

**REVES 2010  
HAVANA, CUBA-  
MAY 19-21, 2010**

***Public Health Intervention to Increase Health Expectancies  
Session 5 - Methods – measures, models, simulations***

**Estimating health expectancy in  
presence of missing data: an application  
using HID survey.**

**Cristina GIUDICI\*, Maria Felice AREZZO\*, Nicolas BROUARD\*\***

·\*La Sapienza, University of Rome

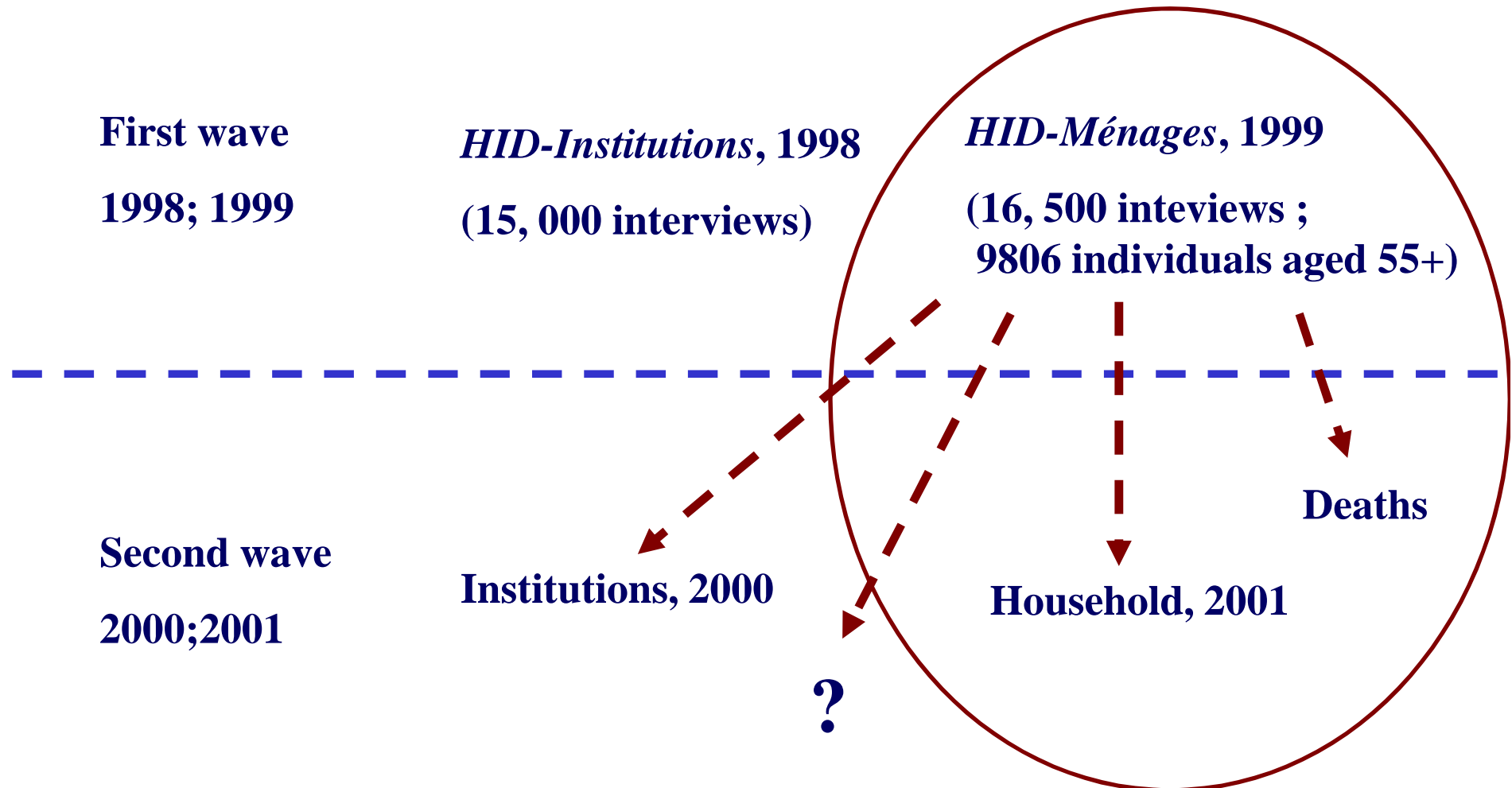
·\*\*Institut National d'Etudes Démographiques (INED)

# *1 - Introduction*

- Aim: Estimation of life expectancy with and without ADL disabilities in France
- Methods: Markov-based multistate life table approach (using IMaCh)
- Data: French National survey on handicaps, disabilities and dependency (HID)

## 2 - Data

### The national survey on handicaps, disabilities and dependency (HID)



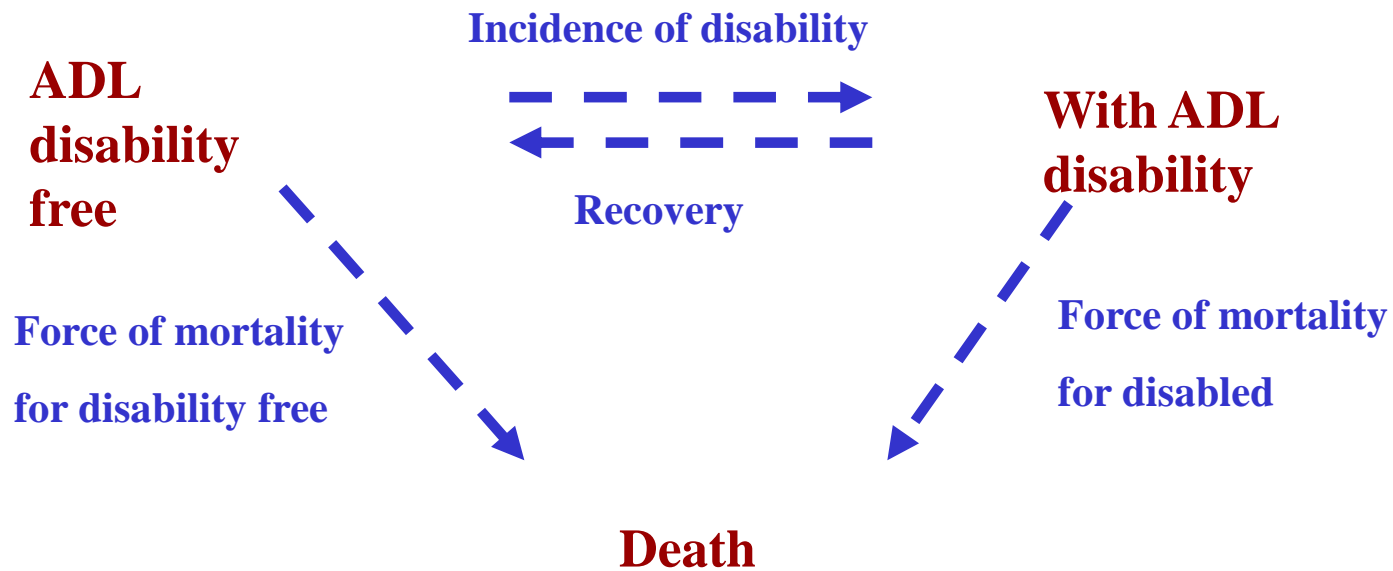
## *Deaths*

- Most of deaths are reported by the interviewer (neighbourhood of former household) (665)
- From vital statistics data some deaths are not confirmed
- Globally 115 deaths are added.
- Thus, 775 deaths are recorded from the 9806 individuals interviewed two years before

## 2.1 - Missing data at second wave

<i>Shorter sample</i>	
Excluded for a second wave (budget...)	39%
<i>Standard reasons of attrition</i>	
Refusal to participate	37%
Change of address (institutionalised?)	21%
Institutionalised	3%
<i>Total</i>	<i>100%</i>

### 3 – Estimation of health expectancy

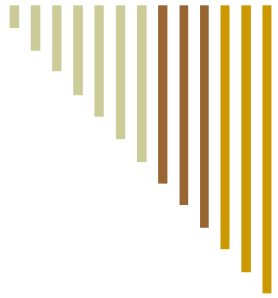


## 3.1 - Methods

*IMaCH (Interpolation of Markov Chains)*

$$\ln \left( {}_h p_x^{ij} / {}_h p_x^{ii} \right) = a_{ij}(h) + b_{ij}(h) * x$$

${}_h p_x^{ij}$  = probability to be in the state  $j$  at time  $h$  for an individual of age  $x$  and in state  $i$  at time 0.

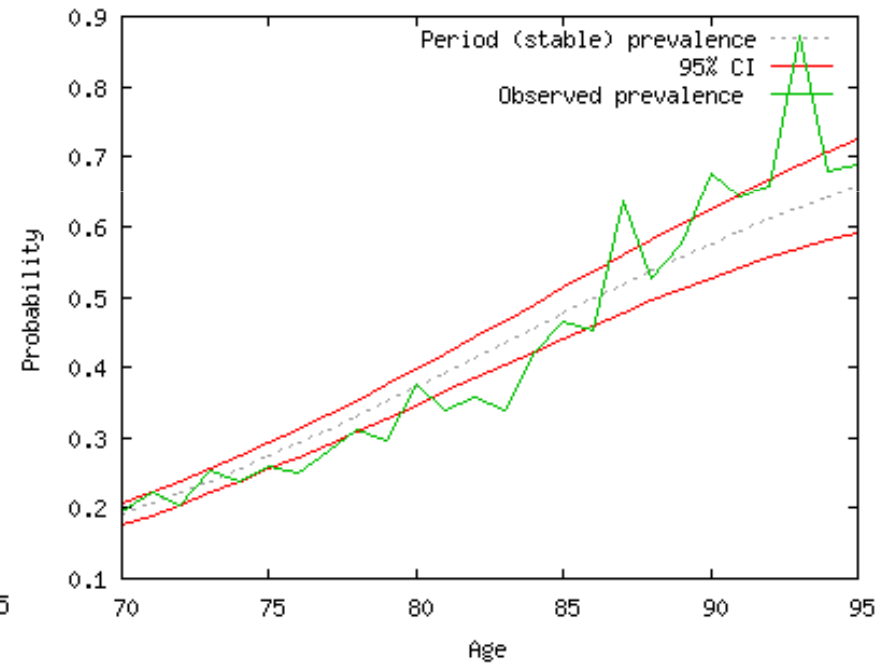
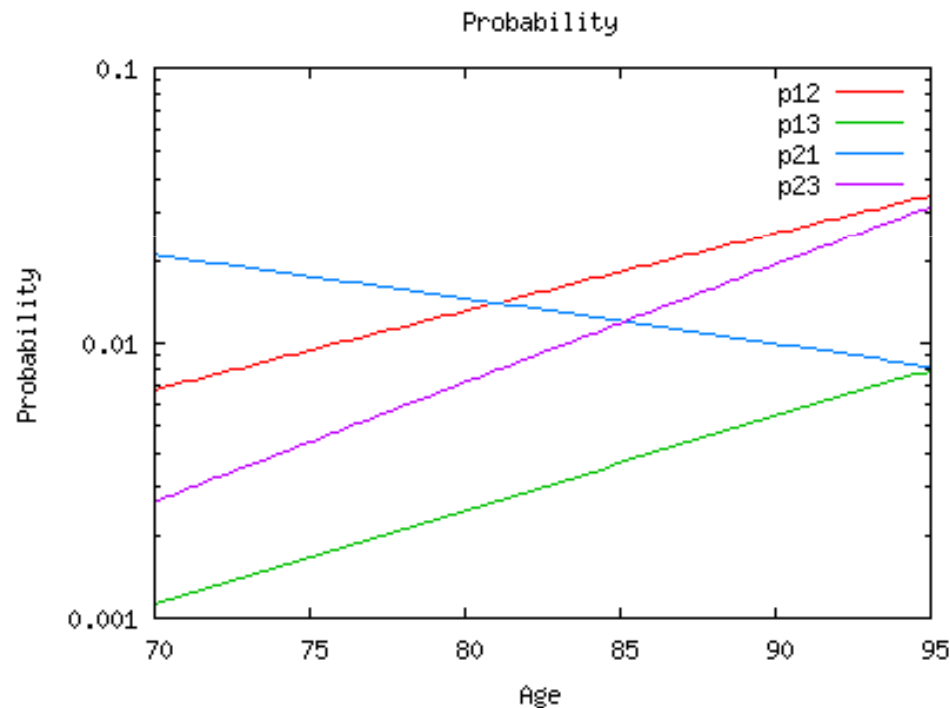


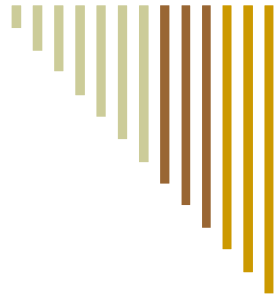
# 1° Analysis

- People whose address changed are supposed to be institutionalized (21%) and all institutionalized people are supposed to be disabled (3%+21%)
- without assumption on unknown health status at the second wave (1923 individuals) which are dropped (by IMaCh)
- Therefore, we introduced a bias by taking into account deaths without compensation for survivors



# 1- Transition probabilities and prevalence of disability using the data set before Vital statistics confirmation

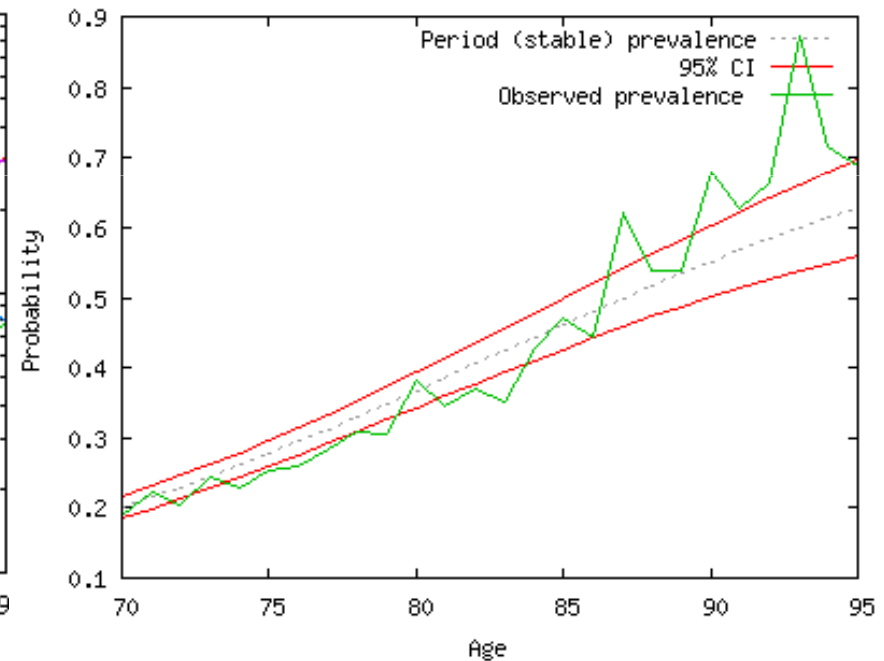
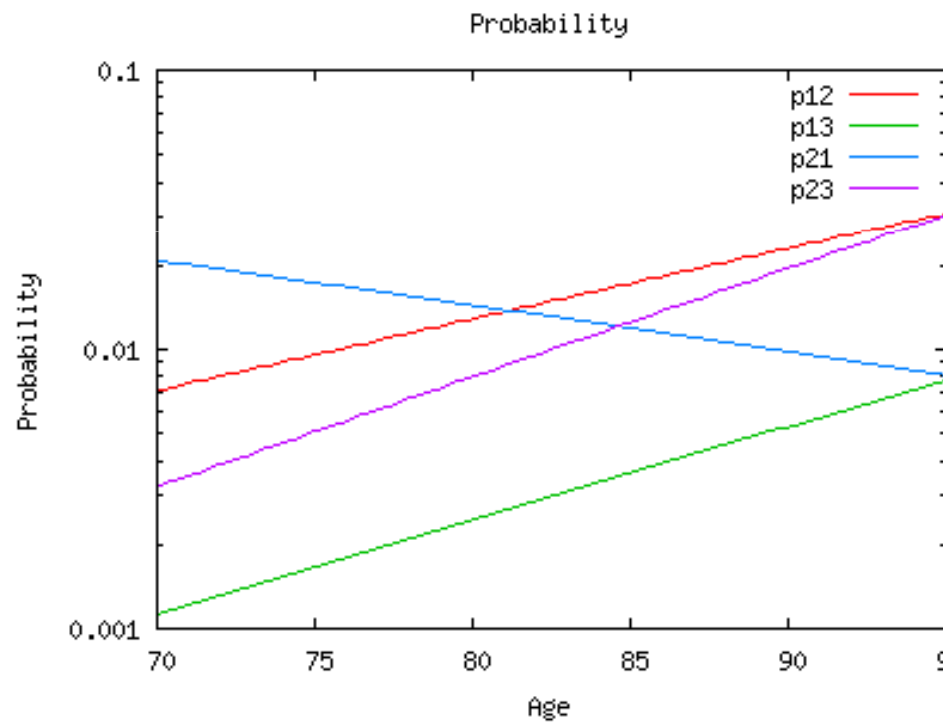




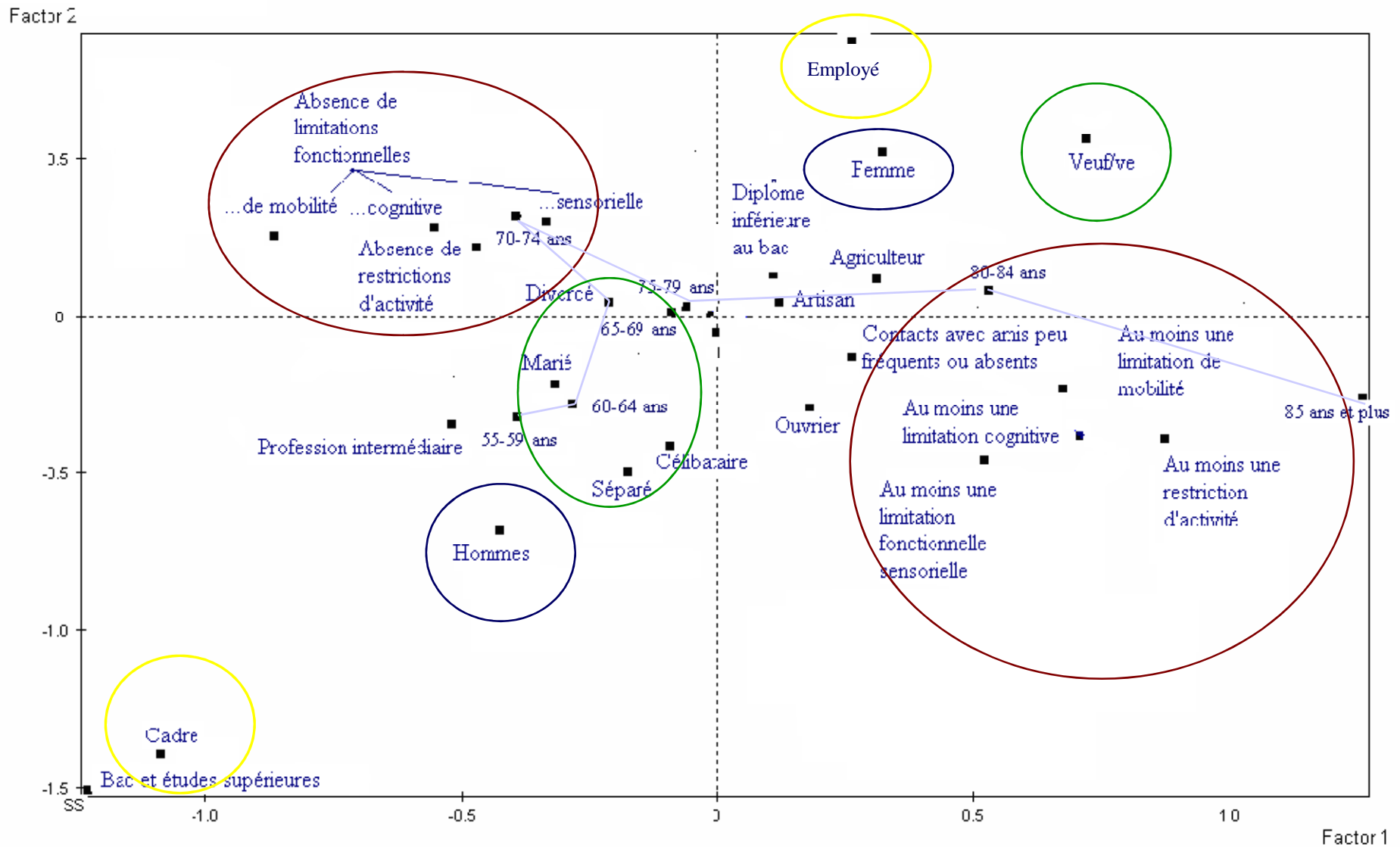
## 2nd Analysis

- The same analysis with vital statistics confirmation of deaths

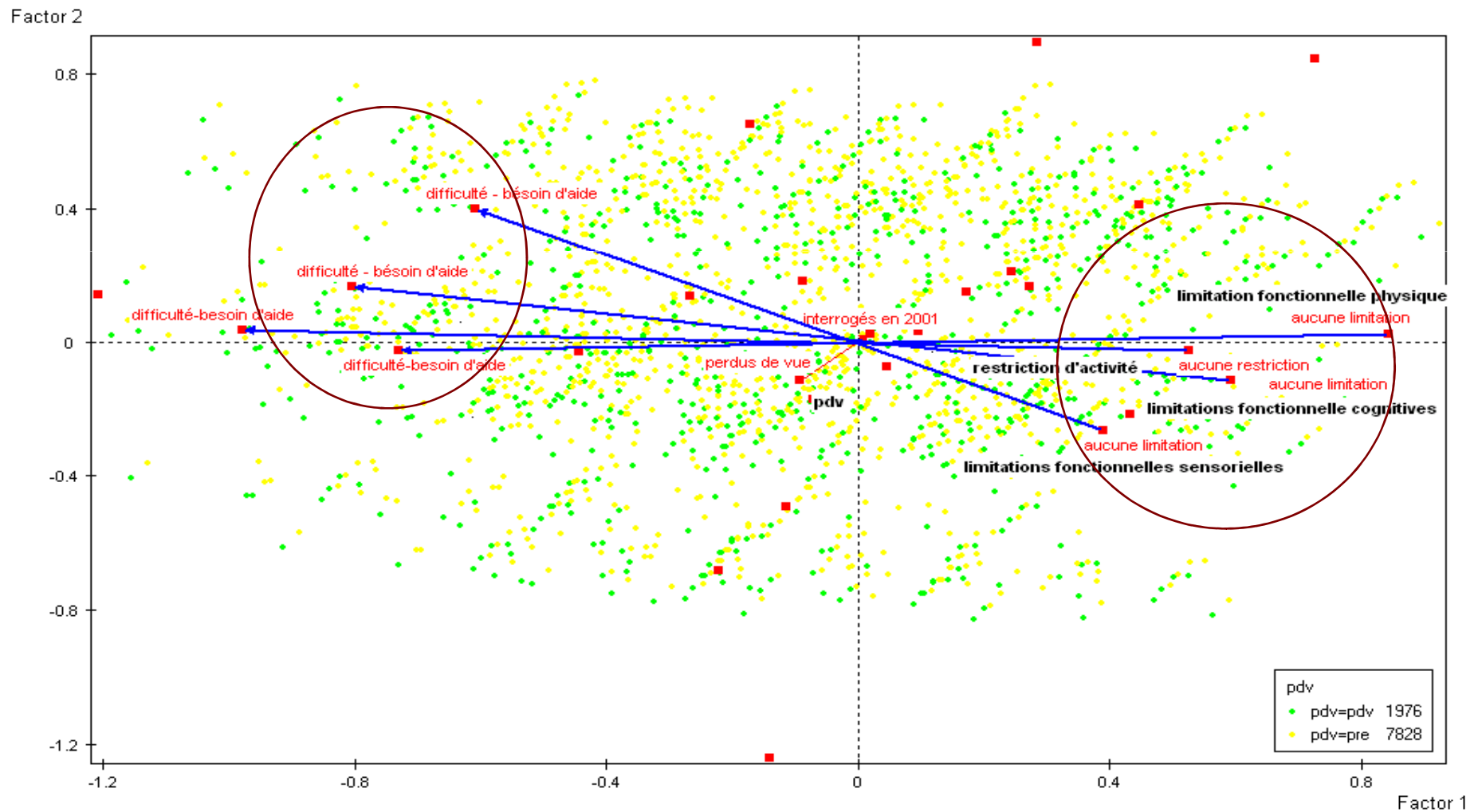
## 2 - Transition probabilities and prevalence of disability adding information coming from Vital Statistics



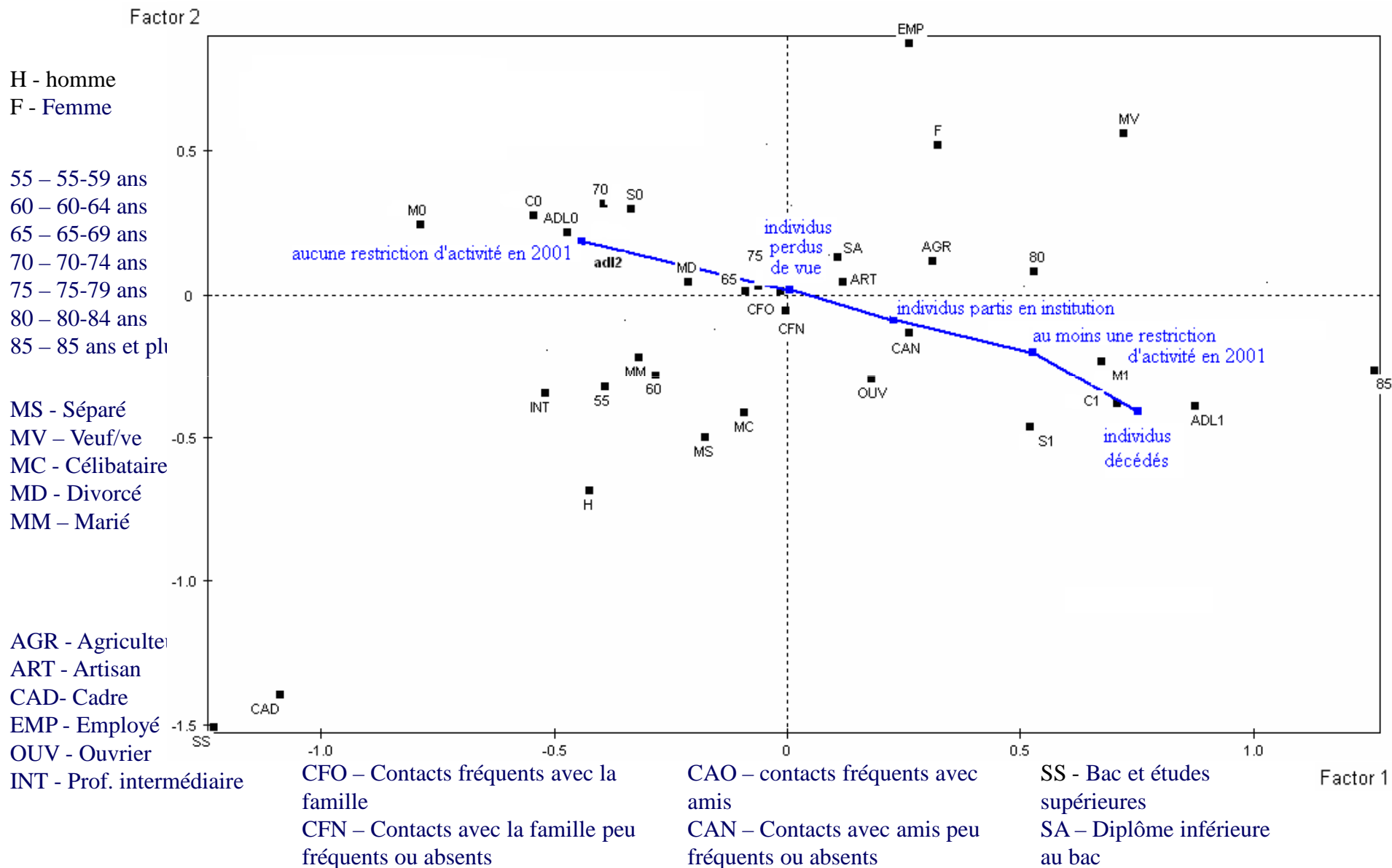
# Multiple correspondence analysis (baseline)



*State of health of missing individuals : evidence from the Multiple Correspondence Analysis (shift of the point 'missing data when institutionalised people are supposed to be disabled*



# The state of health of missing individuals in 2001 could be estimate from available information at the baseline



# Estimation of the health status of missing individuals

## Classification and Regression Trees Analysis (CART)

**It is a supervised classification algorithm: a mathematical rule which assign a new object (individual) to a class  $j$  defined by the algorithm.**

CART allows to estimate the health status of those individuals for which no information on health evolution is available, on the basis of the following variables

**Au moins une restriction d'activité (ADL) en 1999**

**Au moins une difficulté de gestion**

**Etat de santé perçu**

**Age**

**Au moins une limitation fonctionnelle cognitive**

**Vie sociale (vacances)**

**Vie culturelle active**

**Isolement**

✓ **Présence d'un soutien affectif ou moral**

✓ **éducation**

✓ **Catégorie socio professionnelle**

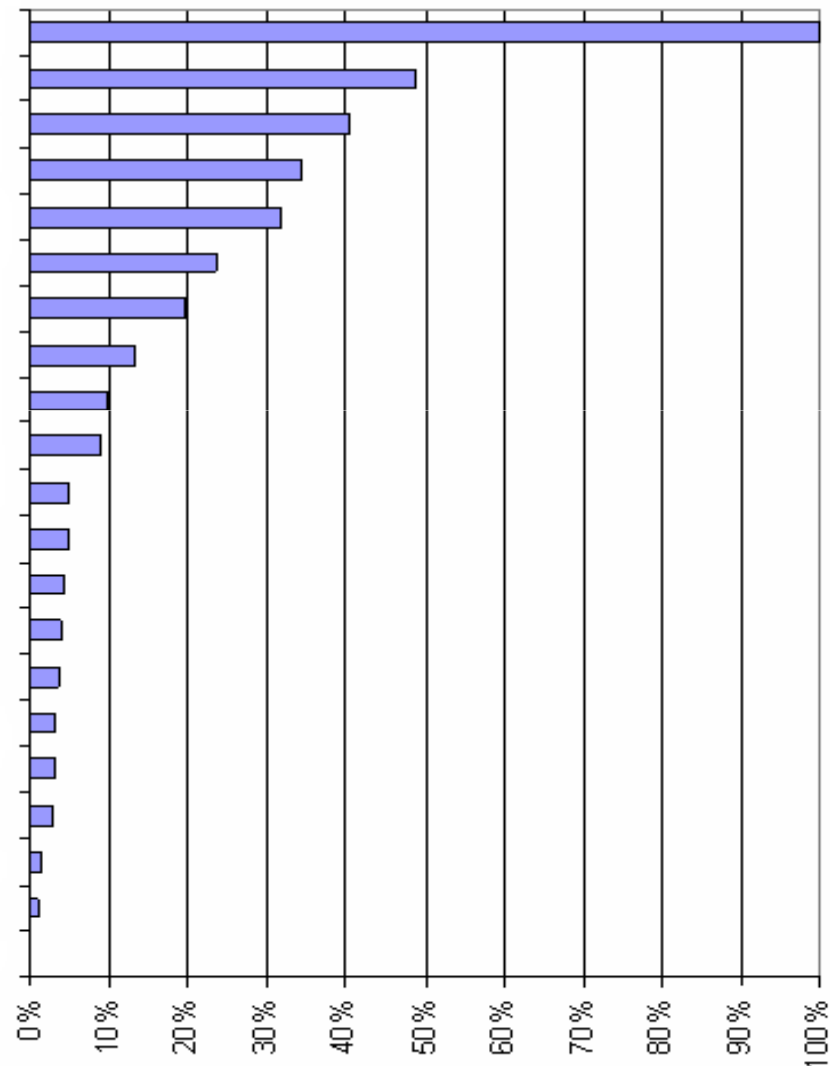
✓ **Nombre d'enfants**

✓ **Limitation fonctionnelle sensorielle**

✓ **Nombre de frères**

...

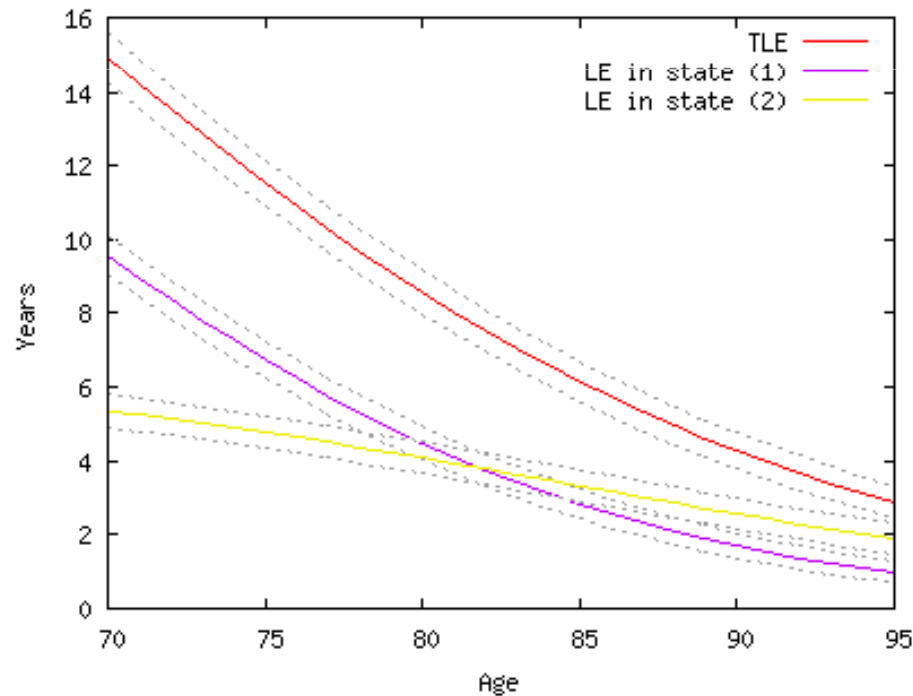
Normalized Importance



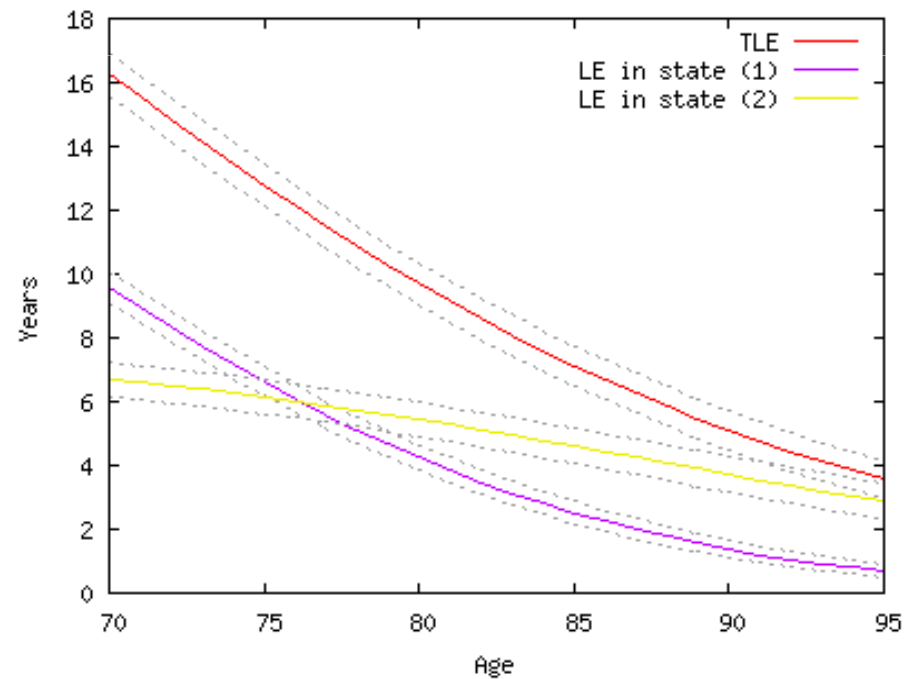


# *Effect of the estimation on total and healthy life expectancy*

Health expectancy  
before the estimation of  
missing data

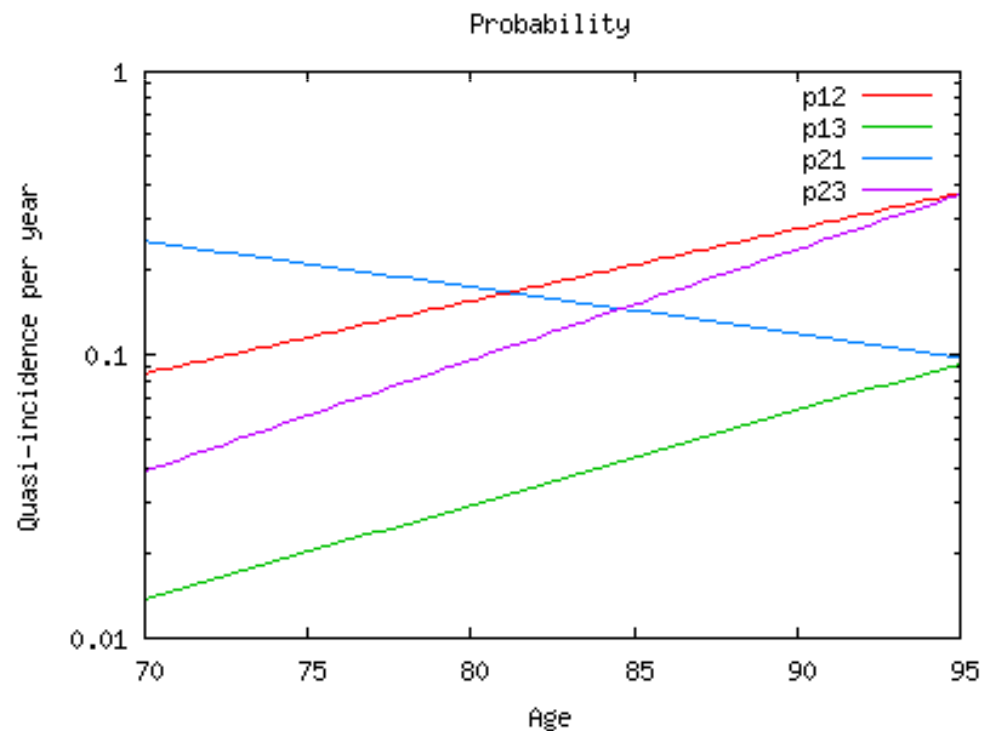


Health expectancy after  
the estimation of missing  
data

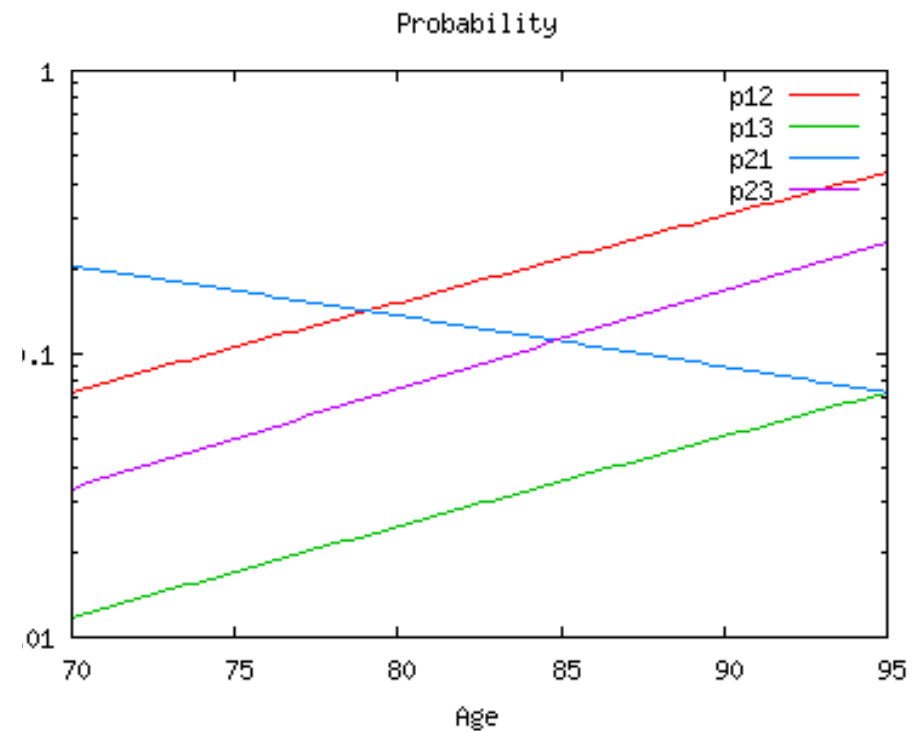


# *Effect of the estimation on transition probabilities*

Before the estimation of missing data

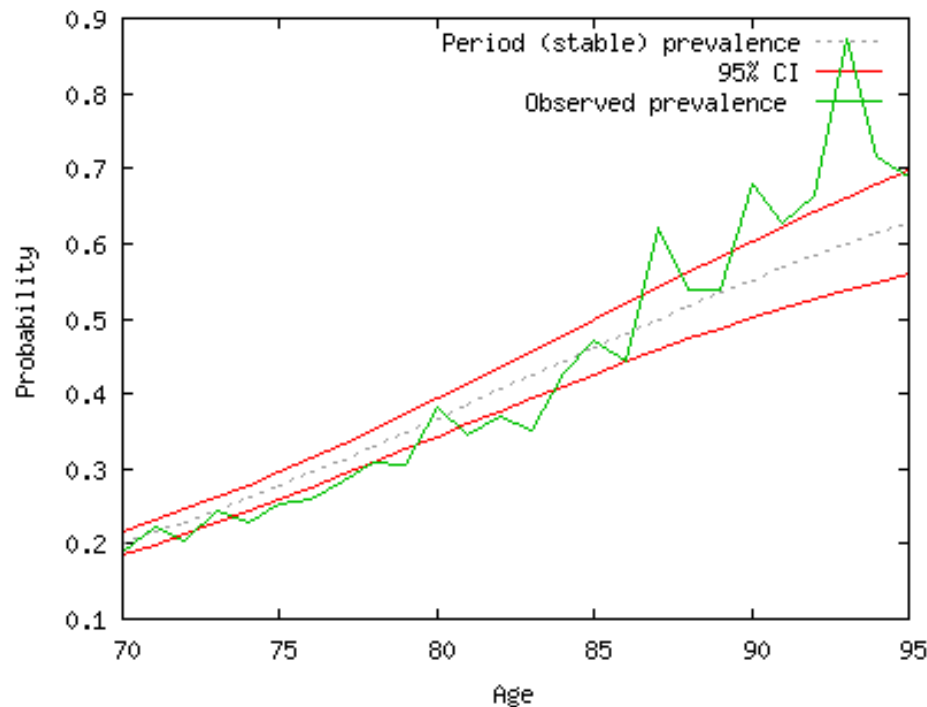


After estimation of missing data

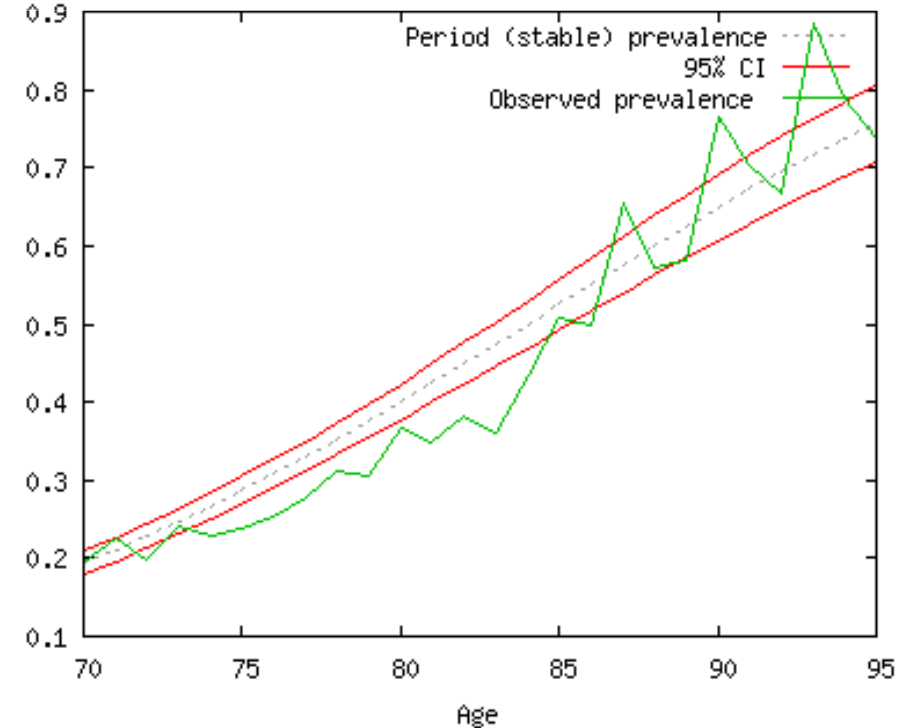


# *Effect of the estimation on cross sectional and stable prevalence*

Before the estimation of missing data



After the estimation of missing data



## *Conclusions*

- The slope of the stable prevalence seems to be always lower the slope of the cross sectional prevalence
- Recent cohort of 75-85 years old seems to accelerate the onset of disability
- There is no evidence of reduction in ADL disability