Methodology changes in the Eustat Survey on the Population in Relation to Activity –(PRA). 1st Quarter 2005

Introduction. Types of Changes and background

The changes that Eustat has introduced in its continuous survey on workforces, Population in Relation to Activity-PRA, are based on three main motives: adapting methodology to international requirements in relation to the demand for information, to take on the social changes that affect the object of analysis from the perspective of researched information and collection methods and, finally, to update the technological framework that supports this operation.

These reasons have resulted in methodological changes that affect from the size and sampling design, to the sampling framework and the treatment of balancing, modifications that affect statistical tools, such as population projections, and to building and migrating to a new IT setting, based on an integrated information system.

Outline of fundamental changes:

a. Methodology changes that affect the sample

   a1. Increase in the sampling size
   a2. Change in sampling type
   a3. Reduction in the rotation of the sample
   a4. Change in the sampling framework
   a5. Modifications in data collection

b. Adaptation of the Questionnaire

   b1. Reduction in response load
   b2. Questionnaire on structure

c. Modifications in the treatment and elevation and post-stratification tools

   c1. Population projections with base year 2001
   c2. Elevation and post-stratification procedures

d. Design, production and migration to a new Data Base setting

   d1. Linking to the population system in the Statistical Population Register
   d2. Data Base design
   d3. Construction of historic tables

1.- Design changes and sample size

The progressive and significant reduction in some key object groups for analysis in the survey over the last five years (unemployed, some economic sectors by province, etc.), has inevitably led to the fall in their representation in the sample or, which amounts to the same, to an increase in the margin of error in the estimation.

Although there are indirect procedures in the research phase to improve the estimations of these populations (Small Area treatments), Eustat also considered approaching two alternative methods: **increase the sample size**, from **3,750 family households per quarter to 5,088** (a 35.7% increase) and **Change from a two-stage sampling system to a simple random system (m.a.s.).**

Since the publication of the first results of the PRA in 1985 to 1998, the sample size was similar to the present sample, 5,000 family households.

Two-stage sampling was maintained to that date: the selection of a subgroup of sections using a proportional stratified system according to the typology, as a first phase, and the extraction of households by systematic circular sampling with equal probability, as a second. The typology of sections presented two fundamental problems: each Census was reviewed periodically –every five years in the case of Eustat-, which means there was a risk of the sample ageing and, on the other hand, there was a higher level of sampling error.

With the elimination of the first phase of the design, the sample seems to be better distributed in the province and it is possible to use more up-to-date sampling frames. Surveying the 4th quarter of 2004 and the first of 2005 is based on a **new sampling framework** configured by the **Statistical Population Register and the Eustat Territory System.**

A medium term aim is to have annually updated frameworks (Survey year –1).

One of the prime objectives of surveys on workforces is to offer information on variations in the study groups over the whole year. This circumstance obliges us to distribute the sample of households over all the weeks in the year and, vis-à-vis offering more consistent evolution data, to build panels of households (permanence in the sample of part of the total of households and constant renewal of the other part).

To date families in the PRA sample remain in the panel 6 quarters and 1/6 of the sample was renewed each quarter. In order to reduce distortions which occur when introducing a new sample and returning to the pre-1998 scheme, **permanence on the panel is prolonged to 8 quarters and renewal involves 1/8 of the sample each quarter.**

The change in the type of sampling and even the increase in size are linked to **giving priority in the collection system to the CATI procedure** (computer assisted telephone interviewing). To date the first interview (entry in the panel) was made through the CAPI system (computer assisted personal interviewing) and successive interviews, unless the family wished otherwise or there was no telephone, were carried out using the CATI system. In the new PRA sample we attempted to locate the families' telephone numbers and unless they could not be found, the family requests otherwise or there was a negative telephone response, personal interviewing was not involved.

Through these changes and some other secondary changes in collection procedures an attempt was made to improve the sample coverage level.

2.- Adapting the Questionnaire

Four criteria were followed in order to review the questionnaires used: analysis of the demand for information on activity, revision of the information required to obtain the Eurostat structural and quality indicators on employment and suggestions by experts and preferential users were taken into consideration from the survey. Furthermore, we considered the possibility of using information from secondary sources.

The proposed aim, in addition to updating the questionnaire, was to reduce the response load. To this effect the **Questionnaire** was also subdivided **in a trend model** (group of items which are always carried out) **and a structure model** (carried out on entry in the panel and/or once a year). This means that the basic variables and variables on the activity of the population will continue to be disseminated quarterly and structural variables (requiring analysis which do not involve seasonalization) will be published as annual averages. Among the latter variables are the relation to activity in the Eustat classification version, the level of education, type of working day, professional situation and type of contract.

It may be said that with regard to the previous version, the trend questionnaire reduces the response load by 41.6 (measured through the number of items) and the structure by 19.4%.

The new questionnaire may be consulted at :

http://www.eustat.es/document/datos/pra05_c.pdf

3.- Analysis of the effects of methodological changes

3.1-Application of projections with Base 2001

As usually occurs, once the Census on Population and Housing is concluded, updating is carried out, based on the new population and its structure, and an analysis is made of the components of the population variation, projections which are used to elevate the sampling results.

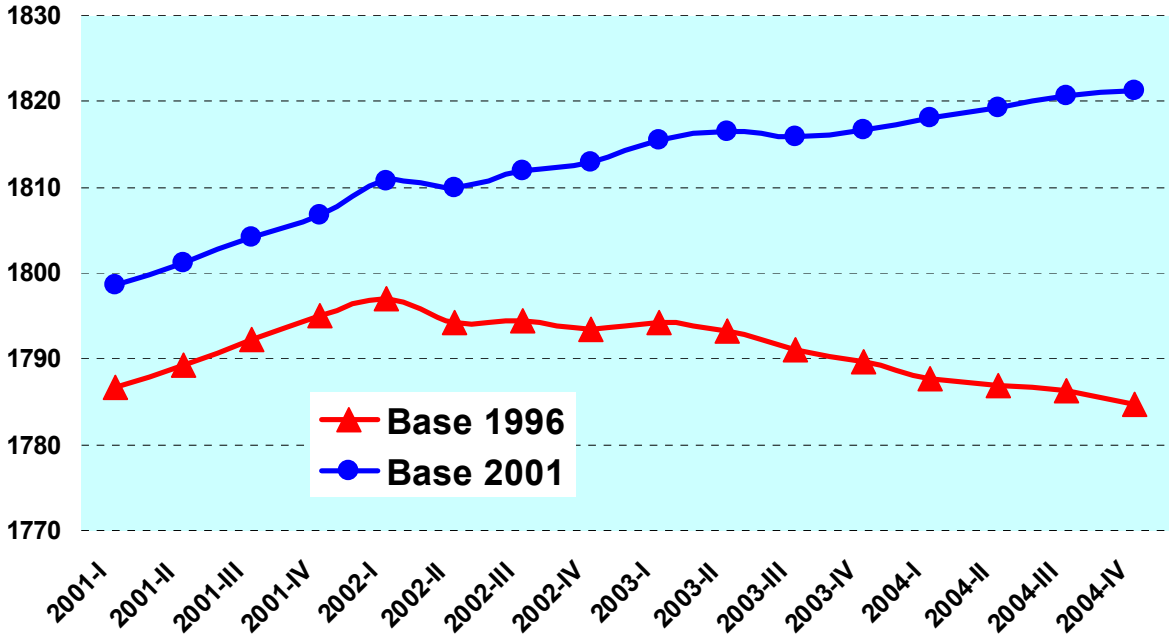
Previously, projections with base 1996 were used, the year when Eustat produced the Population and Housing Statistics, a census operation.

In short, it may be said that the new projections indicate a reversal of the population variation trends that were presented in projections with base 1996: the population tends to grow (as the figures from register updates appear to corroborate) and the component that to a large extent lies behind this increase is migration. The arrival of foreigners is fundamentally producing positive migration balances in recent years.

This type of population, given its demographic characteristics and, of course, the reason for their movements, directly affect the composition of the population according to the activity.

Projections with base 2001 have been applied to the PARA files for 2001 to 2004 and have been compared to data published with the projections with base 1996. In Tables 1 and 2 we find the estimated series with the new projections and those published to date – base 1996-.

Graph 1. Series of population aged 16 and over estimated with projections with base 1996 and base 2001 by quarter. A.C. of the Basque Country. 2001-2004 (thousand).



Source: Eustat

Table 1. Series of population aged 16 and over, of employed and unemployed estimated with projections with base 1996 (published) and base 2001 by quarter.
A.C. of the Basque Country. 2001-2004. (miles)

	Population aged 16 and over			Employed			Unemployed		
	Base 1996	Base 2001	Difference	Base 1996	Base 2001	Difference	Base 1996	Base 2001	Difference
2001									
I	1.786,6	1.798,5	11,9	830,5	839,1	8,6	115,5	116,9	1,4
II	1.789,2	1.801,2	12,0	846,9	856,2	9,3	104,3	105,6	1,3
III	1.792,2	1.804,1	11,9	855,5	865,4	9,9	111,3	112,5	1,2
IV	1.795,0	1.806,8	11,8	864,8	875,1	10,3	94,9	95,9	1,0
2002									
I	1.797,0	1.810,6	13,6	878,9	890,4	11,5	81,0	81,8	0,8
II	1.794,3	1.809,8	15,5	888,4	901,2	12,8	73,0	73,9	0,9
III	1.794,5	1.812,0	17,5	894,5	908,5	14,0	80,3	81,4	1,1
IV	1.793,5	1.813,0	19,5	891,2	906,4	15,2	88,0	89,3	1,3
2003									
I	1.794,2	1.815,5	21,3	887,5	903,9	16,4	88,5	90,0	1,5
II	1.793,3	1.816,4	23,1	895,2	912,8	17,6	82,6	84,1	1,5
III	1.791,0	1.816,0	25,0	905,0	924,0	19,0	79,8	81,4	1,6
IV	1.789,7	1.816,6	26,9	897,3	917,2	19,9	86,8	88,7	1,9
2004									
I	1.787,6	1.818,1	30,5	897,3	919,7	22,4	84,8	86,8	2,0
II	1.786,9	1.819,3	32,4	901,7	925,2	23,5	78,5	80,6	2,1
III	1.786,2	1.820,6	34,4	908,2	933,0	24,8	74,3	76,3	2,0
IV	1.784,8	1.821,2	36,4	912,5	938,6	26,1	68,6	70,7	2,1

Source: Eustat

Table 2. Series of Activity and Unemployment Rates of the population aged 16 and over estimated with projections with base 1996 (published) and base 2001 by quarter. A.C. of the Basque Country 2001-2004. (%)

	Activity rate			Unemployment rate		
	Base 1996	Base 2001	Difference	Base 1996	Base 2001	Difference
2001						
I	52,9	53,2	0,3	12,2	12,2	0,0
II	53,2	53,4	0,2	11,0	11,0	0,0
III	53,9	54,2	0,3	11,5	11,5	0,0
IV	53,5	53,7	0,2	9,9	9,9	0,0
2002						
I	53,4	53,7	0,3	8,4	8,4	0,0
II	53,6	53,9	0,3	7,6	7,6	0,0
III	54,3	54,6	0,3	8,2	8,2	0,0
IV	54,6	54,9	0,3	9,0	9,0	0,0
2003						
I	54,4	54,7	0,3	9,1	9,1	0,0
II	54,5	54,9	0,4	8,4	8,4	0,0
III	55,0	55,4	0,4	8,1	8,1	0,0
IV	55,0	55,4	0,4	8,8	8,8	0,0
2004						
I	54,9	55,4	0,5	8,6	8,6	0,0
II	54,9	55,3	0,4	8,0	8,0	0,0
III	55,0	55,4	0,4	7,6	7,6	0,0
IV	55,0	55,4	0,4	7,0	7,0	0,0

Fuente: Eustat

The new projections show growing increases in the population since 2001 (almost 12,000 persons more than in 2001 to 33,400 more on average in 2004).

These increases are absorbed by the working population to a large extent: with regard to the 1996 estimations, the new estimations in 2001 add a total of 9,500 workers. This figure came to 24,100 on average in 2004. Unemployment also suffered a slight increase: from 1,200 more in 2001 to 2,000 more in 2004.

Nevertheless, changes in the level of employment and unemployment due to the methodological effect of the application of new projections practically disappear if we take the activity and unemployment rates as indicators: the former varies between point 3 and point 5 and the latter remains the same.

With regard to the variables that are used for the post-stratified phase, it may be said that adjustments by sex and age by province are maintained for the population. In relation to the families, the estimated family figures are balanced by the size and province, as this treatment was not carried out until the 1st quarter of 2005 in the new PRA version.

3.2- Effects of the change in design

In order to assess the effect of the methodological changes and obtain bridge data, during the 4th quarter of 2004 and the 1st of 2005 two samples of the PRA were taken, the usual sample with 3,750 households and the new sample with 5,088 and using the new sampling system. If both versions are exploited in all cases applying the projections with base 2001, we notice the changes that were introduced to the new design. We must add others, more difficult to assess, and which must lead to their disappearance (effect of the new field equipment, effect of the new sample to 100%, new Questionnaire, etc.).

In Tables 3 and 4 we see the changes in the main groups in the survey.

Table 3. Population aged 16 and over by relation to activity according to methodolgy PRA 1998 and PRA 2005, for the 4th quarter of 2004 and the 1st of 2005. A.C. of the Basque Country. (thousand)
Data with revised population (base 2001)

	4th QUARTER 2004			1st QUARTER 2005		
	PRA 1998	PRA 2005	Dif	PRA 1998	PRA 2005	Dif
ACTIVE POP.						
Total	1009,3	1007,0	-2,3	1008,7	1006,7	-2,0
Males	580,8	585,3	4,5	579,4	580,8	1,4
Females	428,5	421,7	-6,8	429,4	426,0	-3,4
Álava	147,2	144,6	-2,6	147,0	144,2	-2,8
Bizkaia	530,5	526,7	-3,8	535,1	532,7	-2,4
Gipuzkoa	331,6	335,7	4,1	326,6	329,9	3,3
EMPLOYED POP.						
Total	938,6	937,7	-0,9	938,7	936,1	-2,6
Males	543,8	550,3	6,5	546,2	547,0	0,8
Females	394,8	387,4	-7,4	392,5	389,1	-3,4
Álava	139,7	138,8	-0,9	138,8	139,3	0,5
Bizkaia	489,6	483,1	-6,5	493,9	487,2	-6,7
Gipuzkoa	309,3	315,8	6,5	306,0	309,5	3,5
UNEMPL. POP.						
Total	70,7	69,3	-1,4	70,0	70,7	0,7
Males	37,0	35,0	-2,0	33,2	33,8	0,6
Females	33,7	34,3	0,6	36,8	36,9	0,1
Álava	7,5	5,8	-1,7	8,3	4,8	-3,5
Bizkaia	40,9	43,6	2,7	41,2	45,5	4,3
Gipuzkoa	22,3	19,9	-2,4	20,5	20,4	-0,1

Fuente: Eustat

Table 4. Activity and Unemployment Rates of the population aged 16 and over, and employment (16-64 years), according to PRA 1998 and PRA 2005 methodology, for the 4th quarter of 2004 and the 1st of 2005. A.C. of the Basque Country. (%)
Data with revised population (base 2001)

	4th QUARTER 2004			1st QUARTER 2005		
	PRA 1998	PRA 2005	Diff	PRA 1998	PRA 2005	Diff
ACTIVE POP.						
Total	55,4	55,3	-0,1	55,4	55,3	-0,1
Males	65,6	66,1	0,5	65,4	65,6	0,2
Females	45,8	45,1	-0,7	45,8	45,5	-0,3
Álava	58,0	56,9	-1,1	57,8	56,6	-1,2
Bizkaia	53,8	53,5	-0,3	54,3	54,1	-0,2
Gipuzkoa	57,0	57,7	0,7	56,0	56,8	0,8
EMPLOYED POP.						
Total	65,0	64,9	-0,1	65,0	64,9	-0,1
Males	74,7	75,4	0,7	75,1	75,1	0,0
Females	55,2	54,2	-1,0	54,7	54,4	-0,3
Álava	67,5	67,0	-0,5	67,0	66,9	-0,1
Bizkaia	63,2	62,4	-0,8	63,7	63,0	-0,7
Gipuzkoa	67,1	68,2	1,1	66,3	67,2	0,9
UNEMPL. POP.						
Total	7,0	6,9	-0,1	6,9	7,0	0,1
Males	6,4	6,0	-0,4	5,7	5,8	0,1
Females	7,9	8,1	0,2	8,6	8,7	0,1
Álava	5,1	4,0	-1,1	5,6	3,4	-2,2
Bizkaia	7,7	8,3	0,6	7,7	8,5	0,8
Gipuzkoa	6,7	5,9	-0,8	6,3	6,2	-0,1

Fuente: Eustat

Both for the active population aged 16 and over and the working population and the groups considered (totals, sex and province), and for both test quarters, only in the case of a higher variation than 2% (2.1% increase in the working population in Gipuzkoa with the new method–PRA 2005- in relation to the old –PRA 1998- in the data for the 4th quarter of 2004).

Another two groups in this 4th quarter also get close to 2%: a 1.9% reduction in working women which is produced by the new method. Active women in Álava also appear to be undervalued by 1.8% in relation to the old system. Percentage differences that affect data as a result of the design tend to diminish in the 1st quarter of 2005, except in some cases.

Clearer visions are found for the unemployed group: in the 4th quarter the new method shows 2% less and 1% more in the 1st of 2005. The number of unemployed males fell 5.4% in the first quarter and increased in the last quarter by 1.8%.

The provinces suffer sharper variations. However, the percentages must be relativized to absolute figures. The new method offers 22.7% fewer unemployed (1,700) in the 4th quarter of 2004 in Álava than in the old method. This figure shoots up to 42.2% (3,500 fewer unemployed) in the 1st quarter of 2005.

In Bizkaia in the first quarter 2,700 (6,6%) more unemployed were obtained and 4,300 more in the second (10.4%). Gipuzkoa turns out to be the only province that improves from one quarter to another: in the 4th quarter of 2004 the new method goes from giving 10.8% fewer unemployed (2,400) than the old, to giving practically the same figure in the 1st quarter of 2005.

These important variations, above all in the case of Álava, are due to the smaller size of the group and to the effects of methodological changes.

It is foreseeable that the adaptation of the new survey, as seen in some of the variations between the 4th and the 1st quarter used in testing, eliminates some of the undesired differences resulting from the changes.

4.- A new technological setting for the PRA

Given the speed with which computer tools, programmes and equipment change, in addition to the increasing experience of experts, the PRA applications and information file systems were outdated. Furthermore, we should bear in mind the connections between the operation and the information system that is at the nucleus of the Eustat Statistical Population Register.

These two motives led to a new design in an Oracle Data Base setting, with PL-SQL programming for applications and SAS for tabulations. In addition all the historic files of the PRA were transferred to Oracle Tables, in order to facilitate future treatment or use of series.