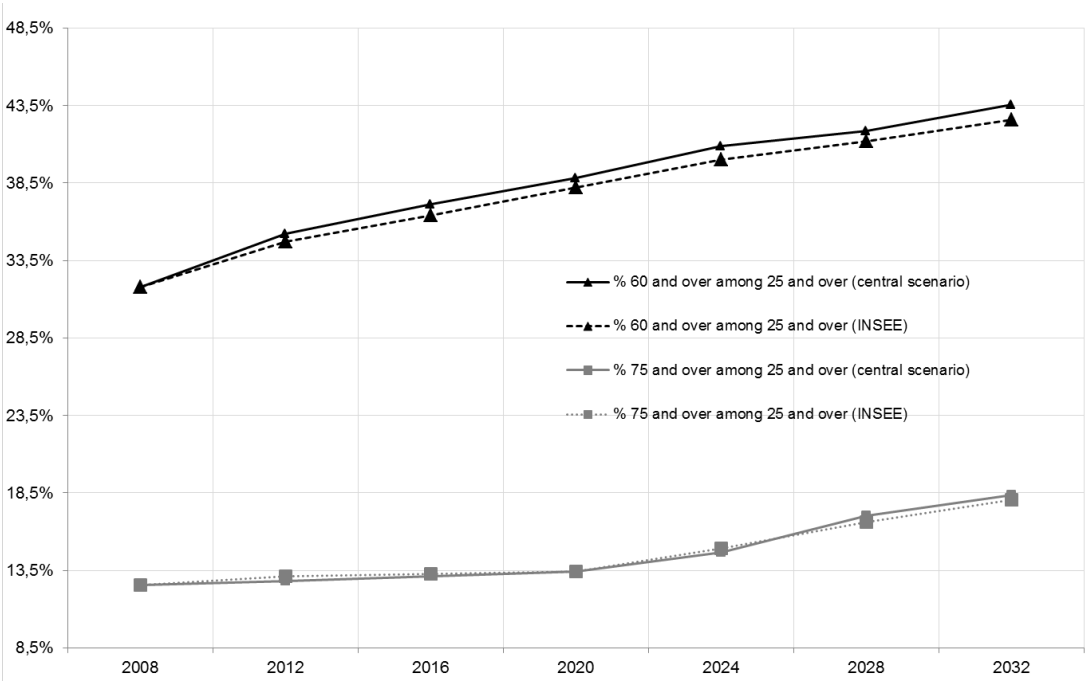
Source : ESPS 2008 and simulation model

4. RESULTS

4.1. Population projections to 2032 horizon

At first we perform a demographic and epidemiological projection of the population aged 25 and over in the 2032 horizon. Under the baseline scenario, we use the probabilities observed in our panel and adjusted to the mortality rates of the National Institutes of Statistics for the Markow process. At the 2032 horizon, the share of 60 years and over in the population is expected to represent approximately 43.5% of 25 years and older against about 31% in 2012 (Figure 7). The share of those aged 75 and over should stagnate until 2020 before experiencing a relatively high growth to reach over 18.5% of 25 years against 13% in 2012.

Figure 7: Population projections in the baseline scenario



Source: ESPS 2008 and micro-simulation model
Population : People of 25 years old and older covered by public health insurance in metropolitan France

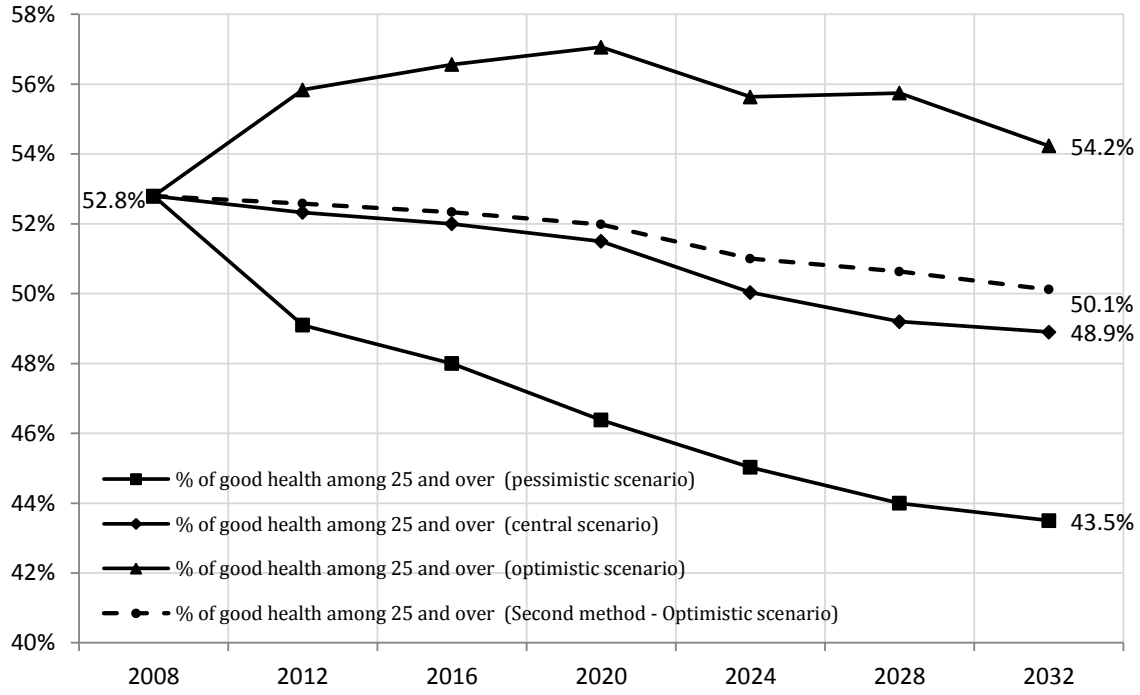
Two alternative scenarios are simulated in order to take into account a possible aging population process less or more favorable in terms of health status than it is tendentiously. To simulate such scenarios two types of hypothesis can be simulated. First in accordance with what is done by Thiébaut et *al.* (2010), the transition probabilities are modified by shifting the probabilities of transitions from one age

group. Thus in the case of pessimistic scenario (EM scenario), the transition probabilities of an individual who is between 60 and 65 years old are those of a person between 65 and 70 years old. Conversely, in the case of an optimistic evolution of the health status of the population, the transition probabilities for an individual who is between 70 and 75 years old are the transition probabilities of an individual who is between 65 and 70 years old. Another way to take into account a possible ageing in good health, is to correlate the shift of the transition probabilities in line with the gain in life expectancy. For instance, if the gain in life expectancy in four years is equal to one year, the transition probabilities for an individual who is between 70 and 75 years old is equal to one fourth of the transition probabilities of an individual who is between 65 and 70 years old and three fourths the transition probabilities of an individual who is between 70 and 75 years old. We call this way to simulate ageing "second method" in the different graphics presented. The first way to simulate ageing is equivalent to simulate a shock in health status in the population. The second way is equivalent to a continuous change in health status in correlation with the increase of life expectancy.

The direct consequence of these assumptions is reflected in various changes of the population in good health over the projection period. Thus, in the optimistic scenario, the share of people in good health reaches 54.2% of the population in 2032 against 52.8% in 2012 (Figure 8)[10]. This share represents only 43.5% of the population in the pessimistic scenario. By shifting the probability of an age group in the pessimistic scenario, the phenomenon of falling in poor-health will be accelerated and the chances of recovery are also reduced. With the second method to simulate the optimistic scenario, the difference with the central scenario is thinner. Indeed, in this scenario the share of people in good health will represent 50.1% of the 25 years old and over against 48.9% in the central scenario.

[10] The share is decreasing in the central scenario because of aging population.

Figure 8: Percentage of people in good health among 25 years old and over population



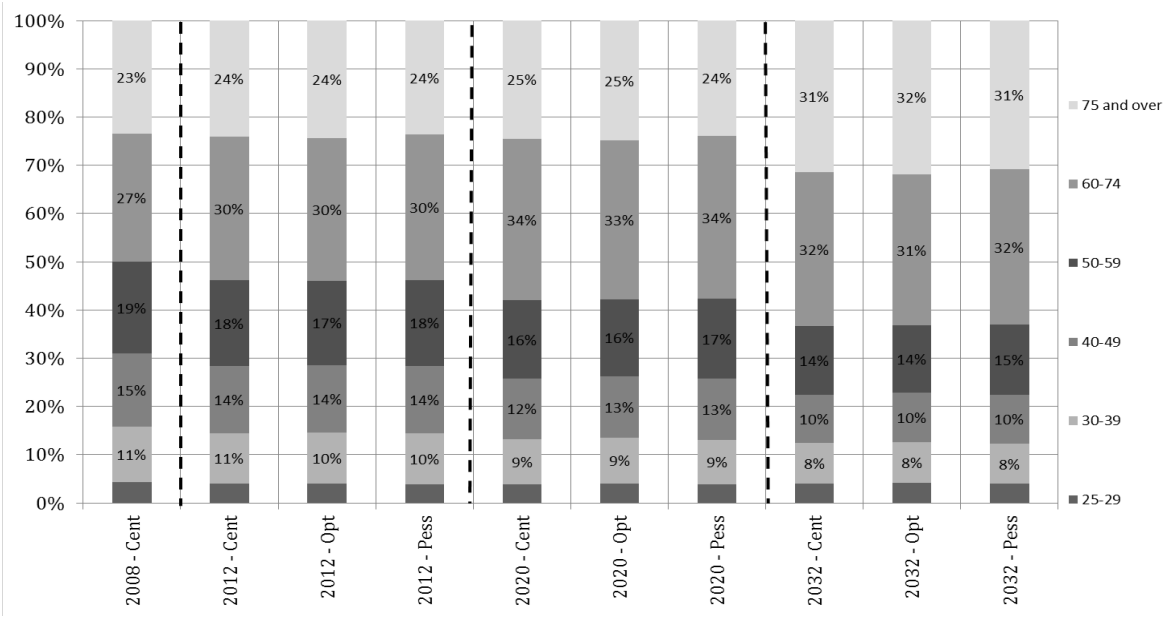
Source: ESPS 2008 and micro-simulation model
Population: People of 25 years old and older covered by public health insurance in metropolitan France

4.2. Financial projections to 2032 horizon

Based on the findings of the demographic projections, two issues related to the evolution of health care costs arise. Firstly, what is the evolution of health care spending of the 25 years old and over insured in the central scenario that is to say, if no epidemiological evolution occurs? Then what is the impact of the divergent evolution of the health status for both optimistic and pessimistic scenarios? The aging of the population contributes to distort the structure of the outpatient healthcare spending in 2032 (Figure 9). While in 2008, 60 years old people and over accounted for about 50% of the cost of care of the population of 25 years old and over, in 2032 this category of the population should concentrate 63% of this expenditure. We can also conclude that this share is quite equivalent between the different scenario. This phenomenon is due to the fact that people aged 60 years old and over will be healthier in the optimistic scenario which tends to reduce their spending. However, we saw that the healthcare expenditure increase with age. This phenomenon combined with the number effect linked to longer

life explains that the share of spending of people of 60 years old and over becomes important even in the case of the evolution of the health status optimistic scenario.

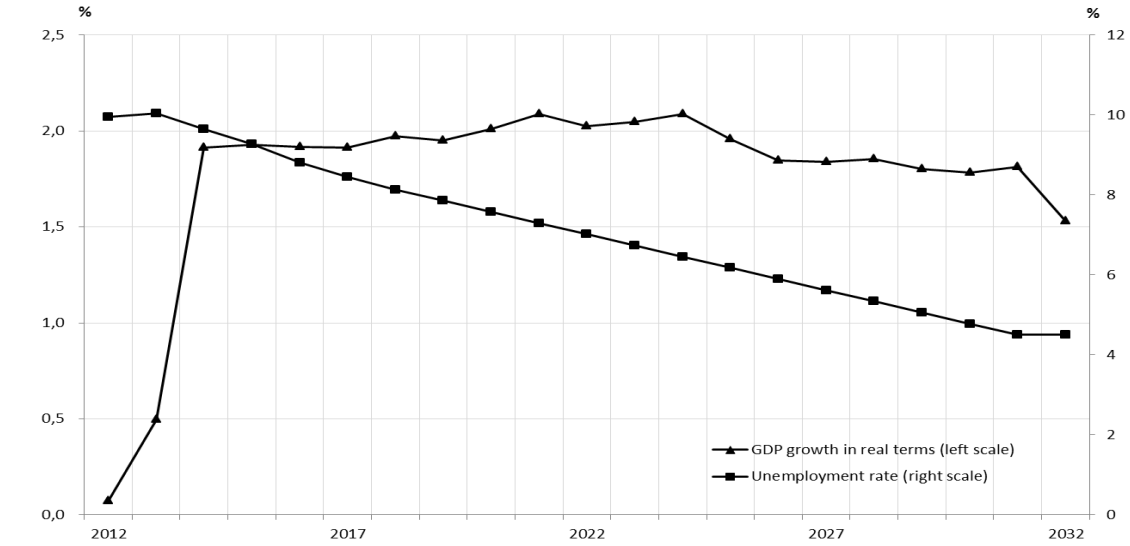
Figure 9: Structure of outpatient healthcare expenditure by age and the evolution of health by year



Source: ESPS 2008 and micro-simulation model
Population: People of 25 years old and older covered by public health insurance in metropolitan France

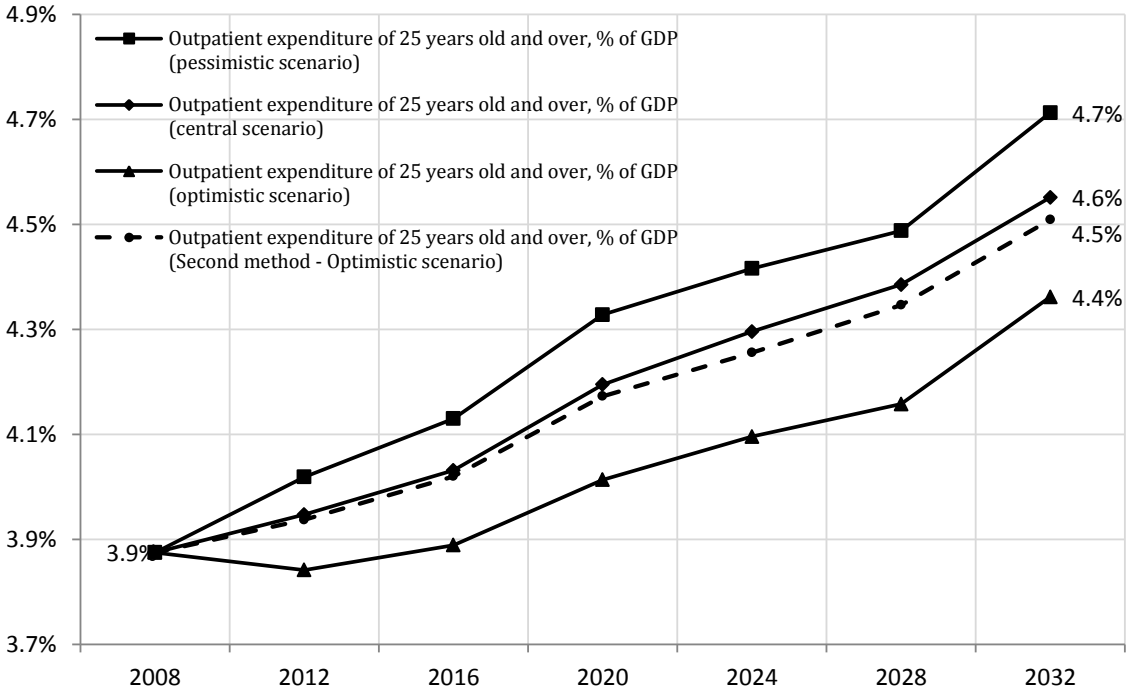
To determine the evolution of spending of the 25 years old and over in the GDP due to aging, we calculate the ratio of outpatient healthcare expenditure on GDP. The chosen scenario for macroeconomic evolution is the central scenario of the Retirement Guidance Council in order to keep coherent assumptions between the different projection exercises of social benefits (Figure 10). While healthcare spending of 25 years old and older represent 3.9% of GDP in 2008, they would reach 4.6% in the baseline scenario in 2032 (Figure 11). A difference in the share of expenditure in GDP appears between scenarios during the projection period. For example, outpatient healthcare spending represents 4.6% of GDP in the central scenario in 2032, against 4.4 % in the optimistic scenario and 4.7% in the pessimistic scenario.

Figure 10: Macroeconomic assumptions for the projections, GDP growth and unemployment rate



Source: Retirement Guidance Council

Figure 11: Outpatient expenditure of 25 years old and over in the four scenarii until 2032



Source: ESPS 2008 and micro-simulation model
Population: People of 25 years old and older covered by public health insurance in metropolitan France

CONCLUSION

Population ageing will be a major challenge in Europe in the coming decades. This phenomenon will raise the question of the sustainability of public spending with increasing healthcare provision costs. In this paper, we present a dynamic micro-simulation model for outpatient healthcare expenditure in France, who projects healthcare costs in the long run. Like all the dynamic micro-simulation models, the method used adopts a life-cycle perspective and projects the population structure over time. The projections are run using a transition process between three states: two non-absorbing (good-health or ill-health) and one absorbing state (death). The outpatient healthcare expenditure are estimated through a two-part model.

While healthcare spending 25 years old and older represents 3.9% of GDP in 2008, it would reach 4.6% in the baseline scenario in 2032. A difference in the share of expenditure in GDP appears between scenarios during the projection period. For example, outpatient healthcare spending represents 4.6% of GDP in the central scenario in 2032, against 4.4 % in the optimistic scenario and 4.7% in the pessimistic scenario. Furthermore, due to aging, outpatient healthcare expenditure of people who are 60 years old and over will account for 63% of the cost of care of the population of 25 years old and over against 50% in 2008.

Even if it is a special purpose micro-simulation model, it should be merged with DESTINIE 2, the micro-simulation model of the National Institute of Statistics and Economic Studies (Blanchet et *al.*, 2011). Thanks to this merge, we should complete the demographic module by simulating birth and professional pathways thanks to the implementation of the model built in DESTINIE 2. The implementation of this module in DESTINIE 2 will also allow taking into account more variables in the simulation of healthcare expenditure but also in the calculation of the transition probabilities. We would be able notably to calculate transition probabilities in function of age, gender but also the level of education thanks to logit estimations. Furthermore, thanks to new data that should be released, we should be able to introduce inpatient care in the model and consequently to reproduce the all healthcare expenditure. Then, the epidemiologic module could be enhanced by dividing the health status in more than three modalities. Furthermore, a module could be developed in which the evolution of risk factors is

linked with the change in health status in the population or with the amount of healthcare consumption. Future research could focus also on how technological progress could be introduced in the model or how the expenditure module could be divided in function of type of care.

REFERENCES

AHCPR (1997), « Trends in Personal Health Care Expenditure, Health Insurance and Payment Sources, community based population, 1987–1995 ». Agency for Health Care Policy and Research, Center for Cost and Financing Studies, National Medical Expenditure Survey data, March 1997, Table 8.

Astolfi R., Lorenzoni L. et Oderkirk J. (2012), « A comparative analysis of health forecasting methods », *OECD health working paper*, n°59.

Bago d'Uva T., Van Doorslaer E., Lindeboom M., et O'Donnell O., (2008), "Does reporting heterogeneity bias the measurement of health disparities?", *Health Economics*, John Wiley & Sons, Ltd., vol. 17(3), pp. 351-375.

Barnay, T., Thiébaut, S. et Ventelou, B. (2009), « L'effet du vieillissement et de l'évolution de la morbidité sur les dépenses de médicaments remboursables en ville. Une micro simulation quinquennale (2004-2029) », *Economie Publique/Public Economics*, pp. 24-25.

Blanchet D., Debrand T., (2007), « Souhaiter prendre sa retraite le plus tôt possible : santé, satisfaction au travail et facteurs monétaires », *Economie et statistique*, N°403-404, pp. 39-62.

Blanchet D., Buffeteau S., Crenner E. et Le Minez S. (2011), «Le modèle de micro-simulation Destinie 2 : principales caractéristiques et premiers résultats », *Économie et Statistique*, n° 441–442.

Blanpain N. et Chardon O. (2010), « Projections de population à l'horizon 2060 », *Insee Première*, n°1320.

Bound J., Schoenbaum M., Stinebrickner T.R. et Waidmann, T. (1999), "The dynamic effects of health on the labor force transitions of older workers," *Labour Econ.*, vol. 6, pp. 179-202.

Breuil-Genier P., Flipo A. et Mahieu R. (1998), «Applications des méthodes de micro-simulation dans les domaines de la santé et de la dépendance », Rapport pour le Commissariat Général au Plan coordonné par Didier Blanchet.

Cambois E., Robine J.-M. et Sieurin A. (2011), « Les espérances de santé en France : une tendance récente moins favorable que dans le passé », *Document de travail INED n° 170*.

Com-Ruelle L., Lecomte T., Le Fur P., Mizrahi A., Mizrahi A. et Sermet C. (1997), « Un indicateur de gravité des maladies – Enquête sur la santé et les soins médicaux », *Documents de travail du CREDES n°1189*.

Devaux M., Jusot F., Sermet C., Tubeuf S., (2008), « Hétérogénéité sociale de déclaration de l'état de santé et mesure des inégalités de santé », *Revue française des Affaires Sociales*, n°1, pp. 29-47.

Disney R., Emmerson C. et Wakefield M. (2006), "Ill-health and retirement in Britain: a panel data-based analysis", *Journal of Health Economics*, vol. 25, pp. 621-649.

Duan N. (1983), "Smearing estimate: a nonparametric retransformation method", *Journal of the American Statistical Association*, vol. 78, pp. 605-610.

Duan N., Manning W.G., Morris C.N. et Newhouse J.P. (1983), "A comparison of alternative models for the demand for medical care," *Journal of Business & Economic Statistics*, vol. 1 (2), pp. 115-126.

Dormont B., Grignon M. et Huber H. (2006), « Health expenditure growth : reassessing the threat of ageing », *Health Economics*, vol. 15, n°9.

Dormont B. (2009), « Les dépenses de santé : une augmentation salutaire ?», Opuscules du CEPREMAP, CEPREMAP, n°15.

Dormont B. et Huber H. (2012), « Vieillissement de la population et croissance des dépenses de santé », Rapport pour l'Institut Montaigne, http://www.institut-montparnasse.fr/wp-content/files/Collection_recherches_n_2.pdf.

Duée M. et Rebillard C. (2004), « La dépendance des personnes âgées : une projection en 2040 », *Document de travail Insee*, INSEE.

EC (2012), The 2012 Ageing Report. Joint Report prepared by the European Commission (DGECFIN) and the Economic Policy Committee (AWG).

Etilé F., Milcent C., (2006), «Income-related reporting heterogeneity in self-assessed health: evidence from France», *Health Economics*, vol. 15, pp. 965-981.

Fries J. (1980), "Aging natural death and the compression of morbidity", *New England Journal of Medecine*, vol. 303.

Geay C., Koubi M., de Lagasnerie G. (2013), « Projections des dépenses de soins de ville construction d'un module pour DESTINIE », Document de travail de l'INSEE et du Trésor, INSEE, Direction Générale du Trésor.

Geoffard P.-Y. (2000), « Dépenses de santé: l'hypothèse d'aléa moral », *Economie et Prévision*, n°142.

Greene W. (2003), "A Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models," *Working Papers 03-19, New York University*, Leonard N. Stern School of Business, Department of Economics.

Gruenberg E.M. (1977), "The failures of success", *Milbank Memorial Foundation Quarterly, Health and Society*, vol. 55.

Hagan R., Jones A.M. et Rice N. (2009), "Health and Retirement in Europe." *Int. J. Environ. Res. Public Health,* vol. 6, n°10, pp. 2676-2695.

Hahn B. et Lefkowitz J. (1992), « Annual expenses and sources of payment for health care services », National Medical Expenditure Survey Research Findings 14. AHCPR Pub 93-0007, Public Health Service, Rockville, MD.

HCAAM (2011), Rapport annuel pour 2011, http://www.securite-sociale.fr/Rapport-annuel-du-HCAAM-pour-2011.
HCAAM (2011), Rapport annuel pour 2011, http://www.securite-sociale.fr/Rapport-annuel-du-HCAAM-pour-2011.
van Imhoff E. et Post W. (1998), « Micro-simulation methods for population projection », *Population: An English Selection*, vol. 10.

Jones A. M. (2000), "Health econometrics," Handbook of Health Economics, in: A. J. Culyer & J. P. Newhouse (ed.), Handbook of Health Economics, edition 1, volume 1, chapter 6, pages 265-344 Elsevier.

Jones A. M., Rice N. et Roberts J. (2010), "Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS", *Economic Modelling*, vol. 27 (4), pp. 866-880.

Leung S.F. et Yu S. (1996), "On the choice between sample selection and two-part models", *Journal of Econometrics*, vol. 72, pp. 197-229.

Maddala GS. (1985), "A survey of the literature on selectivity bias as it pertains to health care markets, " *Adv Health Econ Health Serv Res.*,vol. 6, pp. 3-26.

Marbot C. et Roy D. (2012), « Projections du coût de l'APA et des caractéristiques de ses bénéficiaires à l'horizon 2040 à l'aide du modèle Destinie », *Document de travail Insee G 2012/10*, INSEE.

Manning W.G., Duan N. et Rogers W.H. (1987), "Monte Carlo evidence on the choice between sample selection and two-part models" *Journal of econometrics, vol.* 35 (1), pp. 59-82.

Manning W.G., Mullahy J. (2001), "Estimating log models : to transform or not to transform ?" *Journal of Health Economics*, vol. 20, pp. 461-494.

Manton K.G. (1982), "Changing concepts of morbidity and mortality in the elderly population," *Milbank Memorial Fund Quarterly, Health and Society*, vol. 60, no. 2.

Mizrahi, A. and A. Mizrahi (1985), « Indicateurs de morbidité et facteurs sociodémographiques », CREDES. 657

Mizrahi An., Mizrahi Ar, Com-Ruelle L., Lecomte T., Le Fur P., et Sermet, C. (1997), « Un indicateur de gravité des maladies : enquête sur la santé et les soins médicaux, France 1991-92 », *Document de travail du CREDES n° 1189*.

Mullahy J. (1998), "Much ado about two: reconsidering retransformation and the two-part model in health econometrics", *Journal of health Economics*, vol. 17, pp. 247-281.

Mundlak Y. (1978), "On the pooling of time series and cross sectional data", *Econometrica*, vol. 46 (1).

van Praag B., Frijters P. et Ferrer-i-Carbonell (2002), "The anatomy of subjective well-being", *Journal of economic behaviour & organization*, vol. 1493, pp. 1-21.

Paret A.C. (2012), « Estimations des dépenses de santé à l'horizon 2028 : Modèle Linéaire Généralisé », Mémoire de Master Economics and Public Policy sous la direction de Lise Rochaix et Koen Jochmans.

Robine J-M. et Mormiche P. (1993), « L'espérance de vie sans incapacité augmente », *Insee Première*, n°281.

Rochaix L., Tubeuf S. et Perronnin M. (2006), « Construction d'un indicateur continu d'état de santé agrégeant risque vital et incapacité », *Question d'économie de la santé*, n°107, IRDES.

Spielauer M. (2007), "Dynamic micro-simulation of health care demand, health care finance and the economic impact of the health behaviours: survey and review", *International Journal of Micro-simulation*, vol. 1(1), pp. 35-53.

Thiébaut S., Barnay T., Ventelou B., (2013), "Aging, chronic conditions and the evolution of future drugs expenditures: A five-year micro simulation from 2004 to 2029", *Applied Economics*, n°45, 13, 1663-1672.

Yang Z., Norton E.C. et Stearns S.C. (2003), « The real reasons older people spend more », *Journal of Gerontology Science* , vol. 58B (1).

Wagstaff A., van Doorslaer E. et Paci P. (1991), "Equity in the finance and delivery of health care: some tentative cross-country comparisons", in: McGuire, A., Fenn, P., Mayhew, K. éds.., Providing Health Care. Oxford, 1991.

Willekens F. (2009), "Continuous-time micro-simulation in longitudinal analysis," in A. Zaidi, A. Harding et P. Williamson (dir.), New frontiers in micro-simulation modeling, Farnham, Ashgate, pp. 353-376.

Wooldridge J. M. (2002), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, vol. 20, pp. 39–54.

Zucchelli E., Harris M. et Zhao X. (2012), "Ill-health and transitions to part-time work and self-employment among older workers", *Health, Econometrics and Data Group (HEDG) Working Papers 12/04*, HEDG, c/o Department of Economics, University of York.

Zucchelli E., Jones A.M. et Rice N. (2012), "The evaluation of health policies through dynamic micro-simulation methods,"International *Journal of Micro-simulation, Interational Micro-simulation Association*, vol. 5(1), pp 2-20.

Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex).
Ce projet est distingué par le jury scientifique international désigné par l'Agence nationale de la recherche (ANR).
Il est financé dans le cadre des investissements d'avenir.

(ANR-11-LABX-0091, ANR-11-IDEX-0005-02)

www.sciencespo.fr/liepp

Directeurs de publication:
Etienne Wasmer & Cornelia Woll

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
+33(0)1.45.49.83.61
liepp@sciencespo.fr