ESTIMATING HEALTH EXPECTANCY IN PRESENCE OF MISSING DATA: AN APPLICATION USING HID SURVEY

Cristina Giudici *

Maria Felice Arezzo †

Nicolas Brouard ‡

October 8, 2012

Abstract

In this article we estimate health transition probabilities using longitudinal data collected in France for the survey on handicaps, disabilities and dependencies (HID) from 1999 to 2001. We examine the sample attrition of the survey, and reduce it through a model based imputation method. Life expectancies with and without activities of daily living (ADL) disabilities are calculated using a Markov-based multi state life table approach with two non-absorbing states: able to perform all ADLs and unable or in need of help to perform one or more ADLs, and the absorbing state of death. The loss of follow-up between the two waves induce biases in the probabilities estimates: mortality estimates are biased upwards; also the incidence

*Affiliation: Department of Methods and Models for Economics, Territory and Finance; Sapienza University of Rome; Via del Castro Laurenziano, 9 - 00161 ROME (email: cristina.giudici@uniroma1.it).

†Methods and Models for Economics, Territory and Finance; Sapienza University of Rome; Via del Castro Laurenziano, 9 - 00161 ROME (email: mariafelice.arezzo@uniroma1.it).

‡National Institute of Population Studies in Paris. Boulevard Davout, 133 - 75020 PARIS (email: brouard@ined.fr)

of recovery and the onset of disability seems to be biased. Since individuals were not missing completely at random, we decided to correct this bias by estimating health status for drop-outs using a non parametric model. After imputation, we found that at the age of 70 disability-free life expectancy decreases by 0.4 of a year, whereas the total life expectancy increases by 1 year. The slope of the stable prevalence increases, but it remains lower than the slope of the cross sectional prevalence. Globally there is no evidence of a general reduction in ADL, as defined in our study. The gender and relational differences on life expectancy did not change significantly after the imputation, but expected life free of disability decreases. The added value of the study is the reduction of the bias induced by sample attrition.

Classification JEL: I10, C14
keywords: Healthy life expectancy, Classification and regression trees, Sample attrition

1 Introduction

The debate on aging in Europe is currently paying considerable attention to the healthy life expectancy (HLE) of the elderly. Following the approach of the World Health Organization (WHO), health should be considered as having a dynamic nature[1], and should be taken into consideration in the context of life, as the ability to fulfill actions or to carry out a certain role in society. This is the so-called functional approach, taken by the WHO in the elaboration of the international frame of reference on the matter.

The most suitable indicator to measure the state of health of a population is health expectancy, which measures the length of life spent in different states of health. The term is often used in a general sense for all indicators of health expressed in terms of expectancy,

[1]Social, economic and environmental consequences of illness can be summarized in the sequence: illness or disorder - impairment or invalidity - disability - handicap. According to this sequence, handicap has its origins in a disease (including accidents or other causes of moral or physical traumas) which, as a consequence, causes problems in body functions or structures such as significant deviation or loss (impairment or invalidity). Invalidity constitutes in turn greater or lesser difficulty in performing daily activities (disability). Every dimension of handicap is effectively defined in relation to a norm: for example a disability consists in the reduction of the ability to carry out determined tasks in the way considered normal for a human being.

but the definition most frequently used in Europe is that of disability-free life expectancy (25), where disability is defined as the impact of disease or injury on the functioning of individuals. In other words, a disability is a restriction in the ability of accomplish tasks of daily living which someone of the same age is able to perform. According to (10; 32) disability is strictly dependent on the social and economic background someone belongs to.

The calculation of health expectancy is often based on on a method, pioneered by Wolfbein on the length of "working life" (33), described in details by (31), which combines prevalence of disability obtained through cross-sectional surveys and period life tables. Following this approach, the incidence of incapacity in the period of reference is not taken into account; the prevalence observed at a given moment derives from past health transition, and therefore depends on the history of the cohorts which make up the sample group. Age-specific cross-sectional prevalences are analogous to age-specific proportions of survivors from the corresponding cohorts (4; 15) in the sense that they are not subject to current mortality trends but delayed trends.

A possible alternative is the method of multi-state tables pioneered by Rogers (29) and Willekens for migration and marital status (34), Hoem (11) for the multi-state table of working life and Brouard for the introduction of the period prevalence of labor participation (5). Multi-state models are based on the analysis of the transitions between states in competition with the probabilities of dying from each state. The information necessary for this type of analysis derives from longitudinal surveys. The result, in our case, is the so-called period (or stable) prevalence and can be interpreted analogously to the stationary population of a period life table, as the proportion of the disabled amongst the survivors of successive fictitious cohorts, subject to the flows of entry on disability, recovery and death observed in the period under examination. Then the period health expectancy is the expectancy, in its statistical sense, of the distribution of the duration spent in the healthy state by this fictitious cohort. The analogy with the period life expectancy or simply "life expectancy' 'which is the expectancy of the distribution of deaths by age is obvious. The combination of a cross-sectional prevalence, instead of a period prevalence, with a period life table yields to a mixed Sullivan index which is often and improperly called health expectancy too. Such an health expectancy based on the so-called Sullivan method is not satisfactory in order to monitor the evolution of the current health conditions of a population and to forecast its future development. We believe that a cross-national comparaison within the European Union using the Sullivan index reflects more the history of each countries according to the wars than the real current situations which should be more homogeneous and in a similar way as the life expectancies are.

Computational issues on estimation of health expectancies from cross-longitudinal surveys have been developed by (12), (3) while (21) provided a complete solution with standard errors. The authors developed the embedded Markov chain maximum likelihood procedures pioneered by Laditka and Wolf (1995, (20)). They estimate parameterized transition probabilities following the Interpolation of Markov Chain approach (IMaCh). This approach has been recently applied in several analyses dealing with health (22; 9; 1), including studies based on the French HID survey (8). In these studies, information on health status is given by the interviews at different time, but loss of follow-up "within" successive waves can induce biases in the statistical results.

In our study we estimate the probability of transition between different states of health for the population of 70 years old or older in France, during the period 1999-2001, and following the Markov Chain approach. We based the analysis on the French HID survey, taking into account the loss of follow-up within the two survey waves, and imputing a health state for those who are lost through a non-parametric model named Classification and regression tree (CART).

Taking into account the heterogeneity of mortality due to health states, we compute life expectancy in different states of health and the period prevalence of disability implied by the estimated health transitions. We examine how health transitions are influenced by socio-demographic variables, in order to calculate differences in health expectancy linked to social and relational factors. The added value of the study is the reduction of the bias induced by the loss of follow-up within the two waves of the HID survey.

2 Data and methods

2.1 Data

Our study is based on the national survey on handicaps, disabilities and dependency (HID), carried out in France, between 1998 and 2001, by the National Institute on Statistics and Economical Studies (INSEE), in collaboration with the Institut National d'Etudes DÃľmographiques (INED). The survey began in 1998 with interviews of about 15,000 people living in institutional settings (and particularly in what are considered as "medico-social" institutions in France) and went on in 1999 with interviews of about 16,500 people living in ordinary settings. Most respondents were re-interviewed a second time, starting respectively in 2000 and 2001. The two samples were representative respectively of institutionalized and non-institutionalized population in France. Our study is limited to the population aged 55 and over at the baseline (16,964 interviews, 7,160 and 9,804 in istitutional and ordinary settings respectively).

Some characteristics of HID survey should be emphasized: as far as ordinary settings are concerned, in case of change of address within the two waves only people remained in ordinary settings have been followed: institutionalized people (52 individuals) have not been followed, and in our study they are supposed to be disabled. Moreover, some survey design changes occur within the two waves[2], but these changes are reflected in the weights, and sample weights are used in all our analysis. Analogously, as far as institutional settings are concerned, only people remained in the same istitution have been followed[3].

In addition, respondents to the first wave were followed through the Vital Statistics (*Etat Civil*), so the exact information on the possible date of death have been recorded. More in detail, after the cross check, a total of 3,198 deaths have been recorded from the 16,964 individuals aged 55 years and over interviewed at the baseline, 2,458 of which in istitutional settings and 740 in househods.

[2]985 people resident in the Department of HÃľrault have been excluded from the second wave; 462 individuals have been recorded as deceased; 4,844 did not participate in the first wave, whereas 23 people refused to participate in the second wave

[3]93 out of 100 of the initial sample

On the basis of the HID survey, health is measured through a functional approach: disability refers to the activities needed for independent living and personal care and has been operationalized as the difficulty or inability to perform one of the five activities of daily living (ADL): bathing, dressing, eating, getting in/out of a bed or chair and toileting. Three states are used in the analysis: 1- able to perform all ADLs, 2- unable or in need of help to perform one or more ADLs, and 3- deceased.

For every observation, disability status is known at the first wave. If people are interviewed only once and no other information is known, they will not be included in the estimate without imputation: missing data at the second wave are automatically dropped out by IMaCh. In order to reduce the bias due to the attrition, missing data for individual known to be alive in the second wave, but not interviewed, were assigned through CART as explained in detail in section 2.2.1. In five cases death was recorded by the interviewer, but no information was available by vital statistics. In this cases individuals have been coded as deceased.

2.2 Methods

2.2.1 Sample correction

Let $I(ADL2w)$ be an indicator function taking value 1 if ADL at the second wave is missing and 0 otherwise. First of all we studied the distribution of the drop-outs conditional on some covariates (i.e $f(I(ADL2w)|X_k)$). We found many differences, meaning that the drop-out mechanism was not random, but depended on the covariates values. We therefore decided to input the ADL at the second wave using a model which exploits the influence of the covariates X.

This simply means to build a model for ADL at the second wave using only the not dropped out individuals (7,179). This can be done in many different ways, for example using a logit or a probit model. We decided to use a non-parametric model for reasons that will be discussed at the end of the paragraph.

CART is a supervised classification algorithm. A supervised classification problem can be summarized as follows: for n objects, characterized by a set of k features $X = (X_1, X_2, ..., X_k)$,is known a priori the class $j = 1, 2, 3...J$ to which they belong Classes are generally indicated with variable Y. The scope is to predict which is the class a new object belong to, given its characteristics. A supervised classification algorithm is a mathematical rule which assign a new object to a class j. A function $d(X)$, called classifier, is built in a way that it generates a partition of the feature space X into J non overlapping subsets.

CART is a binary recursive partitioning procedures capable of handling both continuous and nominal characteristics. Starting with the entire sample (parent node), it divides it into two children codes; any of them are then divided into two grandchildren. A node is said to be final if it cannot be divided. The procedure stops when the tree reaches at its maximum size. The full grown tree is then pruned back in order to look for the best final tree. This is the one that minimize the so called cost-complexity function which is a function that takes into account at the same time the misclassification rate of individuals and the total number of final nodes.

The original data has a certain level of heterogeneity: if all individuals belong to the same class, there is no heterogeneity in the data. Conversely, if individuals are uniformly distributed among the J classes heterogeneity reaches its maximum level. Heterogeneity can be measured according to different method; one of the most common is the Gini index which is the one we used. Any split is done according to a variable Xi: the algorithm searches over all feature space looking for the optimal division that is for the binary split that reduces data heterogeneity most. Impurity reduction can be measured and it gives variables ranking based on their capability to separate objects. This is called variable importance.

An important issue is the capability of a tree to correctly classify a new individual. A measure of this generalization power is the misclassification rate which is simply the number of misclassified individuals out of all observed individuals. If the original sample is big enough, a good estimate of the true misclassification rate is obtained by randomly splitting the sample in two sub sample sand using the first part of the data (normally 70% of it) to grow the tree and the second to test it.

A very appealing aspect of CART is that it is able to handle missing values among independent variables through the mechanism of the surrogate variables. The missing data algorithm accomplishes two purposes at the same time: first it uses as much data as possible, complete or not, during the tree construction; second it classifies a new individual even if it has some variables values missing. If a case has a missing value, the algorithm proceed as follows. Suppose that the best split on a node has been found and that it is on variable X_b. Suppose that a new individual has to be classify and that it has a missing on X_b. Among all non-missing variables X in the case, it finds that one, say X_m, with the split having the highest measure of predictive association with the best split found in absence of missing data. In other words, the algorithm splits on the variable X_m which give the most similar classification to the best one X_b.

As we briefly mentioned, we used CART for two reasons: the first one is that it generally classifies more accurately than other models and the second is for the missing values management in covariates X. To confirm the first statement we tried several logistic models and found that the best rate of correct classification was 77.8% whereas for CART was 86%. Table 1 shows the variable importance in predicting the health status at the second wave: CART shows that ADL at the baseline is by far the most important variable. Once we estimated the model, we proceed to imputation for the 2,940 people whose health status was unknown (1,879 for menage group and 1,061 for institution). Table 2 shows the predictive ability of CART and tells how reliable the performed imputation is: results are good because the global error rate is about 19%. In order to provide a first indication of state changes in the study, table 3 shows the sample distribution by status in both waves, before and after imputation: most people began and ended disability-free; recovery percentages changes slightly after the imputation, whereas the percentage of those who remain disable increases.

2.2.2 Transition probabilities estimation method

We calculate the age-specific flows of entry into and exit from disability, and the matrix of the transition probabilities between good health (coded 1), disability (coded 2) and deceased (coded 3) employing the Interpolated Markov chain approach (IMaCh), developed by Brouard and LiÃÍvre, following work by Ladikta and Wolf (1998).

The probability for an individual aged x, observed in the state i during the first wave, to find him/herself in state j at the second wave is indicated by p^x_{ij}, and the transition

Table 1: Importance of independent variables

Independent variable	Importance	Normalized importance
ADL status (at first wave)	.199	100.0%
Age	.076	38.3%
Self-perceived health	.061	30.5%
Mental health	.060	30.1%
Sample (menage or istitution)	.040	19.8%
Activity	.021	10.4%

Table 2: CART misclassification rate on training and test samples

Sample	Observed value	Predicted value		
		Disability free	Disability	Correct percentage
Training	Disability free	5,277,782	1,048,902	83.42%
	Disability	470,800	1,497,614	76.08%
	Overall percentage	69.30%	30.70%	81.68%
Test	Disability free	2,260,641	449,453	83.42%
	Disability	201,536	642,467	76.12%
	Overall percentage	69.28%	30.72%	81.68%

Table 3: Distribution of people interviewed (menage and istitution) at the baseline by state at the beginning and end of the interval

	Menage		Institution	
	Before imputation	After imputation	Before imputation	After imputation
Disability-free at both intervals	3768	4686	733	886
Disability-free to disability	966	1280	599	684
Recovered from disability	658	663	162	162
Remained disabled	1787	2426	2730	2969
Died from disability-free	276	276	343	343
Died from disability	464	464	2115	2115
Missing from disability free	1232	0	238	0
Missing from disability	644	0	239	0
Information on health missing at the base line	9	9	1	1
Total missing	1879	0	477	0
Total	9804	9804	7160	7160

probabilities are estimated based on a series of 3x3 matrices:

$$p_{ij}^{x} = \begin{pmatrix} p_{11}^{x} & p_{12}^{x} & p_{13}^{x} \\ p_{21}^{x} & p_{22}^{x} & p_{23}^{x} \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

The first and the second row represent transitions for individuals who begin the interval respectively non disabled and ADL disabled. The third row represents the absorbing state of death. The probabilities of transition are then parameterized using the following logistic multinomial logit:

$$\ln \frac{p_{ij}^{x}}{p_{jj}^{x}} = \alpha_{jk} + \beta_{jk} x \qquad j \neq k \tag{2}$$

The software IMaCh is able to provide standard errors for the estimated parameters, which are then used to derive standard errors for the life expectancies implied in the transition probabilities. This is an important characteristic which allows for the assessment of whether results are statistically meaningful.

On the basis of transition probabilities estimates, IMaCh provides the so-called period (or stable) prevalence, which can be interpreted, analogously to the stationary population of a life table, as the proportion of the disabled amongst the survivors of successive fictitious cohorts, subject to the flows of entry on disability and recovery observed in the period under examination. In other words, the stable prevalence is implied in the health transitions observed during the survey, whereas the observed prevalence synthesize the history of disability onset, recovery and mortality of the population. Thus, the comparison between the stable and observed prevalence allows to make hypothesis on the future trend of health prevalence for cohorts under examination. (LiÃÍvre et al., 2003).

3 Results

Probabilities of transition For each age we calculate the probability of death within a year from each initial health status and compare the results with the 1998-2000 national age-specific mortality, as shown in figure 1. Total mortality rate is obtained by weighing each status-based probability of death with the proportion of people in each health status,

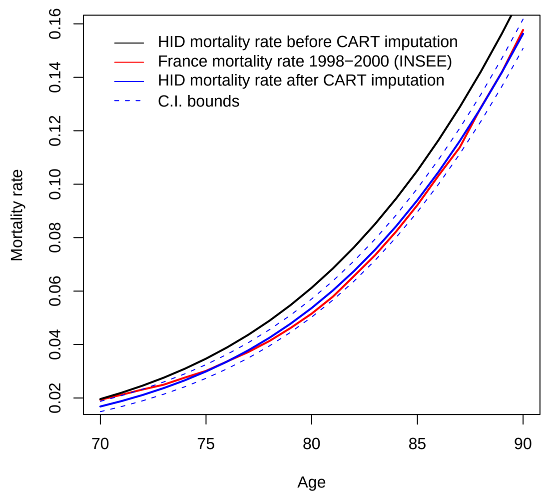

Figure 1: Death Rates by Age for Total Population with 95% confidence interval and comparison with annual national probability of death obtained from French vital statistics

given by the observed HID prevalence. Before CART imputation, mortality seems to be overestimated: the reason is that, since IMaCh automatically excludes individuals with missing ADL, the denominator of mortality rate is biased downward. The bias is corrected after the imputation. Figure 2 shows the transition probabilities from different initial state of health. As expected, the probability of dying is higher among the disabled. Regardless of the initial health state, the slope decreases after imputation, but the reduction is larger for those who were disabled at the baseline. The imputation modifies mainly the transition rates in older ages, except for recovery. In this case the intercept is reduced, and the slope did not change significantly.

Health Expectancies

As shown in figure 3, at all ages, our estimates of LE overlap those based on national statistics: at age 70, our estimates after CART correction are 15.21 years (95% CI [14,67-15.75]) compared to the 1998-2000 French life table of 15.17 years. Estimation before imputation was lower, due to the overestimation of mortality.

According to our model, people aged 70 can expect to live 9.37 years in disability-free

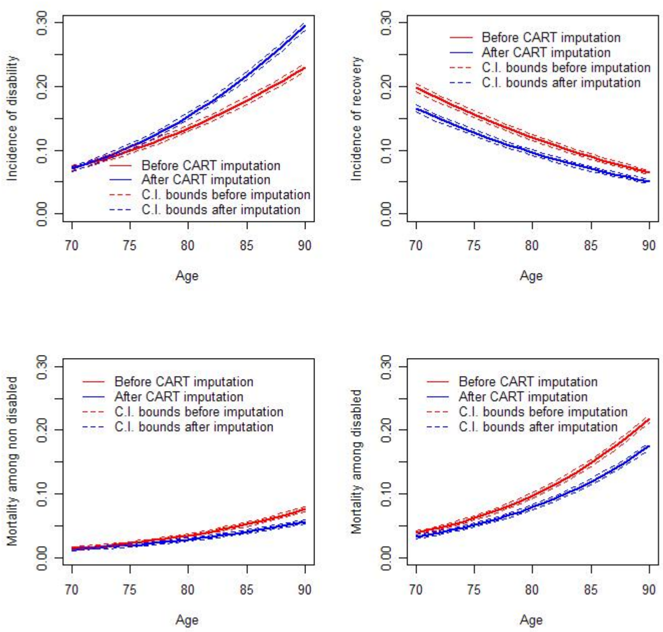

Figure 2: Transition Probabilities by Age for Disabled and Non disabled with 95% confidence interval

state, given that they were in that state initially, but the expectation is reduced to 5.53 years if they were in the disabled state at age 70. The corresponding health expectancies for the disabled state are 6.10 and 8.64 years respectively (figure 4).

Implied prevalence

The impact of continuing the rates of disability onset, recovery and death on ADL prevalence is shown in figures 4 and 5: as expected, the transition probabilities from both initial states (disability free and disabled) to a final state of disability at age x+h (and h=12 months), converge to the so called period, or stable prevalence of disability. The period prevalence is obtained by simulating cohorts aged 70 years and over which experience over time the observed transitions of health. As widely stressed in the literature, the comparison of the stable with the observed prevalence provides an indication on the evolution of age-specific prevalence of disability, if current transition rates of disability onset and recovery continue indefinitely (LiÃÍvre et al. 2003, Jagger et al. 2003, Laditka and Laditka, 2006, Manton and Land 2000, Minicuci et al. 2004, Reynolds, Saito and Crimmins 2005, Crimmins, Hayward, Hagedorn, Saito and Brouard, 2009).

Figure 4 compares the observed prevalence of disability (the broken line) with the stable prevalence (the straight line) before and after imputation. In the first case the two curves overlap, whereas in the second case differences between stable and cross-sectional prevalence become more significant at certain ages. Our imputation of a health state for lost individuals modifies the slope of the curves, but the effect on the stable prevalence is stronger than the effect on observed prevalence. Figure 5 focuses on results after the estimation of missing health status: the slope of the stable prevalence seems to be always lower than slope of the cross sectional prevalence, and globally there is no evidence of a general reduction in ADL.

Gender disparities

As shown by (12) The gender differences on expected life free of disability did not change significantly after imputation : figure 6 shows the transition probabilities for each sex from different initial states of health before and after imputation. Before imputation,the probability of death for disabled men at age 70 is close to that of women at age 78. But, if men are disability free, their probability of dying at 70 is close to that of

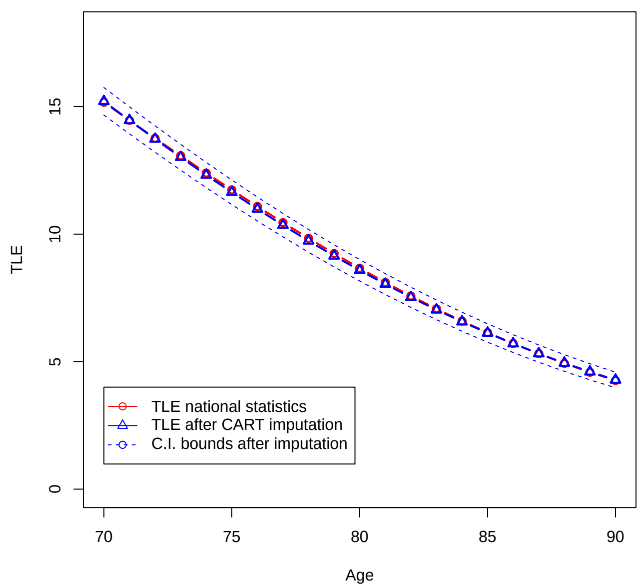

Figure 3: Total life expectancies from HID survey compared with 1998 national life expectancy

Table 4: Life expectancies according to the initial state of health before and after the imputation of an health state (disability free is coded 1 and disabled is coded 2)

Age	TLE (e..)	SE	DFLE (e.1)	SE	DLE (e.2)	SE	e11	e12	e21	e22
	Before imputation									
70	14.77	(0.32)	9.51	(0.26)	5.26	(0.22)	10.14	4.83	6.82	7.11
72	13.35	(0.32)	8.28	(0.25)	5.06	(0.22)	9.02	4.58	5.68	6.77
74	11.99	(0.31)	7.15	(0.25)	4.84	(0.22)	7.99	4.32	4.67	6.39
76	10.72	(0.30)	6.11	(0.24)	4.6	(0.22)	7.06	4.04	3.79	5.98
78	9.52	(0.30)	5.18	(0.23)	4.34	(0.22)	6.22	3.76	3.04	5.54
80	8.41	(0.29)	4.34	(0.22)	4.07	(0.22)	5.47	3.48	2.4	5.09
82	7.39	(0.29)	3.61	(0.21)	3.78	(0.21)	4.8	3.2	1.88	4.64
84	6.47	(0.27)	2.97	(0.20)	3.5	(0.21)	4.21	2.92	1.45	4.2
86	6.04	(0.27)	2.69	(0.20)	3.35	(0.21)	3.7	2.66	1.11	3.77
88	5.64	(0.26)	2.43	(0.20)	3.21	(0.21)	3.25	2.41	0.84	3.37
90	4.9	(0.25)	1.98	(0.18)	2.92	(0.21)	2.86	2.18	0.63	3
	After imputation									
70	15,21	(0,22)	8,59	(0,16)	6,62	(0,17)	9,37	6,10	5,53	8,64
72	13,73	(0,21)	7,32	(0,15)	6,41	(0,17)	8,24	5,83	4,46	8,21
74	12,32	(0,20)	6,16	(0,15)	6,16	(0,16)	7,21	5,53	3,55	7,72
76	10,98	(0,20)	5,11	(0,14)	5,88	(0,16)	6,29	5,21	2,79	7,18
78	9,74	(0,19)	4,17	(0,13)	5,57	(0,16)	5,47	4,88	2,17	6,63
80	8,58	(0,18)	3,36	(0,12)	5,22	(0,16)	4,74	4,54	1,66	6,06
82	7,52	(0,18)	2,67	(0,11)	4,86	(0,15)	4,11	4,20	1,26	5,50
84	6,56	(0,17)	2,09	(0,11)	4,48	(0,15)	3,56	3,87	0,95	4,95
86	5,71	(0,17)	1,61	(0,09)	4,10	(0,15)	3,08	3,54	0,71	4,43
88	4,95	(0,16)	1,23	(0,08)	3,72	(0,15)	2,68	3,24	0,52	3,95
90	4,28	(0,16)	0,93	(0,07)	3,35	(0,15)	2,33	2,95	0,38	3,51

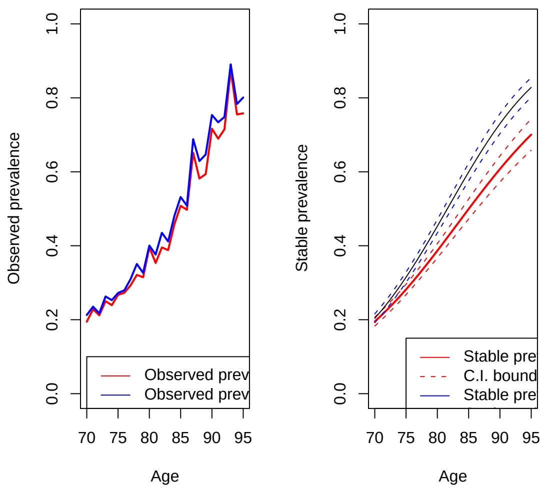

Figure 4: Observed and stable prevalence before and after the estimation of a state of health for those who are lost between the two waves of the HID survey

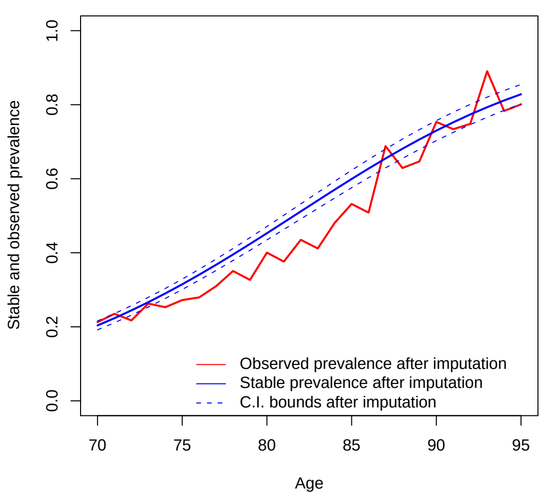

Figure 5: Observed and stable prevalence after the estimation of a state of health for those who are lost between the two waves of the HID survey with 95% confidence interval

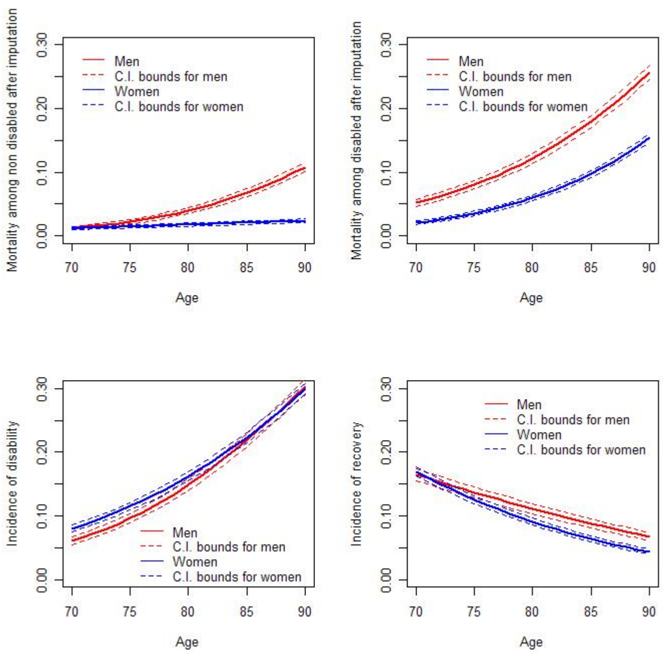

Figure 6: Age specific yearly incidences of mortality for men and women before and after the imputation of a health state for lost individuals known alive, with 95% confidence interval

women at the same age. After imputation, mortality decreases for both sexes, but the gender gap at different ages is almost the same. Globally, for both sexes the probability of dying is higher among the disabled than among the non-disabled. In both cases women show higher onset of disability and lower recovery incidences than men.

These results are reflected on the estimation of health expectancies and stable prevalence implied in the computed probabilities: table 5 shows gender differences in health expectancies after the imputation: the extra years lived by women (about 3.6 years at age of 70) are spent in disability.

Table 5: Life expectancies for men and women according to the initial state of health after the imputation of an health state (disability free is coded 1 and disabled is coded 2)

Age	TLE (e..)	SE	DFLE (e.1)	SE	DLE (e.2)	SE	e11	e12	e21	e22
			Men							
70	13.4	(0.32)	8.6	(0.25)	4.9	(0.23)	9.3	4.5	5.1	7.7
80	7.1	(0.28)	3.3	(0.20)	3.8	(0.23)	4.4	3.3	1.5	4.6
90	3.4	(0.23)	0.9	(0.12)	2.5	(0.22)	2.0	2.1	0.4	2.7
			Women							
70	17.26	(0.30)	8.7	(0.21)	8.56	(0.25)	9.5	7.9	6.0	10.7
80	10.16	(0.26)	3.44	(0.16)	6.72	(0.22)	5.0	5.9	1.8	7.6
90	5.08	(0.22)	0.95	(0.09)	4.14	(0.21)	2.6	3.8	0.4	4.3

4 Summary and conclusions

The HID survey, as other surveys dealing with health, is characterized by quite an important loss of individuals between its waves. This attrition induces a bias in transition probability estimates and, consequently, health expectancies in different states of health is biased.

In this work, health is measured through a functional approach, and people are considered disabled if they are unable or in need of help to perform one or more ADLs.

In order to reduce the bias due to the attrition, we assigned a state of health to individuals known to be alive in the second wave, whose state of health was unknown, through CART.

The correction allows to reduce the bias due to the overestimation of mortality and recovery on the one hand, and to the underestimation of onset of disability on the other hand.

According to our model, people aged 70 can expect to live 9.37 years in disability-free state, given that they were in that state initially, but the expectation is reduced to 5.53

years if they were in the disabled state at age 70. The corresponding health expectancies for the disabled state are 6.10 and 8.64 years respectively. Regardless the initial state of health, people aged 70 can expect to live 15.2 years, of which 6.6 in disability. The main effect of CART imputation on health expectancies is related to the increase of life expectancy of 0.62 of a year, due to the increase of disabled life expectancy of almost 1.2 years, associated to the reduction of disability free life expectancy of 0.5 of a year.

After the imputation, the slope of the stable prevalence seems to be always lower than the slope of the cross sectional prevalence, and globally there is no evidence of a general reduction in ADL.

The gender differences on expected life free of disability did not change significantly after imputation. Nevertheless, women show higher onset of disability and lower recovery; and these results are reflected on the estimation of health expectancies and stable prevalence.

References

[1] Andrade, F.C.D. (2010)Measuring the Impact of Diabetes on Life Expectancy and Disability-Free Life Expectancy Among Older Adults in Mexico J Gerontol B PsycholSciSocSci(2010) 65B(3).

[2] Breiman, L, Friedman J.H., Olshen R.A., Stone C.J. Classification and Regression Trees.Wadsworth International Group, Belmont, California (1984).

[3] Bonneuil, N. Brouard, N. Methods of calculation of health expectancy: application to the LSOA surveys (1984-86-88). In: 5th meeting of the International Network on Health Expectancy (REVES-5): Future Uses of Health Expectancy Indices, Ottawa (1992).

[4] Brouard, N. Structure et dynamique des populations. La pyramide des annees a vivre, aspect nationaux et exemples ragionaux. Espace Populations Societes (2), 157-168 (1986).

[5] Brouard, N. Esperance de vie active, reprises d'activie feminine : un modele. Revue economique (1980), vol 31 pp. 1260-1287.

[6] Cagiano de Azevedo R.(ed.), The European Welfare in a Counteraging Society, ed. Kappa, Roma. (2004)

[7] Cambois, J.M. Robine, and N. Brouard, 1999, "Life Expectancies Applied to Specific Statuses. A Historyof the Indicators and Methods of Calculation", in Population, an English Selection, vol. 11 (1999)

[8] Cambois E., A. Lievre, (2004), Risques de perte d'autonomie et chances de recuparation chez les personnes agees de 55 ans ou plus : une evaluation a partir de l'enquete Handicaps, incapacite, dependance in : Etudes et Resultats, n.349, DRESS, Paris

[9] Crimmins E.M, Hayward M.D, Hagedorn A, Saito Y.; Brouard N.. Change in Disability-Free Life Expectancy for Americans 70 Years Old and Over. Demography 2009;46(3):627-646.

[10] Freedman VA. Late-Life Disability Trends: An Overview of Current Evidence. Workshop on Disability in America: A New Look. Summary and Background Papers Field MJ, Jette AM, Martin L, editors. Washington, DC: National Academies Press, 2006; Also available online at http://www.nap.edu/catalog/11579.html.

[11]

[12] Giudici C., " Les determinants socio-demographiques de la sante aux grands ages " workingpaper, Les Lundis de l'INED, Parigi, 22 maggio 2006

[13] Giudici C, Arezzo M F. Social inequalities in health expectancy of elderly: evidence from the HID Survey. In: IUSSP-UIESP , Ed. XXXVI International Population Conference / XXVI eme Congres international de la population; Marrakech: IUSSP-UIESP. P. 60.(2009)

[14] Giudici C, Arezzo M F., Brouard, estimation of health expectancy in presence of missing data - evidence from the HID Survey. Reves 2010 meeting, La Havana, Cuba, May, 19-21 2010

[15] Guillot, M. The cross-sectional average length of life (cal): a cross-sectional mortality measure that reflects the experience of cohorts. Population Studies 57(1), 41-54 (2003).

[16] Hastie T., Tibshirani R., Friedman, J. The Elements of Statistical Learning. Springer. (2001).

[17] Jagger, C., E. Goyder, M. Clarke, N. Brouard, and A. Arthur (2003) "Active life expectancy in People With and Without Diabetes", in: journal of public health medicine n. 25 p.42-46.

[18] Laditka, S.B. and J.N. Laditka. Effects of Diabetes on Healthy Life Expectancy: Shorter Lives With More Disability for Both Women and Men. In: Longer Life and Healthy Ageing, edited by Zeng Yi, E.M. Crimmins, Y. Carriere, and J.-M. Robine. Dordrecht, The Nederlands, Springer p. 71-90. (2006).

[19] Laditka J.N. and S.B. Laditka. Increased hospitalization risk for recently widowed older women and protective effects of social contacts. Journal of women and aging, vol. 15 n. 2/3 (2003)

[20] Laditka, S. B., Wolf, D. New method for analyzing active life expectancy, Journal of Aging and Health, 10(2) , p. 214-241 (1998).

[21] Lievre A, Brouard N, Heathcote C.R. The estimation of health expectancies from cross-longitudinal surveys. Mathematical Population Studies 10:211-248.(2003).

[22] Lievre A., F. Jusot, T. Barnay, C. Sermet, N. Brouard, J.M. Robine,M.A. Brieu, F. Forette. Healthy Working Life Expectancies at Age 50 in Europe: a New Indicator. In: The Journal of Nutrition, Health & Aging, v.11 n. 6. (2007)

[23] Manton, K.G. and K.Land. Active Life Expectancy Estimates for the U.S. Elderly Population: A Multidimensional Continuous Mixture Model of Funtional Change Applied to Completed Cohorts, 1982-1996. In: Demography n.37 p.253-65.(2000).

[24] Minicuci, N., M. Noale, S.M.F. Pluijm, M.V.Zunzunegui, T. Blumstein, D.J.H. Deeg, C. Bardage, and M. Jylha. Disability free life expectancy: a Cross National Comparison

of Six Longitudinal Studies on Ageing. The CLESA Project. In: European Journal of Ageing n.1, p. 37-44. (2004).

[25] Perenboom R.J.M., Health expectancies in European Countries, in J.M. Robin, C. Jagger, C.D. Mathers, Determining health expectancies, John Wiley and Sons Ltd, Chichester, 2003

[26] Reynolds, S.L., Y. Saito and E.M. Crimmins. The impact of obesity on Active Life Expectancy in Older American Men and Women. The Gerontologist n.45 p.438-444. (2005).

[27] Robine J.M., Jagger C. Allongement de la vie et etat de sante de la population. Population et Societe, Volume VI, Editions de l'INED, 2004, p. 51-84. (2004).

[28] Robine J.M., C. Jagger, C.D. Mathers, E.M. Crimmins, R.M. Suzman, (2003) Determining Health Expectancies,ed. Wiley, England

[29] Rogers, A. Introduction to multiregional mathematical demography. (1975) Wiley.

[30] Sermet C., E. Cambois. Measuring the State of Health, in G. Caselli, J. Vallin, G. Wunsch,Demography, Analysis and Synthesis, vol. 2, Academic Press Elsevier. (2006).

[31] Sullivan, D. A single index of mortality and morbidity. HSMHA health report 86(4), 347-354 (1971).

[32] Verbrugge LM. Recent, Present, and Future Health of American Adults. Annual Review of Public Health. 1989;10:61.

[33] Wolfbein, S. The length of working life. Population Studies (3), 286-294 (1949).

[34] Willekens, F. Computer Program for Increment-decrement (multistate) Life Table Analysis: A User's Manual to Lifeindec. Working papers of the International Institute for Applied Systems Analysis (1979).

[35] Wu, X. Kumar, V. The Top Ten Algorithms in Data Mining. Chapman & Hall/CRC.(2009).