SAMPLE DELIVERABLE 7.2.

OVERSAMPLING DESCRIPTION-PART II

Grant agreement No:	SSH - CT - 2007 – 217565
Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Funding Scheme:	Collaborative Project - Small or medium scale focused research project
Deliverable n.:	D 7.2.
Deliverable name:	Oversampling Description-Part II
WP n.:	1.2.
Lead beneficiary:	1
Nature:	Report
Dissemination level:	PU
Status:	Final
Due delivery date:	March 31th , 2010
Actual delivery date:	May 18th, 2010
Project co-ordinator name:	Monica Pratesi
Title:	Associate Professor of Statistics - University of Pisa
Organization:	Department of Statistics and Mathematics Applied to Economics of the University of Pisa (UNIPI-DSMAE)
Tel:	+39-050-2216252, +39-050-2216492
Fax:	+39-050-2216375
E-mail:	coordinator@sample-project.eu
Project website address:	www.sample-project.eu

OVERSAMPLING DESCRIPTION-PART II

Project Acronym:	SAMPLE
Project Full title:	Small Area Methods for Poverty and Living Conditions Estimates
Grant agreement No:	SSH - CT - 2007 – 217565
WP no.:	1.2.
Deliverable n.	D7.2.
Document title:	Oversampling Description-Part II
Status:	Final
Editors:	Monica Pratesi, Alessandra Coli
Authors:	Caterina Giusti (caterina.giusti@ec.unipi.it), UNIPI-DSMAE
	Paolo Consolini (consolini@istat.it), ISTAT
	Michela Casarosa (m.casarosa@provincia.pisa.it), PP-UROPS
Document editing:	Marta Garro (garro@sample-project.eu), UNIPI-DSMAE
Language revision:	Marta Garro

Summary

1. Introduction

The first part of Deliverable 7 (D7.1. *Oversampling Description* - Part I) was devoted to the description of the state-of-the-art of the EU SILC oversampling for the Province of Pisa; the synthetic overview of the local contacts led by the Province of Pisa in order to gain access to administrative data; and the description of the methodology for the combined estimation at LAU2 and LAU1 levels.

In this second part, the oversampling description is completed through a synthetic view of the use of administrative data in order to improve the quality of the survey data (§2). The following section outlines the demographic an social profile of the territory of the Province of Pisa (§3,4,5) as it results from the administrative data sources. Direct estimates of some income and discomfort indicators based on Eusilc-oversampling are then provided at a sub-provincial level(§6). Finally, a contextual reading of administrative and survey indicators is provided (§ 7)

2. Data quality enhancement and record linkage strategies in It-Silc and Pisa survey.

 Paolo Consolini (Istat)

This section is devoted to the description of the imputation and correction process carried out by Istat on the income target variables. The leading idea is to exploit administrative data (particularly the tax agency and the pension database) in order to fill in the survey missing values, correct outliers or unreliable values, improve the quality of income estimates. Whenever the administrative data source cannot provide useful information, imputation is made through the IVEware model[1][2] .
Imputation and correction affect the final estimates of income. A measure of such impact is provided both for Italy and the Province of Pisa.

The Italian Eu-Silc survey (briefly It-Silc) uses administrative micro-data in order to reduce the survey measurement errors. Errors can arise in any step of the survey process, i.e. from the questionnaire, the respondent and the interviewer, as well as from the selected methods of collecting data (Kish, 1965). Moreover, errors can occur because some of the sampled units have not been actually observed (non-observation error). Particularly, the "selective non-response" bias is the typical non-observation error in income surveys. According to Van der Laan et al (1997), for instance, because of "selective-non response" bias the Netherlands income is underestimated by as much as 10%.
Istat uses administrative data in order to support the editing and imputation processes of It-Silc survey thus improving the quality of data on income . The process is based on the integration of the It-Silc and administrative datasets at a micro level (P. Consolini, 2008, 2009).
Some basic requirements have to be satisfied. Statistical units are to be defined uniformly in all sources (harmonisation of units), all sources should cover the same target population (completion of populations), all variables have to be defined and classified in the same way among the considered data-sources (harmonisation of the variables and classifications), all data should refer to the same period or the same point in time (Paul van der Laan, 2000). In other terms, administrative data need to be comparable with the It-Silc survey data.

---

1 IVEware was developed by the Survey Methodology Program at the University of Michigan's Survey Research Center, Institute for Social Research.. It uses a multivariate sequential regression approach to imputing item missing values. For details see "A multivariate technique for multiply imputing missing values using a sequence of regression models" by Raghunathan, Lepkowski, Van Hoewyk and Solenberger (*Survey Methodology*, June 2001)

The technique used to link the administrative units to those from sampling surveys is the so-called "exact matching". This technique allows to combine the information related to the same statistical units by means of a collection of identifiers called "match keys", provided that each unit is associated with a unique identifier not affected by errors. Although different typologies of exact matching techniques could be used, we refer here to the simplest "one-to one" relationship, where every statistical unit is associated with only one record among the data-sources involved. In different data-sources, records are matched using the Personal Tax Number. Once the integration has been carried out identification numbers are eliminated and replaced with an internal codes system according to the Italian rules and regulations that protect the "confidentiality" of the individuals.

2.1. The integration of survey and administrative data at micro level

The integration process can be described through the following steps:

i) Input data: the survey.

Istat checks whether Personal tax numbers are coherent with respect to the personal items which are used to generate them (i.e. name, date of birth, etc.). Non coherent tax codes are substituted with automatically generated codes.

ii) Input data: the administrative archives.

Tax agency data and Pensions data (Inps) are the administrative data sources engaged in the matching process. Personal tax numbers are checked and corrected following the procedure described above. Furthermore information coming from multiple records and relating to the same person is organized in order to avoid duplications.

iii) The exact matching procedure

At this step the survey and the administrative sources are matched using the Personal tax code number as the key variable. The result is a file (matched file) where information on income both from the survey and the administrative archives are provided.

iv) *Detecting and solving incoherencies on income in the matched file*.

Sometimes the survey and the administrative data source assign a different kind of income to the same person. A system of hypothesis and rules has been established in order to choose which income component must be attributed.

v) Reconciliation of incoherent income values.

Analysis of the coherence between administrative and survey data on income and formulation of hypothesis for reconciling incoherent income values.

Figure 1 - A simplified sketch of data integration process in *It-Silc*
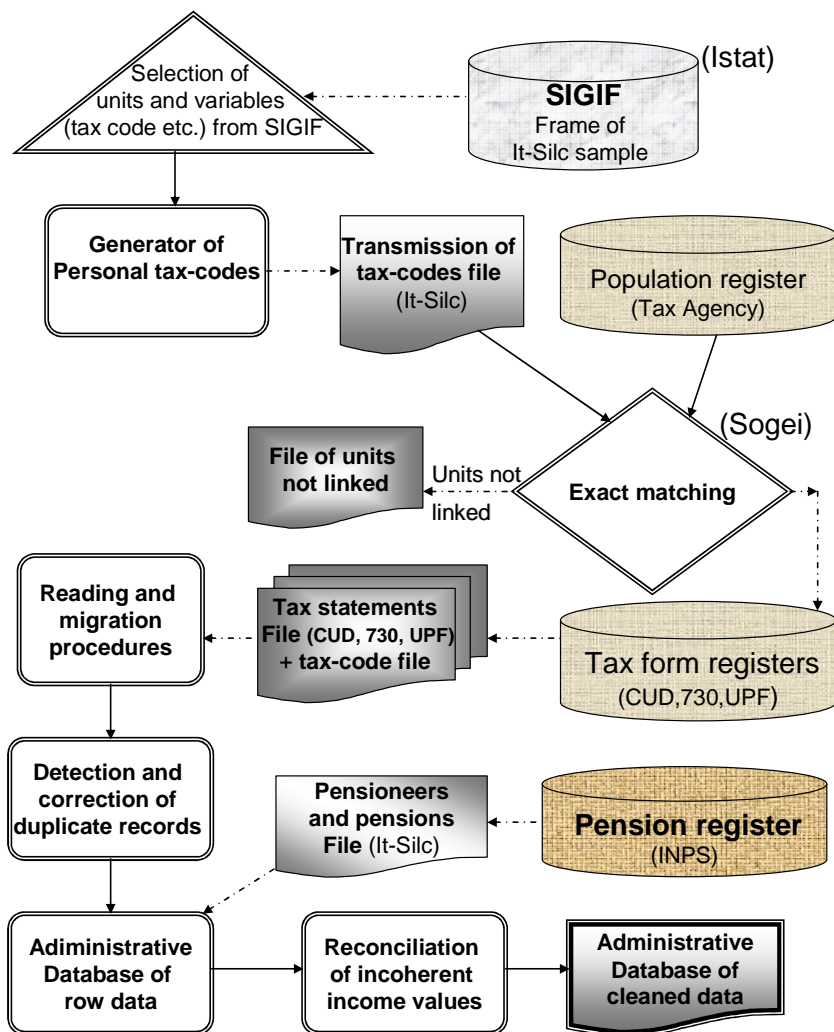
2.2. The impact on the Province of Pisa oversampling: main results

The 2008 sample size is 818 households in the province of Pisa. The units of analysis of Pisa survey, as well as It-Silc, are households and individuals (persons aged 15 and over), whilst the main variables collected concern income and basic variables (poverty, social exclusion and other living condition). In 2008 edition there are 1,984 members belonging to these households units. In particular, 676 individuals are also included in It-Silc 2008 sample frame and 1,308 represent the persons belonging to the Pisa over-sampling units. At the end of data collection 1,479 persons (aged 15 and over) are required to complete the individual questionnaire in the province of Pisa. This is the reference group for the analysis of procedures for editing/imputation, and integration of data sources.

At a first stage the It-Silc units are linked to the Population Register through the Personal tax number. For the Pisa province, the matching rate is about 98.4%. Subsequently a record linkage with the Tax statement Register (Cud, 730, Upf) is performed. Around 78.3 % of Pisa sample is successfully matched with at least one fiscal form, or, in other words, has at least one tax declaration. The rate of It-Silc interviewees that report a tax declaration is slightly lower: 76,2 %. On the other hand, who has got a tax code linked with the Italian Population Register but does not

file a tax statement is represented by 19.1 and 20.2 per cent respectively for Pisa and It-Silc. Finally, the remaining 2.6 per cent (3.6 It-Silc) is represented by the interviewees whose tax code is not matchable with Population Register.

Table 1- Main results of the linkage between tax and survey records. It-Silc and Pisa 2008

	SURVEY DATA of PISA				IT-SILC SURVEY DATA			
	Sampled	%	Interviewed	%	Sampled	%	Interviewed	%
Linked with tax codes reporting at least one tax declaration	1,392	70.2	1,158	78.3	41,546	67.1	34,139	76.2
Linked with tax codes no tax declaration	560	28.2	282	19.1	18,887	30.5	9,138	20.2
Not linked	32	1.6	39	2.6	1,506	2.4	1,528	3.6
TOTAL	1,984	100	1,479	100	61,939	100	44,805	100

2.2.1 Detecting and solving incoherencies in the matched data set

Some incoherencies stemmed from the comparison between income components in the different data sources. Generally speaking, an incoherence occurs when two or more datasets report different values on the same object (unit).

Table 2 shows the main results of the coherence analysis carried out on the matched records, both for Pisa and Italy.

Table 2 - Coherence analysis on income – It-Silc, Pisa and Tax Agency of Italy 2008.

		Did you earn self-employment or employed income, pensions or unemployment benefits in 2007?							
		SURVEY DATA of PISA				IT-SILC SURVEY DATA			
Tax Agency		YES	%	NO	%	Total	YES	%	NO	%	Total
LINKED	Income reported	1,127	93.7	12	4.3	1,139	33,178	91.6	170	2.0	33,348
	Income not reported*	6	0.5	13	4.7	19	323	0.9	468	5.4	791
	No tax declaration	45	3.7	240	87.0	285	1, 630	4.5	7,508	87.5	9,138
NOT LINKED		25	2.1	11	4.0	36	1,092	3.0	436	5.1	1,528
TOTAL		1,203	100	276	100	1,479	36,223	100	8,582	100	44,805

* A tax declaration is present but it does not report any main income component

Dark grey highlights the first type of incoherency: income is declared in the survey but not to the Tax Agency (4.2% in Pisa, 5.4% in Italy).

The second type of incoherency (Light grey) occurs when income is recorded in the Tax Agency register but not in the survey (4.3% Pisa, 2% Italy).

The strategy adopted to solve the incoherencies depends on the kind of income. Generally the administrative data source is assumed to be more reliable. By removing inconsistencies of the first and second type is possible to avoid misclassification of income components and double counting. In the following, for each income category details are provided on the impact of inconsistencies and on the method used to solve them.

Employed income

Table 3 shows the results of the linking procedure with respect to employed income. According to the survey results, in Pisa, 558 interviewed individuals have earned employed income in 2007. With respect to these individuals, the Tax Agency archive records employed income as well for around 89% (income reported) whereas it records a different category of income for 5% (income not reported). Furthermore around the 3% does not result to have submitted any tax declaration.

Table 3 - Employed income by result of the linkage and content of the two data sources. It-Silc and Pisa 2008

ADMINISTRATIVE DATA		Did you earn employed income in 2007?									
		SURVEY DATA of PISA					IT-SILC SURVEY DATA				
		YES	%	NO	%	Total	YES	%	NO	%	Total
LINKED	Employed Income reported	498	89.2	51	5.5	549	13,594	85.3	1,863	6.4	15,457
	Employed income not reported*	28	5.0	581	63.1	609	935	5.9	17,747	61.5	18,682
	No tax declaration	16	2.9	269	29.2	285	777	4.9	8,361	29.0	9,138
NOT LINKED		16	2.9	20	2.2	36	631	3.9	897	3.1	1,528
TOTAL		558	100	921	100	1,479	15,937	100	28,868	100	44,805

* A tax declaration is present but it reports a different kind of income

Inconsistencies of the first type account for about 8% whereas inconsistencies of the second type account for 5.5%. Most frequently incoherencies are due to the different kind of income recorded for the same unit. For example an individual earns employed income according to It-Silc, whereas the same individual earns self-employed income according to the fiscal data source. In order to choose the "true" kind of income to be definitely assigned, several analysis have to be carried out taking into account the professional status of the individual as it results in the survey as well as other information drawn from the fiscal data source.

Only for a small part of the matched units employed income has been reclassified to another income category, particularly only 663 units (3.7%) for It-Silc and 34 units (5.6%) for Pisa.

Self-employed income

Incoherencies of the first and second types are to be solved not contradicting the decisions taken for employed income.

Self-employed income includes the remuneration of temporary workers formally hired as independent collaborators (labeled as "co.co.co."). The difficulty to clearly distinguish co.co.co. workers from employees is one of the most relevant causes of the self-employed income incoherencies. For this reason as a first step it is essential to detect and solve incoherencies between the survey and the fiscal data separately for self-employed income (without co.co.co.) and the co.co.co. income. Table 2 shows the final results.

Table 4 - Self-employment income by result of the linkage and content of the two data sources. It-Silc and Pisa 2008

ADMINISTRATIVE DATA		Did you earn self-employment income in 2007?									
		SURVEY DATA of PISA					IT-SILC SURVEY DATA				
		YES	%	NO	%	Total	YES	%	NO	%	Total
LINKED	Employed Income reported	132	66.7	66	5.1	198	3,391	68.7	1,931	4.8	5,322
	Employed income not reported*	40	20.2	920	71.8	960	761	15.4	28,056	70.4	28,817
	No tax declaration	19	9.6	266	20.8	285	602	12.2	8,536	21.4	9,138
NOT LINKED		7	3.5	29	2.3	36	181	3.7	1,347	3.4	1,528
TOTAL		198	100	1,281	100	1,479	4,935	100	39,870	100	44,805

* A tax declaration is present but it reports a different kind of income

The coherence of survey and administrative data is sensibly lower with respect to employed income. Only about 69% (67% in Pisa) of interviewed people declaring self-employed income in the survey, record self-employed income in the administrative archive as well. Incoherencies of the first type accounts for over 27% (30% in Pisa), incoherencies of the second type for 4.8% (5% in Pisa). In order to solve incoherencies, self-employed income is assigned to other income categories for 574 units. Once the integration process is completed, 6220 sampled units result to have earned self-employed income in 2007.

Pensions.
Contrarily to employed and self-employed income, pensions cannot be affected by tax evasion. As a consequence administrative data (Inps) are always considered as the most reliable data source, survey being taken into account only for the not-matched units. In It-silc 2008, 14.987 sampled units (479 in Pisa) receive pensions, of which 479 in the province of Pisa. Over 95% of pensions income (over 96% in Pisa) comes from the administrative data source.

2.2.2 Impact of integration/imputation on income estimates

In the previous section we have synthetically described the procedure adopted by Istat to solve incoherencies between survey and administrative data. In a few cases income is reclassified to a different category (for example from self-employed to employed income).
In the following step income is estimated through a reconciling process between the survey and the administrative data source estimates. This requires obviously a preliminary harmonisation aimed at establishing how to obtain Eu-Silc income categories moving from the income-related items of the Tax agency and Inps archives.
For what concerns employed income, the fiscal value is considered the "true" value unless the survey records a greater value. In this last case an in-depth analysis is carried out. The income value is finally estimated taking into account the results of the analysis as well as the amount of the discrepancy between the survey and administrative values.
Self-employed income values are generally supplied by the data source which records the largest amount of *net* self-employed income. It is worth recalling that Istat is required to estimate both net and gross income. The choice of the best data source is based on the following considerations:
    - Tax returns usually contain exact information on taxable incomes and tax liabilities. They also provide information on social security contributions. Therefore, tax records can be used to measure the net taxable income. In general, neither taxable income is identical to gross income, nor net taxable income is identical to disposable income (see figure 5). In principle, if the deductions from profits are available to the company owners for their personal use, then they should be considered as components of both the gross and the disposable personal incomes. However, not all the tax

abatements allowed are explicitly shown in the tax returns. By definition, tax evasion is also not available in the tax files.

- Survey data on self-employment income may be affected by under-reporting to an unknown extent. Moreover, gross income is usually unknown by the interviewees and the collection of the additional information needed to compute it (taxes and social security contributions) puts an excessive response burden on the respondents. Survey data should therefore be integrated by external sources and/or microsimulations. In both cases, the amount of taxes and contributions could be added to the net income reported in the survey to get a 'survey based' measure of gross income. If taxes and contributions are correctly measured, the result of such an addition gives gross income net of survey under-reporting (this should be kept in mind when assessing the international comparability of the data). Therefore, the 'survey based' measure of gross income would be equal to taxable income only if survey under-reporting equals the sum of tax avoidance and of tax evasion.

Table 5 - Employed income: income by groups of units generated by the matching procedure – It-Silc and Pisa 2008

SOURCE OF DATA			Inclusion/exclusion[*] of records on the employed income database					
			Included			Excluded		Total
			Nr.	%	Mean[**] income	Nr.	%	Nr.
			It-Silc					
LINKED	a.	Employed Income reported both in Survey and Tax data	13,582	79.4	17,487	12	1.7	13,594
	b.	Employed Income reported in Survey but not in Tax data	520	3.0	13,956	415	60.3	935
	c.	Employed Income reported in Survey, No Tax declaration	711	4.2	11,846	66	9.6	777
	d.	Employed Income reported in Tax data but not reported in Survey	1,681	9.8	7,753	182	26.5	1,863
NOT LINKED	e.	Employed income reported in the Survey only	618	3.6	16,100	13	1.9	631
TOTAL			17,112	100	16,139	688	100	17,800
			Pisa					
LINKED	a.	Employed Income reported both in Survey and Tax data	498	86.8	17,725	-	-	498
	b.	Employed Income reported in Survey but not in Tax data	10	1.7	17,121	18	51.4	28
	c.	Employed Income reported in Survey, No Tax declaration	14	2.4	8,962	2	5.7	16
	d.	Employed Income reported in Tax data but not reported in Survey	37	6.5	7,069	14	40.0	51
NOT LINKED	e.	Employed income reported in the Survey only	15	2.6	13,502	1	2.9	16
TOTAL			574	100	16,704	35	100	609

[*] It concerns the decision to include (or exclude) the contribution of a specific record on (from) the integrated database of employed income receivers
[**] Not weighted

Tables 5 and 6 analyse the contribution of a specific record on the employed/self-employed database by source of data. The first distinction is between matched (or linked) and not matched (or not linked) units. Matched units are further grouped according to the data source from which the income value is taken:

Group *a*: income is estimated on the basis of information recorded by both the survey and the administrative data source;

Groups *a* and *b*: income is estimated exclusively on the basis of survey data;

Group *d*: income is estimated on the basis of administrative data only.

The second distinction concerns the decision to include/exclude a specific record on/from the integrated database of employed or self-employed income receivers.

Income is on average higher for the *a* group both for employed and self-employed income. Furthermore, income estimated on the basis of survey data only is higher than income estimated using exclusively fiscal data. Cases excluded from employed income database are mainly referred to incomes reported in survey but not in tax data (60.3 per cent). In other words, tax registers seem to be a more reliable source on employed income than survey data. Cases excluded from self–employed database are equally represented by incomes reported in survey but not in tax data or vice versa.

Table 6 - Self-employed income: income by groups of units generated by the matching procedure-It-Silc and Pisa 2008

SOURCE OF DATA		Inclusion/exclusion of records on the self-employed income database					
		Included			Excluded		Total
		Nr.	%	Mean[*] income	Nr.	%	Nr.
		It-Silc					
LINKED	a. Self-employed Income reported both in Survey and Tax data	3,383	54.4	24,767	8	1.3	3,391
	b. Self-employed Income reported in Survey but not in Tax data	491	7.9	11,783	270	41.8	761
	c. Self-employed Income reported in Survey, No Tax declaration	526	8.4	12,682	76	11.8	602
	d. Self-employed Income reported in Tax data but not reported in Survey	1660	26.7	10,160	271	41.9	1,931
NOT LINKED	e. Self-employed income reported in the Survey only	160	2.6	16,990	21	3.2	181
TOTAL		6,220	100	18,622	646	100	6,866
		Pisa					
LINKED	a. Self-employed income reported both in Survey and Tax data	131	57.7	23,956	1	2.7	132
	b. Self-employed income reported in Survey but not reported in Tax data	23	10.1	9,278	17	46.0	40
	c. Self-employed income reported in Survey with No Tax declaration	16	7.1	10,362	3	8.1	19
	d. Self-employed income reported in Tax data but not reported in Survey	51	22.5	14,215	15	40.5	66
NOT LINKED	e. Only reported in Survey	6	2.6	16,542	1	2.7	7
TOTAL		227	100	19,126	37	100	264

[*] Not weighted

Table 7 analyses the impact of imputation on employed income target variable.
The no-response rate is low both for It-Silc (3.9%) and Pisa (2.5%). Outliers and not reliable values concern a limited number of cases: 4.5% in It-sic and 4.7% in Pisa. Anyhow, the imputation for missing or unreliable values has a valuable impact on the yearly average income. In fact average income is 1200 euro lower (2400 euros in Pisa) after imputation.

Table 7 - The impact of imputation on employed income target variable - It-Silc and Pisa 2008

	Valid	Imputed by ADM* because not reliable or outlier		Imputed by IveWare because missing	Imputed by ADM because missing	Total	
		Before	after			Before imputation	After imputation
It-Silc							
Number of cases	14,118	701		161	451	14,819	15,431
% (after imputation)	*91.5*	*4.5*		*1.0*	*2.9*	*96.0*	*100*
Mean of employed income	16,222	38,805	13,959	13,398	14,342	17,290	16,034
Pisa							
Number of cases	499	25		3	10	524	537
% (after imputation)	*92.9*	*4.7*		*0.6*	*1.9*	*97.6*	*100*
Mean of employed income	16,279	63,388	14,474	14311	14545	18,526	16,151

* ADM stands for Administrative data source

Table 8 - The impact of imputation on self-employed income target variable - It-Silc and Pisa 2008

	Valid	Imputed because not reliable or outlier		Imputed because missing*	Total	
		before	after		Before imputation	After imputation
It-Silc						
Number of cases	3,515	449		885	3,963	4,848
% (after imputation)	*72.5*	*9.3*		*18.2*	*81.8*	*100*
Mean of self-empl. income	18,369	15,227	16,570	12,413	18,013	17,116
Pisa						
Number of cases	128	19		42	147	189
% (after imputation)	*67.7*	*10.1*		*22.2*	*77.8*	*100*
Mean of self-empl. income	16,684	17,536	20,518	13,510	16,794	16,364

* As a result of the integration process 288 units of the It-Silc (13 of Pisa survey) have been reclassified as self-employed and their incomes have been imputed.

Tables 9 and 10 present the main results of combining administrative and survey data. The final (merged) data can be divided into four distinct groups on the basis of the original source of income: 1) income is present solely in the tax archive; 2) income is present only in the survey; 3) income comes from the tax records, being greater than the corresponding survey income; 4) income comes from survey since the value is greater than in the administrative source. For each group, the table displays the annual mean income and the number and the percentage of percipients.

As results from data (table 9), the number of employed income receivers increases of about 11% in It-silc (7% in Pisa) whereas employed income increases of about 0.7% (3.4% in Pisa).

Table 9 - Employed income: the main results of the data source integration - It-Silc and Pisa 2008

	Units with employed income in only one data-source		Units with employed income in both data-sources		Total	
	Only tax	Only survey	Tax≥ Survey	Tax <Survey	before integration	after integration
It-Silc						
Number of cases	1,681	1,849	8,661	4,921	15,431	17,112
% (after integration)	*9.8*	*10.8*	*50.6*	*28.8*	*90.2*	*100*
Employed income (yearly mean)	7,753	13,861	17,448	17,555	16,034	16,139
Pisa						
Number of cases	37	39	320	178	537	574
% (after integration)	*6.4*	*6.8*	*55.7*	*31.0*	*93.6*	*100*
Employed income (yearly mean)	7,069	12,800	17,513	18,107	16,151	16,704

Coming to self-employed income (table 10) we observe that the 1,372 records coming exclusively from the tax file reduce the overall mean income. In effect, the mean of self-employed income present solely in the tax records is lower (9,895 euros) than the average of all survey incomes after editing and imputation (17,116 euros). For 1,177 percipients, self-employment income is reported solely in the survey. The mean income of this group is 12,893 euros, again a value lower than the average income computed on the whole of the survey data. As already noticed, the majority of this group is made of taxpayers who filed a tax return without reporting self-employment incomes or did not file a tax return. Among the group of percipients who have reported self-employment income in both sources, 1,712 display an higher amount in the tax data, whilst 1,959 persons reported a larger income in the survey data. For what concerns the former group, mean income of the (selected) tax data (26,513 euro) is approximately twice that of the (discarded) survey incomes of the same group (15,593 euro). Similarly, among the group of percipients with higher survey incomes, the mean of the (selected) survey income (21,280 euros) is nearly twice as great as the mean income of the (discarded) net taxable incomes (13,764 euros).

Table 10 - Self-employed income: the main results of the data source integration - It-Silc and Pisa 2008

	Units with S.E. income in only one data-source*		Units with S.E. income in both data-sources*		Total	
	Only tax	Only survey	Tax≥ Survey	Tax <Survey	before integration	after integration
			It-Silc			
Number of cases	1,372	1,177	1,712	1,959	4,848	6,220
% (after integration)	*22.1*	*18.9*	*27.5*	*31.5*	*77.9*	*100*
Mean of self-empl. income	9,895	12,893	26,513	21,280	17,116	18,622
			Pisa			
Number of cases	38	45	68	76	189	227
% (after integration)	16.7	19.8	30.0	33.5	83.3	*100*
Mean of self-empl. income	9,519	10,632	29,492	19,683	16,364	19,126

(*) As a result of the integration process 288 units of the It-Silc (13 of Pisa survey) have been reclassified as self-employed and their incomes have been imputed

Merging administrative and survey data brings about a rise of 28.3 % in the number of percipients and an increase of 8.8 % in the average of self-employment income compared to the exclusive use of survey data. When both sources report information on self-employment incomes, there is some evidence of a higher under-estimation rate on the tax data compared to the survey data.

3. The territory of the Pisa Province: an introduction to its demographic and social profile

UROPS, Province of Pisa

3.1. The division in Health Societies

Within the Pisa Province there are 39 municipalities grouped into 5 Health Societies (HS).
The Health Societies are public organisations which have the function of social and sanitary planning and management at supra-municipal level. They represent one of the most important innovations in the integration of health programs and social assistance. Their role is to encourage

the involvement of local communities, social parties, third sector associations, and volunteers in identifying health-related needs (according to the WHO definition) in the planning process. For the HS, local action is an essential element in the regional strategy to promote health. It is not limited just to social and health issues, but also includes the improvement of health through an integrated inter-sectorial policy capable of influencing the determining factors affecting the health of populations and the quality of the environment.

The Health Societies of the province of Pisa are:

- AREA PISANA, which includes 9 municipalities: Calci, Cascina, Fauglia, Lorenzana, Orciano pisano, Pisa, San Giuliano Terme, Vecchiano, Vicopisano.
- VALDERA, which includes 15 municipalities: Bientina, Buti, Calcinaia, Capannoli, Casciana Terme, Chianni, Crespina, Lajatico, Lari, Palaia, Peccioli, Ponsacco, Pontedera, Santa Maria a Monte, Terricciola.
- ALTA VAL DI CECINA, which includes 4 municipalities: Castelnuovo Val di Cecina, Montecatini Val di cecina, Pomarance, Volterra.
- BASSA VAL DI CECINA, which includes 6 Municipalities: Casale Marittimo, Castellina Marittima, Guardistallo, Montescudaio, Ribarbella, Santa Luce.
- VALDARNO INFERIORE, which includes 4 Municipalities Castelfranco di Sotto, Montopoli Valdarno, Santa Croce sull'Arno, San Miniato.

Some particularities:

- Some municipalities of Bassa Val di Cecina are under the competence of another Province (Livorno).
- In the Pisa Province there is a little Municipality (Monteverdi Marittimo) which is under the competence of HS Val di Cornia (Livorno Province)
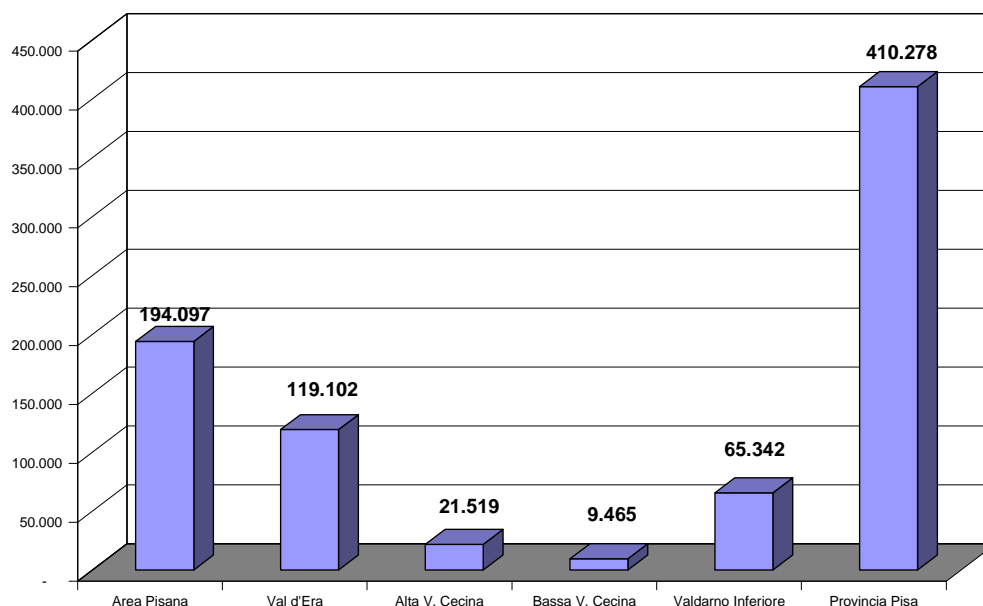
3.2. Population

The Province's population is concentrated in the Area Pisana (47,3%), which is the area with the higher density of population (352,3 inhabitants per kmq). In this Health Society there is the municipality of Pisa, the provincial capital, which hosts an important University.

The second area is Valdera (29,02% of total province's population and a density of 184,1 inhabitants per kmq).

The third area is Valdarno Inferiore (15,9% of total province's population and a density of 330,4 abitanti per kmq), an industrial area with important industrial tanneries.

The last two areas are: Alta Val di Cecina (5,2% of total population) and Bassa Val di Cecina (2,3%). In these areas the population is distributed in many hamlet apart from each other. The density of population is around 40 inhabitants per kmq.

Graph 1. Population of Pisa province by Health's Societies. Year 2008

In terms of population growth the most active areas are Bassa Val di Cecina, which had a crude rate of increase (2007/2008) of +25,8‰ and Valdera (+18,3‰). Alta Val di Cecina is the only area with a crude rate of total increase of -6,5‰. In this area we register the population aging and also the depopulation of little towns.

In the Pisa province, the population growth is due exclusively to immigrants components: the crude rate of natural increase is -1,7‰ and the crude rate of migration increase is 12,5‰.

3.3. Foreign citizens

At the end of 2008, the number of foreign citizens in the Pisa province was 30.524, which is 7,4 per cent of the population. From 2007 to 2008 the increase was 3.702 foreign citizens (+14%).

The area with the higher percentage of foreign citizens is Valdarno Inferiore (9%), followed by Bassa Val di Cecina (7,4%), Area Pisana (7,2%) and Alta Val di Cecina e Valdera (7,1%).

The higher growth between 2007 and 2008 was in Valdera (15% circa); the lower in Alta Val di Cecina (7,74%),

4. Social and sanitary indicators by Health Societies

UROPS, Province of Pisa

The Pisa Province (UROPS) is part of the Regional Network of Social Observatory (RNSO) coordinated by the Regional Social Observatory. One of the aims of the Regional Network was to create a regional common set of indicators in social and sanitary policies and a common methodology for the data acquisition (see "Sample Periodic Report", pag 14 and Annex 1). These indicators will be used to monitor and to evaluate the performance of Health's Societies. After a long work UROPS created two groups of indicators: a minimum set (155 indicators) and a complementary set (151).

The minimum set must be updated every year, the updating of the complementary set is not mandatory. The indicators are calculated at supra municipal level (Health's Societies).

The institutions involved in data collected and data updating are: Social Observatories, Regional Health Agency, Tuscany Region and Zancan Foundation (a private foundation that collaborates with the Tuscany Region).

The indicators are divided into11 sections:

1. Demographic profile;
2. Determinants of Health Status (social determinants like income, labour data, etc, and sanitary determinants);
3. Health Status (death rate, travel accidents, hospitalization, diseases);
4. Territorial Assistance: essential level of territorial health care (indicators about sanitary services);
5. Territorial Assistance: social area (indicators about social services);
6. Social and Sanitary Assistance: elderly persons (health indicators, socio-demographic indicators, not self sufficiency);
7. Social and Sanitary Assistance: families and youngsters;
8. Social and Sanitary Assistance : immigrates;
9. Social and Sanitary Assistance : disability;
10. Social and Sanitary Assistance: mental health;
11. Social and Sanitary Assistance: dependences.

In March 2010, for the first time, UROPS started up data collection and gathered the first results. In the next tables it will be described the territorial profile that results from the collection of some of these indicators. We have selected only the indicators that are more useful in order to compare the results with data coming from the EU SILC 2008 oversampling.

4.1. Indicators for demographic profile

Next table reports some indicators about the demographic profile of Health Societies. It results that the area with a larger ageing population is Alta Val di Cecina (ageing ratio is 274,9 while provincial average is 176,5). The areas with lower ageing population are Valdarno Inferiore e Valdera (respectively 150,4 and 155,2).

Within the Minimum Set there are many other indicators about population ageing because it is a phenomenon that produces important effects in social policies, especially concerning the public social expenditure (see next paragraph).

Table 1: Main demographic index by Health's Societies. 31/12/2008

	Ageing ratio	Age dependency ratio	Indice di struttura pop. attiva	Indice di ricambio pop. attiva
Health Societies				
Area Pisana	190,2	54,4	125,1	173,6
Val d'Era	155,2	52,4	108,0	147,3
Alta Val di Cecina	274,9	63,5	141,0	179,2
Bassa Val di Cecina	199,4	52,6	122,8	177,3
Valdarno Inferiore	150,4	52,2	110,0	134,8
Total (Province)	**176,5**	**53,9**	**118,0**	**159,5**

Ageing ratio: of population aged 65 and over to population aged 0-14, per 100
Age dependency ratio: 1st variant (Population aged 0-14 and 65 and more to pop. aged 15-64)

In table 2 there is another important indicator (average number of people per household). Within the Pisa province the average number is 2,4 people for every household (we go from the 2,2 of Bassa Val di Cecina to the 2,6 of Valdarno Inferiore).

Table 2 - Average number of people per household, by Health Society. Year 2008

	Average number of people
Area Pisana	2,3
Valdera	2,5
Alta Val di Cecina	2,3
Bassa Val di Cecina	2,2
Valdarno Inferiore	2,6
Total (Province)	**2,4**

4.2. Determinants of Health Status

4.2.1. Income and pensions indicators

Within the Minimum Set there are many indicators concerning the determinants of Health Status. Health Status is the result of the incidence of many factors (income, job, housing, environment, etc.) that produce a direct or indirect impact on this status. In Minimum and Complementary sets there are many indicators about these determinants. Next tables report the main ones.

In table 3 and graph 1 we can see that the areas with the higher level of mean taxable income, in 2007, were Area Pisana (more then provincial level) and Valdarno Inferiore. We can also see the tendencies in the last three years.

Table 3 - Mean Taxable Income by Health Society. Years: 2005-2006-2007

	Mean Taxable Income 2005	Mean Taxable Income 2006	Mean Taxable Income 2007
Area Pisana	21.905,3	22.660,4	23.427,4
Val d'Era	18.096,3	18.954,0	20.517,7
Alta Val di Cecina	18.722,3	19.218,9	20.353,5
Bassa Val di Cecina	17.188,3	18.289,8	19.781,7
Valdarno Inferiore	19.020,6	19.806,1	21.323,0
Total (Province)	**20.131,1**	**20.894,1**	**22.031,2**

Mean Taxable Income=Total Taxable Income/Number of taxpayers
Source: Minimum Set of Indicators– Section 2 Determinants of Health

Graph 1 Evolution of mean taxable income. Year 2005-2007
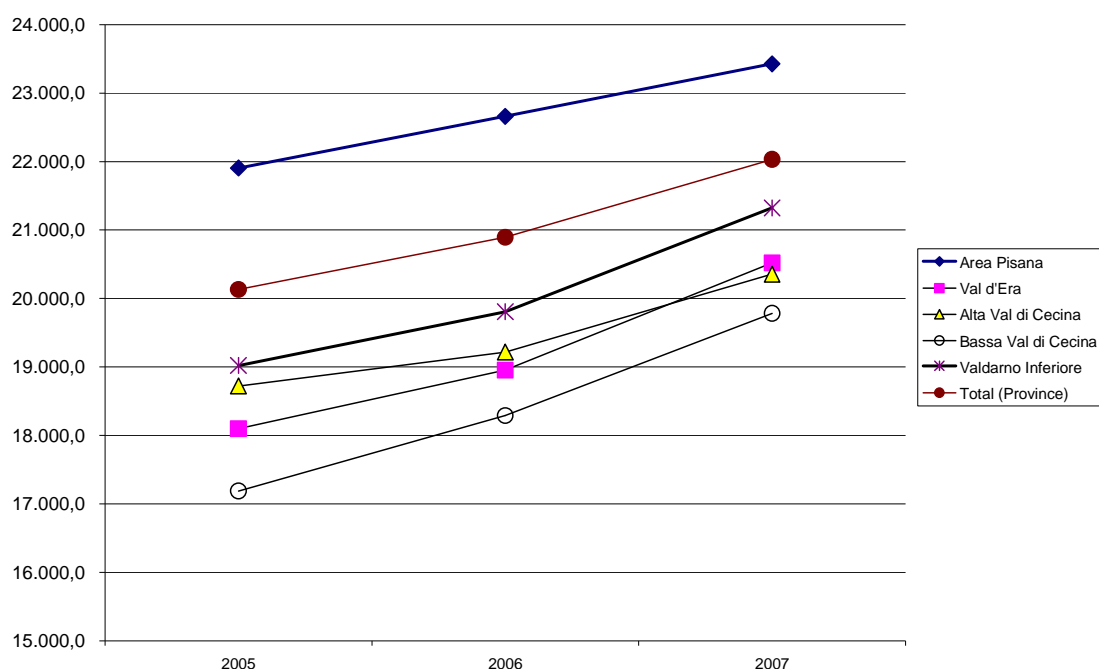
In our Social Security System there is a type of pension assigned to people without a minimum income. The amount of this pension is around 350,00 euro. Next table reports the number of pensions to population over 65 in the five areas. Table 5 reports the mean monthly amount of other pensions: retirement, disability, survivor. Here we can see that the area with the higher level of retirement pensions is Alta Val di Cecina.

Table 4 – Proportion of Pensions supplemented by guaranteed minimum income to population over 65. Year 2008

	Pensions/population over 65
Area Pisana	5,0
Val d'Era	4,4
Alta Val di Cecina	3,0
Bassa Val di Cecina	4,3
Valdarno Inferiore	4,1
Total (Province)	**4,5**

tax pensions supplemented by guaranteed minimum income=pensions/population over 65 years old
Source: Minimum Set of Indicators– Section 2 Determinants of Health

Table 5 –Mean amount of pensions by typology and by Health Society. Monthly amount. Year 2008

	Retirement	Disability	Survivor	Guaranteed minimum income	Total pensions
Area Pisana	974,0	566,9	540,4	356,7	736,7
Val d'Era	874,9	529,7	499,8	317,3	696,8
Alta Val di Cecina	1.033,0	519,4	582,3	332,6	800,8
Bassa Val di Cecina	860,9	580,6	502,2	491,0	695,2
Valdarno Inferiore	851,0	532,8	479,6	359,7	685,6

Source: Minimum Set of Indicators– Section 2 Determinants of Health

4.3. Housing indicators

Tuscany Region every year grants contribution for households that cannot afford to pay a rent. Not all requests may be accepted because the amounts are limited but we do know the number of families that make a request. This may be a good indicator to measure housing problems.

Next table reports this indicator and here we can see that the higher level is in Area Pisana and Valdera.

Table 6 - Requests for rent contribution (1000 households) by Health Societies. Year 2008

Comune	Requests/households*1000
Area Pisana	21,7
Val d'Era	20,6
Alta Val di Cecina	11,0
Bassa Val di Cecina	10,1
Valdarno Inferiore	16,0
Total (Province)	**19,7**

Source: Minimum Set of Indicators– Section 2 Determinants of Health

5. Indicators about local policies

UROPS, Province of Pisa

In order to comprehend the health status of a territory it is also important to know and to measure the policies against social problems realised by local institutions.

Next table and graph report important data about the distribution of the public social expenditure: households and young people, disability, dependences, elders (hospices and residential assistance), immigrates, poverty and different users. How we can see from the table and from the graph, the expenses for elders people absorb the majority of public resources.

The expenditure for poverty is about 2.692.227 at provincial level which correspond to the 5% of the total expenditure.

About this there are some differentiations at local level: in Valdarno Inferiore it is the 12% of the total, in Valdera it is the 0,7%. We have to underline that in Valdarno there isn't the expenditure for elders hospices that is the higher one in the other areas.

Table 1 Distribution of public social expenditure for social. Year 2007

	Area Pisana		Valdera		Alta Val di Cecina		Bassa V. di Cecina		Valdarno Inferiore	
	Expendit..	**%**	**Expendit**	**%**	**Expendit**	**%**	**Expendit**	**%**	**Expendit**	**%**
Households and young people	2.643.775	8,0	944.991	13,1	217.368	4,6	1.056.082	21,1	924.974	22,4
Disability	4.855.466	14,7	2.536.735	35,2	1.095.384	23,3	1.465.001	29,3	1.209.256	29,2
Dependences	150.222	0,5	73.149	1,0	53.043	1,1	47.965	1,0	9.379	0,2
Elders (hospices)	16.564.934	50,0	1.973.798	27,4	2.051.728	43,7	315.823	6,3		
Elders (residential assistance)	1.424.874	4,3	164.198	2,3	144.759	3,1	522.562	10,5	169.273	4,1
Elders (other)	2.050.080	6,2	1.117.126	15,5	713.323	15,2	1.162.313	23,3	951.218	23,0
Immigrates	1.670.651	5,0	61.000	0,8	58.686	1,2	131.641	2,6	46.383	1,1
Poverty	1.759.120	5,3	49.938	0,7	113.487	2,4	274.345	5,5	495.337	12,0

	Area Pisana		Valdera		Alta Val di Cecina		Bassa V. di Cecina		Valdarno Inferiore	
	Expendit.	**%**	**Expendit.**	**%**	**Expendit.**	**%**	**Expendit.**	**%**	**Expendit.**	**%**
Different users	1.988.439	6,0	276.010	3,8	252.375	5,4	20.000	0,4	331.469	8,0
Total expenditure	**33.107.561**	**100,0**	**7.196.945**	**100,0**	**4.700.153**	**100,0**	**4.995.732**	**100,0**	**4.137.289**	**100,0**

Graph 1 Distribution of public social expenditure. Total Province. Year 2007
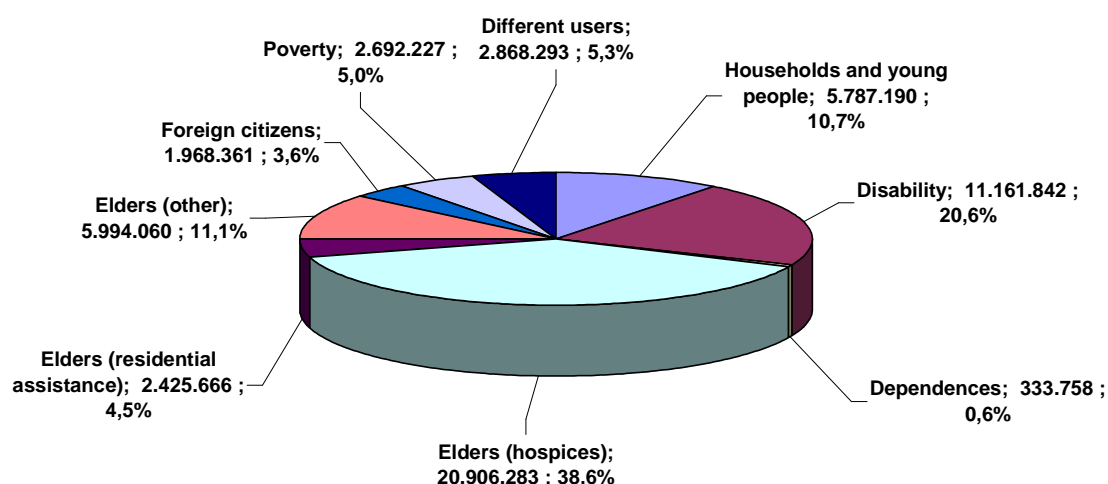
6. *Direct estimates using oversampling data*

Caterina Giusti, UNIPI-DSMAE

In this section we present and comment the direct estimates computed using data coming form the EU SILC 2008 oversampling for the Province of Pisa. The estimates comprehend some income and some economic discomfort indicators, such as the mean and median equivalised household income in 2007, the rate of households declaring to be unable to face unexpected financial expenses, and so on. All these estimates are computed at provincial level, that is for the Province of Pisa, and also at a finer geographical level, namely for the five "Società della Salute" or "Health Societies" of the Province of Pisa. To be noted that for the moment it is not possible to compute the estimates of interest also for the Tuscany Region and for the entire Italian territory using data coming from the same survey. It will be possible to compute also these estimates, comparing them with those regarding the Province of Pisa presented in this report, only when all the data coming from EU SILC 2008 will be available.

In order to compute the direct estimates, the household weights have been calibrated on the distribution of the households in the domains of interest, the Province and the Health Societies.

As concerns the variance of the direct estimates, we have considered different approaches: a simpler approach based on the linear approximation of the estimators' variances, and a more refined approach based on resampling techniques (see section 4 of SAMPLE Consortium - Deliverable 7.1). Since the variances computed using the different techniques resulted to be equivalent, the variances presented in this section are those computed under the first approach.

Table 1 – Mean of the equivalised household income in Euros, by Health Society and household characteristics, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	18279,76	1149,43
Valdera	19406,27	899,54
Alta Val di Cecina	17744,79	979,5
Bassa Val di Cecina	15205,46	927,29
Valdarno inferiore	20579,19	2140,38
Educational level of the head of the household		
Primary school	15016,19	656,72
Secondary school (10-13 years)	19142,39	1129,01
Secondary school (14-19 years)	20013,59	762,58
University degree or Phd	28194,79	4265,77
Number of household members		
1	17242,12	1316,12
2	18298,42	760,67
3	22361,31	2373,97
4	18681,26	933,25
5 or more	15859,40	1202,94
Gender of the head of the household		
Male	19566,06	809,97
Female	15475,41	1152,84
Total (Province of Pisa)	18819,62	695,85

The mean of the equivalised household income, according to the Eurostat guidelines (see the Manual UDB variables description version 2007-1 from 01-03-09), was equal to 18820 Euros, in the Province of Pisa in 2007 (Table 1). If we consider some income quantiles (Table 2) we can see that the 20% of the households in the Province had an equivalised income under 11000 Euros, the 50% under 16707 Euros (median value) and the 80% under 23576 Euros.

Considering the same estimates for the five Health Societies in the Province we can see that for "Area Pisana", "Valdera" and "Valdarno inferiore" the mean and percentiles estimates are in line with those at provincial level, especially taking into account the estimated standard errors. On the contrary the household equivalised income estimates obtained for "Alta Val di Cecina", and in particular for the "Bassa Val di Cecina" Health Societies, seem to be slightly lower with respect to the provincial ones.

Table 2 – 20th percentile, median and 80th percentile of the equivalised household income in Euros, by Health Society, year 2007.

	Percentile	Estimate	Standard Error
Health Society			
Area pisana	20th	11162,38	828,77
	Median	16900,04	757,99
	80th	22434,01	834,02
Valdera	20th	10703,25	882,18
	Median	17486,95	862,32
	80th	24761,65	1203,7
Alta Val di Cecina	20th	10502,11	687,13
	Median	15553,79	1039,58
	80th	23678,66	1408,33
Bassa Val di Cecina	20th	9203,92	664,46
	Median	13453,65	638,26
	80th	19793,88	1416,62
Valdarno inferiore	20th	11044,17	898,89
	Median	16027,63	1093,59
	80th	25157,04	1552,71
Total (Province)	20th	10981,23	416,76
	Median	16706,96	564,74
	80th	23575,20	561,91

As far as concerns some of the main household characteristics, we can see form Table 1 and 3 that in the Province of Pisa an increasing level of education of the head of the household corresponds to an higher level of household income, both in terms of mean and percentile estimates.

The estimated mean household income for the Province is also differentiated when we consider the gender of the head of the household: where the head is a male, the estimated mean household income is significantly higher, around 19500 Euros, with respect to the nearly 15500 Euros estimated for the households where the head is a female. This gap is confirmed also in terms of estimated income distribution values (see Table 3): the 50% of the households with a female as the head had in 2007 an income under 13400 Euros, against the 17600 Euros obtained for the households with a male head. Note that the estimated standard errors allow to consider these differences as significant ones.

Finally, the number of household members does not suggest any particular conclusion with respect to the estimated values, especially when considering the estimated standard errors.

Table 3 – 20th percentile, median and 80th percentile of the equivalised household income in Euros, by household characteristics, year 2007.

	Percentile	Estimate	Standard Error
Educational level of the head of the household			
Primary school	20th	8764,26	489,21
	Median	13657,67	642,04
	80th	18897,27	891,34
Secondary school (10-13 years)	20th	11542,13	649,31
	Median	17434,58	990,33
	80th	23598,69	1237,6

	Percentile	Estimate	Standard Error
Educational level of the head of the household			
Secondary school (14-19 years)	20th	14349,86	894,08
	Median	19347,87	566,24
	80th	24082,51	1022,11
University degree or Phd	20th	14013,01	1345,65
	Median	22281,99	1836,88
	80th	38037,88	4192,58
Number of household members			
1	20th	8537,47	533,65
	Median	14037,73	982,91
	80th	20544,66	1042,38
2	20th	11812,46	641,22
	Median	16417,14	854,89
	80th	24037,62	1085,37
3	20th	12865,34	1344,4
	Median	18992,77	1095,79
	80th	25143,84	1193,8
4	20th	11399,38	1194,66
	Median	18608,41	978,79
	80th	23103,75	1282
5 or more	20th	10171,4	2136,37
	Median	14604,77	1699,31
	80th	20386,54	2666,58
Gender of the head of the household			
Male	20th	11465,01	503,32
	Median	17569,58	475,39
	80th	24030,44	513,87
Female	20th	9077,05	828,68
	Median	13381,56	704,76
	80th	20917,72	2268,36
Total (Province)	20th	10981,23	416,76
	Median	16706,96	564,74
	80th	23575,2	561,91

Table 4 reports the estimated values of the Head Count Ratio (HCR) or *at-risk-of-poverty-rate*, that is of the percentage of households with an equivalised income under the poverty line. If we consider as poverty line the 60% of the median equivalised income in the Province of Pisa in 2007 (10024,18 Euros), we obtain for the Province an HCR equal to 15,83%. The HCRs estimated for four out of the five Health Societies in the Province are very similar to the provincial estimate, even if characterised by higher estimated standard errors. The only exception is the HCR estimated for the "Bassa Val di Cecina": the value is over the 24% while the corresponding standard error is only slightly higher than the others. To better investigate this gap we also computed the HCRs using as poverty line the 60% of the median equivalised income in the Tuscany Region in 2006 (EU SILC survey 2007). The results, not reported here, are essentially the same of those in Table 4. Thus, we can say that considering the HCR and the household equivalised income estimates, the "Bassa Val di Cecina" Health Society represents a critical area in the Province.

Table 4 – Head Count Ratio (%) by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	15,52	3,05
Valdera	15,46	3,08
Alta Val di Cecina	15,94	4,16
Bassa Val di Cecina	24,77	4,78
Valdarno inferiore	15,87	3,54
Total (Province)	15,83	1,82

Table 5 reports the estimates of the proportion of households with the house owned (without a mortgage) by one or more household members. As we can see, a higher proportion of households with house owned by household members do not always correspond to a higher household income in the area. For example, the relative high mean and household income estimates of the "Area pisana" Health Society correspond to a relative low proportion of households with ownership of the house; note however that the standard error of the "Area pisana" estimate in table 5 do not allow to consider this estimate as significantly different from the mean provincial value (65.8%) .

Table 5 – Proportion of households with house owned by household members (without a mortgage), by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	61,17	4,17
Valdera	75,38	3,44
Alta Val di Cecina	77,43	4,57
Bassa Val di Cecina	68,34	5,14
Valdarno inferiore	58,56	4,60
Total (Province)	65,84	2,41

As concerns some of the main EU SILC economic discomfort indicators, Table 6 highlights that the "Area pisana" is characterized by a high proportion of households with arrears on utility bills, around the 17% against a provincial value of 12,6%. Lower estimated proportions characterize the remaining Health Societies, with the exception of the 11,3% of the "Alta Val di Cecina".

Table 6 – Proportion of households with arrears on utility bills, by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	16,93	3,12
Valdera	7,96	2,09
Alta Val di Cecina	11,33	3,79
Bassa Val di Cecina	7,77	3,16
Valdarno inferiore	8,18	2,50
Total (Province)	12,59	1,71

Also in terms of ability to make ends meet (Table 7) and to afford to go for a week's annual holiday away from home (Table 8), the "Area pisana" is characterized by the higher estimate, that is by the higher expressed discomfort. On the opposite, the "Alta Val di Cecina" is characterized by the lower estimates. Note that the proportions of households that cannot afford to go away from home for a week's annual holiday is particularly high for all the areas, with a provincial estimate equal to nearly 41%.

Table 7 – Proportion of households able to make ends meet with great difficulty, by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	18,41	3,36
Valdera	13,49	2,95
Alta Val di Cecina	12,11	3,68
Bassa Val di Cecina	13,01	3,89
Valdarno inferiore	16,55	3,66
Total (Province)	16,27	1,94

Table 8 – Proportion of households that cannot afford to go for a week's annual holiday away from home, by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	44,55	4,23
Valdera	36,29	4,01
Alta Val di Cecina	33,25	5,18
Bassa Val di Cecina	40,52	5,32
Valdarno inferiore	40,47	4,57
Total (Province)	40,91	2,48

The 15,89% of the households in the Province declared to not be able to keep home adequately warm (Table 9). For what concerns the Health Societies, we can see that the estimated proportions are differentiated, even if the estimated standard errors suggest to be cautious in this discrimination process. Anyway, in this case the higher level of discomfort is expressed by households living in the area "Valdarno inferiore" and "Area pisana", while "Bassa Val di Cecina" has the lower estimate.

Table 9 – Proportion of households not able to keep home adequately warm, by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	17,60	3,59
Valdera	11,56	2,75
Alta Val di Cecina	9,52	3,25
Bassa Val di Cecina	7,93	3,05
Valdarno inferiore	22,23	4,06
Total (Province)	15,89	2,04

Another indicator of economic discomfort expressed by the interviewed households is represented by the ability to face unexpected financial expenses, corresponding in Italy for EU-SILC 2008 to 750 Euros. As we can see from Table 10, the estimated proportion of households that declared to be unable to face these expenses is equal to 31% in the Province. Note that in this case a relative low proportion characterizes the two Health Societies situated in the Val di Cecina.

Table 10 – Proportion of households that cannot face unexpected financial expenses (750 Euros), by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	36,96	4,19
Valdera	23,63	3,53
Alta Val di Cecina	18,77	4,45
Bassa Val di Cecina	25,17	4,73
Valdarno inferiore	32,70	4,49
Total (Province)	31,29	2,42

Table 11 – Proportion of households that cannot afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day, by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	8,30	2,70
Valdera	2,94	1,32
Alta Val di Cecina	6,03	2,95
Bassa Val di Cecina	7,59	2,95
Valdarno inferiore	14,47	3,52
Total (Province)	7,57	1,49

Finally, the 7,57% of the households living in the Province declared to not be able to afford a meal with meat, chicken, fish (or vegetarian equivalent) every second day. For this indicator the results are really differentiated between the Health Societies: note in particular the low estimate obtained for "Valdera" (2,94%) and the high estimate for "Valdarno inferiore" (14,47%). In this case, the differences are significant even considering the estimated standard errors.

As concerns social benefits, in Tables 12, 13, 14 and 15 we finally consider some pension indicators, that is some indicators based on the percentage of population living in the Province of Pisa declaring they were receiving a pension in 2007.

As we can see, the estimated proportion of individuals receiving a pension in 2007 is equal to approximately the 35% of the total population living in the Province. This percentage is poorly differentiated among the Health Societies, with the exception of the "Bassa Val di Cecina", which seems to be characterized by a higher proportion of retired people.

Table 12 – Proportion of individuals declaring they were receiving a pension, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	33,58	2,78
Valdera	38,21	2,76
Alta Val di Cecina	39,44	3,62

	Estimate	Standard Error
Health Society		
Bassa Val di Cecina	43,44	4,02
Valdarno inferiore	31,13	2,79
Total (Province)	35,09	1,61

Considering the different types of social benefits, over the 70% of the individuals receiving any type of social benefit declared they were receiving old age benefits[3] (Table 13). In this case the estimated standard errors suggest to be cautious in discriminating among the different areas.

The percentage of retired people declaring they are receiving survivors' benefits is approximately equal to 25% in the Province (Table 14), while the corresponding percentage is lower in the "Alta Val di Cecina" and "Valdarno inferiore" areas. Finally, we can note that considering the percentage of retired people declaring there are receiving a disability pension (Table 15), the "Alta Val di Cecina" Health Society is again characterised by a lower estimate if compared with the other areas. Thus, this area seems to be characterized by lower amounts of social benefits especially dedicated to survivors and people with disabilities.

Table 13 – Proportion of individuals receiving any type of social benefits declaring they were receiving old age social benefits, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	70,10	4,53
Valdera	77,54	3,90
Alta Val di Cecina	78,20	4,93
Bassa Val di Cecina	73,88	5,20
Valdarno inferiore	73,32	4,70
Total (Province)	73,52	2,48

Table 14 – Proportion of individuals receiving any type of social benefits declaring they were receiving survivors' social benefits, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	28,48	4,60
Valdera	25,59	4,33
Alta Val di Cecina	18,31	4,61
Bassa Val di Cecina	26,98	5,50
Valdarno inferiore	19,91	4,35
Total (Province)	25,68	2,57

---

[3] Old age benefits include old age pensions, anticipated old age pensions, partial retirement pensions, care allowances, survivor's benefits paid after the standard retirement age, disability cash benefits paid after the standard retirement age, lump-sum payments at the normal retirement date and other cash benefits. More details are available in Eurostat, 2008.

Table 15 – Proportion of individuals receiving any type of social benefits declaring they were receiving disability social benefits, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	35,05	4,88
Valdera	29,65	4,25
Alta Val di Cecina	12,87	3,92
Bassa Val di Cecina	21,93	4,93
Valdarno inferiore	27,14	4,70
Total (Province)	30,49	2,67

Summing up the main results standing from the oversampling direct estimates, we can say that the economic, poverty and social benefits indicators of the five Health Societies of the Province of Pisa are characterized by certain variability. Thus, computing the direct estimates only at provincial level would mask the important differences emerging when repeating the analysis at a more detailed geographical level.

Another interesting result is the variability between the different computed indicators. In particular, areas characterized by low mean and median household income estimates can be characterized by low estimated discomfort indicators. That is, direct income estimates and indicators of perceived economic discomfort can give different indication of the poverty and living conditions in a given area. Thus, it is important to always consider both types of indicators when analysing the areas of interest.

7. Comparison between EU-SILC indicators and administrative data
UROPS, Province of Pisa

The poverty of a territory and people who live in it is deeply influenced by demographic, social and economic development of this territory. Indicators from administrative data, referred to in paragraph 5, identify the local context of EU-SILC survey. These are environmental variables that describe the discomfort situations of municipalities or aggregation of municipalities but can not measure the individual or household discomfort.

The results of EU-SILC sometimes confirm the strengths and weaknesses of the territory that emerge from territorial analysis and sometimes contrasting with it.

If we try to compare EU-SILC indicators with administrative data and indicators can detect a certain correspondence that concerns the economic indicators: the lower income areas are Alta and Bassa Val di Cecina.

But the oversampling analysis can emphasize an important point which is the divergence between economic indicators and discomfort indicators and, sometimes, this discrepancy is confirmed by administrative data such as for Alta and Bassa Val di Cecina.

For example, Valdarno is a typical area with good income but with high indicators of discomfort, especially with regard to the proportion of households that are unable to keep home adequately warm and the proportion of households who can not afford a meal with meat, chicken, fish every two day. This situation is particularly obvious for Bassa Val di Cecina which is the area with the highest poverty rate and a lower level of distress indicators.

Looking at the regional indicators, we realize that Alta and Bassa Val di Cecina are the areas with the lower percentages of requests for rent contribution (another indicator of the regional network). This is a good indicator of social disadvantage because the contribution is restricted to households with very low income.

About pensions, there is a good consistency between the percentage of people who declared they receive a pension (35,09% at provincial level) and the percentage reported by INPS data (31,35%). The same correspondence exists when comparing the percentages of people with survivor's pensions (25,68% in EU-SILC and 21,5% in INPS data).

Finally, it may be interesting to analyze the social distribution of public expenditure in the light of survey results.

Figure 1 Public social expenditure per habitant (Euro). Year 2007

Public social expenditure per habitant (Euro). Year 2007
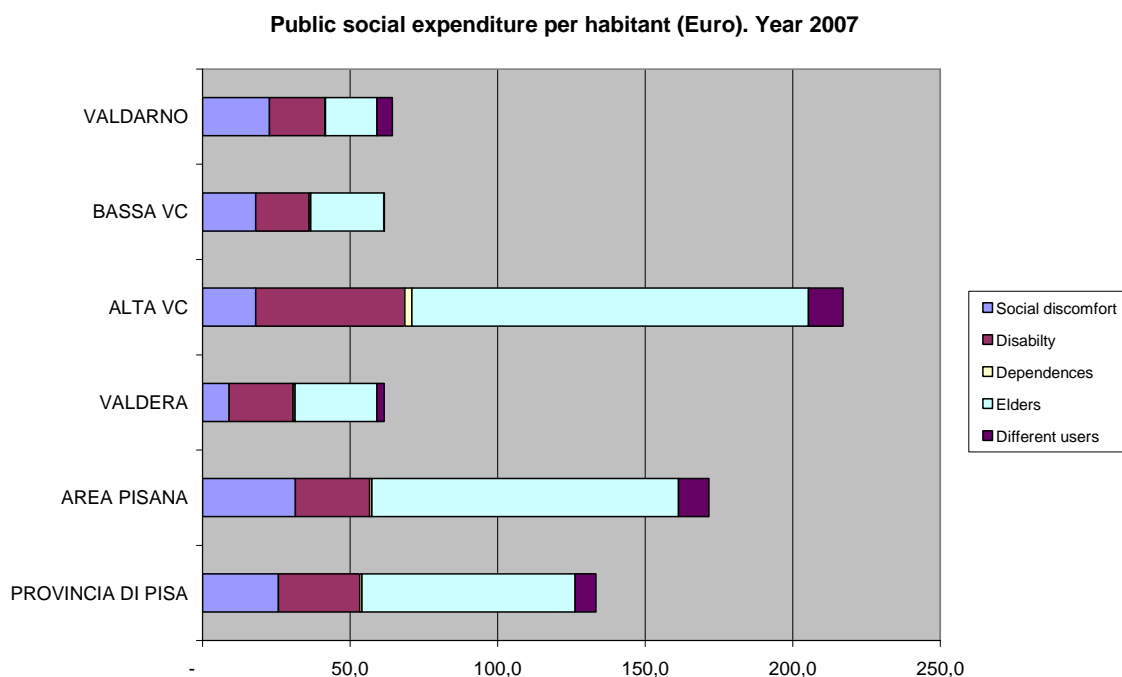
From EU-SILC we can see that in the Province of Pisa in 2007, there was a 15,83% of households with an income below the poverty line. That same year public spending on "Poverty" was € 2.692,77 at provincial level that correspondsto the 5% of total expenditure, to which we must add the cost for Families and Young People (10,7%), which include also social assistance spending and Immigrants (3,6%). In total, the public expenditure on social disadvantage was approximately 19,3% (€ 10.447.778). Instead 54,2% of the expenditure was absorbed by the interventions for the elderly.

In Figure 1 we can observe the distribution of public expenditure per capita. Compared with Table 1, we find that the Bassa Val di Cecina has a low capacity for action on social discomfort

Table 1– Head Count Ratio (%) by Health Society, year 2007.

	Estimate	Standard Error
Health Society		
Area pisana	15,52	3,05
Valdera	15,46	3,08
Alta Val di Cecina	15,94	4,16
Bassa Val di Cecina	24,77	4,78
Valdarno inferiore	15,87	3,54
Total (Province)	15,83	1,82

In this context, it is important to evidence that the social expenditure of Bassa Val di Cecina includes also the expenditure for 4 municipalities that are of the competence of Province of Livorno (Bibbona, Cecina, Castagneto Carducci, Rosignano Marittimo). It is possible that the area of discomfort of Bassa Val di Cecina relates to a greater extent than the municipalities of the Province of Pisa.

References

Consolini P. (2008), "Practical aspect on the harmonization of the definitions, of the variables and units for the EU-SILC project in Italy", Report of WP2, *ESSnet Statistical Methodology Project on the Area: Integration of survey and administrative data,* downloadable from the website: http://cenex-isad.istat.it

Consolini P. (2009), "Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa", *Collana Metodi e Norme Istat*, 38, Roma.

Eurostat (2007), *Manual UDB variables description*, ver 2007-1 from 01-03-09.

Eurostat (2008), *Description of target variables: Cross-sectional and Longitudinal, 2008 operation*. EU-SILC 065 (2008 operation)

Kish L. (1965), *Survey Sampling*, New York: Wiley.

Raghunathan, T.E., Lepkowski, J.M., van Hoewyk, J., Solenberger, P. (2001), "A multivariate technique for multiply imputing missing values using a sequence of regression models", *Survey Methodology*, 27, pp. 85-95.

SAMPLE Consortium - Deliverable 7.1

Van der Laan, P. and H.K. van Tuinen (1997), "Increasing the Relevance of Income Statistics: Experiences and Plans in the Netherlands", *Papers and Final Report of the First Meeting on Household Income Statistics*, Canberra, Australia, 2-4 December 1996, Expert Group on Household Income Statistics, p. 193-220. Belconnen: Australian Bureau of Statistics.

Van der Laan P. (2000), ""Integrating administrative registers and household survey", *Netherlands Official Statistics*, 15, Summer 2000