

## WORKSHOP ON DATA INTEGRATION

Madrid, 24-25 November 2011

Session 5 of the Agenda: Register-based statistics

## THE INTEGRATION OF THE SPANISH LABOUR FORCE SURVEY WITH THE ADMINISTRATIVE SOURCE OF PERSONS WITH DISABILITIES

---

<b>INDEX</b>	<b>page</b>
<b>1. INTRODUCTION .....</b>	<b>2</b>
1.1. NECESSITY OF INFORMATION .....	2
1.2. LOOKING FOR A SOLUTION .....	3
1.3. WORKING GROUP .....	3
<b>2. DESCRIPTIONS OF THE PROJECT .....</b>	<b>4</b>
2.1. OBJECTIVES .....	4
2.2. CHARACTERISTICS .....	5
2.2.1. DEFINITIONS (established by the sources of information used).....	5
2.2.2. SCOPE (established by the source of information used).....	6
<b>3. TREATMENT OF THE INFORMATION .....</b>	<b>6</b>
3.1. DATA PREPARATION .....	7
3.1.1. STANDARDIZING IDENTIFIERS (IN SDPD DATA):.....	7
3.1.2. CHECKING DUPLICATIONS (IN SDPD DATA): .....	8
3.1.3. CHECKING VALID VALUES (IN SDPD DATA): .....	9
3.1.4. DELETING REGISTERS:.....	9
3.2. MATCHING .....	10
3.2.1. MATCH 1: THROUGH THE STANDARD IDENTIFIER .....	10
3.2.2. MATCH 2: THROUGH THE BIRTH DATE AND THE LOCALITY CODE ..	11
3.2.3. MATCH 3: THROUGH THE STANDARD IDENTIFIER .....	11
3.2.4. MATCH 4: INVESTIGATION OF STRANGE IDENTIFIER IN SMALL LOCALITIES .....	11
3.3. ESTIMATION AND WEIGHTS .....	12
<b>4. RESULTS. COMPARISON WITH OTHER SOURCES .....</b>	<b>16</b>
<b>5. CONCLUSIONS .....</b>	<b>17</b>

## 1. INTRODUCTION

---

The “*Employment of persons with disabilities*” investigates the situation, with regard to the labour market, of the group of persons between the ages of 16 and 64 years old who hold disability certificates.

This operation provides data regarding the labour force (employed persons, unemployed persons) and for the population outside the labour market (inactive persons) within the group of persons with disabilities.

It is formed as a periodic operation of an annual nature that uses information deriving from integrating statistical data supplied by the Economically Active Population Survey (EAPS, that is the Spanish Labour Force Survey, LFS) with the administrative data recorded in the State Database of Persons with Disabilities (SDPD).

### 1.1. Necessity of information

The group of persons with disabilities has formed an axis for priority action in social policies carried out in recent years in order to achieve integration of these persons in the workplace.

In particular, it is an essential point of interest for:

- The Spanish Committee of Representatives of Persons with Disabilities (CERMI) that is the Spanish umbrella organisation representing the interests of more than 3.8 million women and men with disabilities in Spain. The mission of CERMI is to guarantee equal opportunities of women and men with disabilities and to protect their human rights, ensuring they are fully included in society.
- The ONCE Foundation, whose main objective is to implement integration programmes of work-related training and employment for people with disabilities, and universal accessibility, promoting the creation of universally accessible environments, products and services.
- The Elderly and Social Services Institute (IMSERSO) that is the Social Security Administration Body responsible for handling Social Services supplementing Social Security System provisions and which deals with older and dependent persons.
- The Spanish National Institute of Statistics (INE), that, as coordinator of the official statistics, has the mission of completing the lack of information and of promoting the utilization of administrative sources to produce data without increasing the budget and without overburdening the informants.

In 2009, the available information on Spanish employment and disability was:

1. SDIH-1986: the Survey on Disabilities, Impairment and Handicaps
2. SDIHS-1999: The Survey on Disabilities, Impairments and Health Status
3. Ad-hoc Module-2002 on employment of disabled people for the labour force sample survey provided for by Council Regulation
4. DIDSS-2008: Disabilities, Independence and Dependency Situations Survey.

These surveys are carried out in wide, irregular and infrequent periods, so that a continuous monitoring of the situations of the persons with disabilities cannot be performed. But information is required for evaluating the effectiveness of the policies

and the current situation of persons with disabilities, mainly in 2008, year in which there is special interest in measuring the crisis effects.

## 1.2. Looking for a solution

On June 2009, a meeting between INE, CERMI and ONCE-foundation took place. Both organizations demand to INE periodic information about the employment of persons with disabilities, and to obtain it, they suggest to introduce the variable “disability” into all the social surveys.

INE shows several considerations against this proposal:

- On one hand, the reduction of the burden to the informants is a primary objective for the institute.
- On the other hand, the study of any phenomenon requires a set of questions in the questionnaires to assure the correct definition of the concept. So that an overburdening in the questionnaires is required. But it would provide a lack of quality of the responses, mainly in the EAPS that have a big questionnaire with a great deal of questions (labour, familiar relations, education, annual modules...)

In this situation INE and CERMI propose looking for alternative ways. There are two possible sources of information:

- *The Economically Active Population Survey (EAPS):*  
It is the Spanish Labour Force Survey (LFS) and involves the most complete source of information about the situation of the labour market. The main objective of the EAPS is to reveal information on economic activities as regards their human component.
- *The State Database of Persons with Disabilities (SDPD):*  
It is a registration system, with national scope, for persons with disabilities. It provides information regarding the features of citizens who have officially been recognised as persons with disabilities by the State administrative bodies with jurisdiction. It is managed by the IMSERSO.

Therefore it is decided that the best solution to get information is to use information deriving from integrating statistical data supplied by the EAPS with the administrative data recorded SDPD.

This solution will permit getting the required information without overburdening the informants and without increasing the INE budget (given that the introduction of this system would be partially financed by the ONCE Foundation).

## 1.3. Working group

For accomplishing this new register-based statistical operation regarding employment and disability, the CERMI, ONCE-Foundation, IMSERSO and INE have created a working group with the aim of sharing work, knowledge and experience.

During 2010 this working has evaluated 2008-data to study the feasibility of getting relevant, reliable and periodical information on persons with disability and their labour market relations on the base of merging the EAPS with the SDPD. So that along this

year, “The employment of the persons with disabilities 2008” is developed as pilot study for evaluating the possibility of getting information on “Disability and Employment” by crossing 2008-EAPS versus 2006-SDPD (and previously 2008-EAPS with the POPULATION REGISTER from INE, to assign the identification number to the EAPS registers).

The IMSERSO has provided the 2006-SDPD that at that moment was the latest available update.

The CERMI contributes with its knowledge, experience and technical support.

The ONCE-Foundation, besides technical support as CERMI, provides financing for hiring supporting staff.

The INE supplies the 2008-EAPS data and assumes the technical and operative tasks for establishing the methodological bases and the procedures to get periodical results.

The success of this study has led INE to establish it as a periodic operation of an annual nature and to extend it by merging with other administrative sources (Social Security, Pensions and Dependency)

## **2. DESCRIPTIONS OF THE PROJECT**

---

### **2.1. Objectives**

The overall objective of the project is to fulfil the demand of information on the situation of the persons with disabilities with respect to the labour market.

The focuses of the study are:

- To estimate the number of employed, unemployed and inactive persons inside the collective of persons with disabilities, as well as their comparison with the persons without disabilities.
- To carry out the analysis of the disability and labour market from the perspective of gender.
- To implement the analysis of the participation in the labour market from the perspective of the type of impairment.

Also, there are complementary objectives derived from the available information from EAPS and SDPD data:

- To obtain the evolution and variation along the time of the number of active/inactive persons with disability and their comparison with the evolution inside the persons without disability.
- To ascertain the characteristics of persons with disabilities relating their personal, familiar and geographical features with their labour market situation.
- To determine the features of the type of disability and its severity (SDPD variables) versus the employment and households variables (EAPS variables).

## 2.2. Characteristics

### 2.2.1. DEFINITIONS (established by the sources of information used)

The variables *employed/unemployed/inactive* are defined by the LFS regulation.

The variable *disability* is defined by the Spanish legislation.

- *Employed population*: All persons 16 years old and older who, during the reference week:
  - a) Either worked for at least one hour, even sporadically or occasionally, in exchange for a salary, wages or another form of remuneration in cash or in kind.
  - b) Or were employed but not working (due to illness, holidays, leaves, work conflicts, bad weather...)
- *Unemployed population*: Those persons 16 years old or older who combine the following conditions simultaneously:
  - a) Without work
  - b) Seeking work
  - c) Available to work
- *Economically active population*: Employed and unemployed population
- *Inactive population*: Those persons 16 years old or older who in the reference week cannot be classified neither employed nor unemployed.
- *Disability*: It is a wide concept that can be analysed from several points of view.

For the World Health Organization (WHO) *disability* is considered an umbrella term encompassing impairments, limitations of activity and restrictions of participation. The WHO expands the concept of health incorporating environmental factors (physical, social and attitudinal environment in which the persons live and carry out their lives).

But from the legal point of view, there is an administrative procedure, defined by regulation<sup>[1]</sup> in which persons with disability voluntarily participate to be evaluated by qualified equipment that determine the degree of disability in function of different parameters defined by law. The evaluated persons that have a degree of disability equal or bigger than 33% are those who are “legality” considered as persons with disabilities and who receive an official certificate of disability.

The SDPD includes all the persons who have been evaluated through this procedure.

As the project is developed over the base of the SDPD, the definition of disability is delimited by the legal definition used for the SDPD, that mainly check for illnesses and impairments and it is different from the WHO recommendation, though it is important to remark that in many cases both definitions coincides.

---

[1] “*Real Decreto 1971/1999, de 23 de diciembre, de procedimiento para el reconocimiento, declaración y calificación del grado de discapacidad. Real Decreto 1856/2009, de 4 de diciembre, de procedimiento para el reconocimiento, declaración y calificación del grado de discapacidad, y por el que se modifica el Real Decreto 1971/1999, de 23 de diciembre*”.

### 2.2.2. SCOPE (established by the source of information used)

The scope for the project is delimited by the EAPS and the SDPD scopes.

- EAPS scope and sampling: The EAPS is a quarterly survey whose scope considers the population living in family dwellings, excluding group or collective dwellings (hospitals, residences, barracks, etc.) and secondary or seasonal dwellings (used during holiday periods, at weekends, etc.)

The survey uses a two-stage sampling with first stage unit stratification. The First stage units are the census sections (areas established for electoral purposes). The second stage units are family dwellings: information is collected on all persons regularly living in the dwelling.

The total sample, formed by around 65.000 households, is divided into six subsamples. Family dwellings are renovated partially every quarter of the survey, in order to avoid tiring the families. Each quarter, the dwellings in the sections of a specific subsample are renewed.

- The SDPD scope: Only includes people that have freely and voluntarily asked the evaluation procedure. In compensation, it is a census that includes the whole population with legally and officially recognized disability, and it has a high degree of liability because, by law, the evaluating equipment is formed by doctors, psychologists and social workers.
- “The employment of persons with disabilities” scope: because of the nature of the project, it is defined by EAPS and SDPD scopes (defined above), so it includes:
  - Persons with certificated of disability
  - Persons living in family dwellings (not collective dwellings)
  - Persons between 16 and 64 years old.

## 3. TREATMENT OF THE INFORMATION

---

The objective is to link EAPS and the SDPD to get information on “Disability and Employment”.

The information is going to be joined at microdata level. It means that every unit in the EAPS sample is completed with information about disability (degree, impairment, severity) from the SDPD, so that the EAPS estimators could be applied either to the collective of persons with disabilities or to the population without disabilities.

Unfortunately, when integrating data from different sources, several problems and conflict may appear:

Firstly, SDPD data lack an unique and global identifier that permits the linking operation. Furthermore, the data are neither carefully controlled by many factors, including data entry errors, lack of standard format, incomplete information or any combination of these factors.

Finally, as the initial sample was chosen to cover the EAPS objective but not for estimating data on disability, as the disability is a phenomenon that affect to a small

percentage of the population (2,8% of population between 16 and 64 years old), the weights should be readjusted to get reliable estimations on disability and employment. So to get the objective of the project, several tasks should be undertaken:

- Data preparation
- Field matching
- Choose appropriated weights

### 3.1. Data preparation

- EAPS data:

The survey data are consistent and correct, because they have been previously checked.

The only unresolved question in EAPS is the lack of an identifier for each person. There is a specific project in INE (managed by EAPS area) that has the mission to assign identifiers to EAPS fields by matching the survey records with the Population Register. So these identifiers are provided on time to undertake the project “The employment of the persons with disabilities”.

In summary, EAPS information is prepared to be linked with other sources

- SDPD data:

Data from SDPD must be standardized and cleaned. In particular the outstanding task for preparing the SDPD link with EAPS are:

#### 3.1.1. STANDARDIZING IDENTIFIERS (IN SDPD DATA):

The SDPD identifier (IDEN) is recorded in a not standard format. For standardizing it, an unique number of order is assigned to each field in the SDPD.

After that, every numerical code inside the not standard identifier is extracted from the SDPD and considered as a candidate for crossing with EAPS and with the Population Register<sup>[2]</sup> (PR).

Hence, every SDPD field can have zero, one or more candidate to be standard identifiers and all of them are used for the joins with different sources.

These joins are made taking into account not only the standard identifier candidates but also the birth date, the presence of duplications and the locality. When the join is accepted, then the standard candidate that has been matched is chosen as definitive and it is stored for futures queries or joins (COD\_FIN).

---

<sup>[2]</sup> The Population Register (PR) is the administrative register in which inhabitants are recorded. Its data constitutes proof of residence in the municipality. Everybody who resides in Spain is obliged to register in the Municipal Register in which they habitually reside. Anyone who lives in several municipalities will have to register only in the one in which they spend more time in over the year.

Amongst others, the PR includes the variables identifier (Identity card, Foreign identity number or passport), name and surname, direction, incidences (that include death), type of dwelling (normal/collective).

## OBTENTION OF AN STANDARD IDENTIFIER

Number of order	Not standard identifier	Standard identifier (cross candidates)			(Chosen standard identifier )
NORDER	IDEN	NUM1	NUM2	NUM3	COD_FIN
0100000001	'LLLLLLL'	0			
0100000002	'0'	00000000			
0100000003	'12345678L'	12345678			12345678
0100000004	'123/12345678L'	00000123	12345678		12345678
0100000005	'12345/1234L'	00123456	00001234		

Crosses

None crosses

## 3.1.2. CHECKING DUPLICATIONS (IN SDPD DATA):

The SDPD data are managed by the autonomous communities (Spanish regions). There can be cases in which the same person with disability would be evaluated in two different moments in two different autonomous communities. These cases are duplicated registers in the SDPD.

For detecting duplications, it should be taken into account as the standard as the non standard identifier together with the birth date (BD):

## DUPLICATED BY IDEN &amp; BIRTH DAY (BD)

NORDEN	REG	IDEN	BD
3900000306	06	2707608	19570723
3900000307	06	2707608	19570723
1100000579	01	D015382294	19640107
2100000284	01	D015382294	19640107
3300032878	03	D015382294	19640107

## DUPLICATED BY COD\_FIN &amp; BD

NORDEN	REG	IDEN	COD_FIN	BD
3600021773	12	X00104798	104798	19520214
3600021786	12	X0104798	104798	19520214
0300000057	10	377512	377512	19530622
1800000018	01	D000377512	377512	19530622
2800003408	13	00377512J	377512	19530622

## DUPLICATED BY COD\_FIN or IDEN &amp; BD

NORDEN	REG	IDEN	COD_FIN	BD
0400008988	01	D031789529	31789529	19450328
1200008515	10	31789529	31789529	19450328
4600048258	10	31789529	31789529	19450328
1100016693	01	D031789529	.	19450328

To solve duplications, the next ordered rules has been disposed:

It is chosen the register whose...

1. SDPD location coincides with the update PR location



2. SDPD location coincides with the 2006-PR (2006 is the reference period for SDPD. June the 31<sup>st</sup>, is the reference date used for 2008-PR)
3. SDPD location coincides with the birth location in the PR.
4. SDPD province coincides with the PR province
5. Date of update in the SDPD is the most recent.
6. ...The rest of duplications are randomly solved.

Note: the latest update is used in the last place (before random selection) because the variable 'update' in the SDPD is missing in most of the registers.

With this process, only the 5,7% of the duplications are solved randomly and the solutions can be used from one year to another.

### 3.1.3. CHECKING VALID VALUES (IN SDPD DATA):

The errors detected are:

- No valid identifiers:  
In some cases the IDEN variable contains the birth date. These cases are detected and marked as possible wrong identifiers that should be treated carefully in the case of being matched with some other source.
- Local standardization:  
The SDPD uses an old and obsolete system of codes for the variable locality (LOCAL). So INE assumes the task of translating this old-fashioned system to the update standard one.
- Age and sex correction:  
For all the registers linked with the PR, the inconsistencies between the variables Birth Date (BD) and SEX are reviewed.

a) If the register does not cross with the EAPS then it is given priority to the POPULATION REGISTER information versus the SDPD one.

Besides, for SEX variable, the inconsistencies between SDPD and PR links are solved by kind or name: a list with the names that have inconsistencies is obtained and each name is classified in 'female name', 'male name', 'unisex name'.

For female /male names the sex is assigned in corresponding to the gender of the name. For unisex names a detail revision one by one permit to establish, in function of the relationships between the members of the household, the sex of the person. Finally, in case of doubt, PR value is chosen.

b) If the register cross with EAPS then:

- b1) The information that coincides in two of the three sources is the chosen one.
- b2) If there is no coincidence in any of the three sources:  
Each disagreement is revised for taking a choice, always under the next order of priorities: first EAPS information, secondly POPULATION REGISTER information, and finally SDPD information.

### 3.1.4. DELETING REGISTERS:

Finally, there are three classes of registers that have been removed from the SDPD

- Out of age-range deletions:  
After the sex and age variables are corrected, persons with age out of range [16,64] are deleted
- Deleting dead persons:  
For all the registers linked with the PR, it can be determined if the person is alive. In case of dead the register is considered out of scope.
- Removing persons living in collective dwellings:  
EAPS scope exclude people living in collective dwellings, so to keep consistency, the project should delete from SDPD people living in this kinds of dwellings.

INE has a DIRECTORY of CENTERS that was utilised in the Disabilities, Independence and Dependency Situations 2008 Survey and that now it is crossed with the SDPD through the variable street-code. This code is assigned to the SDPD through a distance function between the SDPD literal for the street and the code-list literal for the street. Finally the cross is reviewed by using the PR value on collective/normal household.

**CLEANING SDPD**

Total number of registers (initially)	1.048.838
Persons living in collective establishments	17.393
Removed duplicates	19.039
Deads (until 2008)	96.326
Age out of range (16-64)	33.537
Final number of registers	892.455

**3.2. Matching**

The objective is to merge the EAPS registers with the SDPD ones to obtain the subsample of persons with disabilities inside the sample of EAPS.

Four ways for matching these registers are considered:

**3.2.1. MATCH 1: THROUGH THE STANDARD IDENTIFIER**

The SDPD registers which have a numerical code as standard identifier are merged with EAPS through this code (COD\_FIN)

For deciding if the matching is or not valid, the variables birth date (BD) and locality code (LOCAL) are taken into account as shown below

**SUMMARY OF SDPD REGISTERS MATCHED WITH EAPS THROUGH CODFIN**

Freq	Coincidences (by CODFIN-BD-LOCAL)	Description and treatment
<b>5.035</b>	<b>Coincidences by CODFIN</b>	Total or matches between BEPD and EAPS
3.380	CODFIN+AAAAMDD+PPMMM	=> Matches correct: revised through a sample
752	CODFIN+AAAAMDD+PP	
222	CODFIN+AAAAMDD	
133	CODFIN+AAAAMM+PPMMM	

Freq	Coincidences (by CODFIN-BD-LOCAL)	Description and treatment
30	CODFIN+AAAAMM+PP	=> Matches correct: revised exhaustively
11	CODFIN+AAAAMM	
56	CODFIN+AAAA+PPMMM	
24	CODFIN+AAAA+PP	
6	CODFIN+AAAA	
54	CODFIN+AAA+PPMMM	
13	CODFIN+AAA+PP	
40	CODFIN+AAA	
100	CODFIN+PPMMM	=> Familiar relations: revised exhaustively (to find the correct CODFIN through the MR)
37	CODFIN+PP	
177	CODFIN	=> Not valid matches, rejected

CODFIN= standard identifier

BD= Birth date (year=AAAA+ month=MM + day=DD)

LOCAL= code of locality (province=PP + locality=MMM)

### 3.2.2. MATCH 2: THROUGH THE BIRTH DATE AND THE LOCALITY CODE

The SDPD registers for which there are not standard identifiers, are investigated through the birth date (BD) and the locality code (LOCAL) as follows:

- a) For each pair (BD,LOCAL) in EAPS, it is obtained its frequency in this survey.
- b) The SDPD registers without standard identifier are merged with EAPS through the variables BD and LOCAL
- c) The cases from b) that have a frequency of (BD, LOCAL) in EAPS equal to 1 are investigated in the Population Register (PR). If in the PR the frequency of these cases is also 1, this means that in Spain there is only one person with this features and the match is correct.
- d) The cases from b) that have a frequency of (BD, LOCAL) in EAPS equal to 2 and correspond to persons of different sex are investigated in the Population Register. If in PR the frequency is also 2, the sex permit to distinguish which of them is the correct and again, this means that in Spain there is only one person with these features so the match is correct.

### 3.2.3. MATCH 3: THROUGH THE STANDARD IDENTIFIER

The SDPD registers that have not a standard identifier are merged with EAPS through this non-standard code (PASSPORTS). In spite of being a small quantity of possibilities, some of them can result correct matches.

### 3.2.4. MATCH 4: INVESTIGATION OF STRANGE IDENTIFIER IN SMALL LOCALITIES

For SDPD registers with "strange identifiers" that are situated in localities that are in EAPS sample and whose population is less than 30.000, can be investigated in the Population Register to determine if some of them really match with EAPS.

Strange identifiers are those that in the SDPD have:

- Identifier equal to missing value or
- Birth date inside the identifier code or
- Identifier that contains “O” instead “0” or
- Identifiers with code numbers whose length is bigger than 8 digits (8 is the length of the standard identifier in Spain)

As shown bellow, the principal part of the sample of persons with disabilities obtained from the merging between SDPD and EAPS is obtained through the standard identifier (match 1), but it is obvious that there is part of the sample that can be obtained from other kind of investigation and, as the sample size of persons with disabilities is small, it is worth to carry out all the explained methods.

#### EAPS ∩ SDPD SAMPLE

MATCH 1 =>	<b>98,0</b> % of the total match
MATCH 2 =>	<b>0,7</b> % of the total match
MATCH 3 =>	<b>0,2</b> % of the total match
MATCH 4 =>	<b>1,1</b> % of the total match

### 3.3. Estimation and Weights

#### 3.3.1. ESTIMATION:

The expression of the EAPS estimator for a specific characteristic Y in a certain quarter of survey is as follows:

$$\hat{Y} = \sum_h F_h Y_h = \sum_h F_h \sum_{i=1}^{n_h} y_{hi}$$

the sum h is extended to the strata of a province, an autonomous community or the national total, in which

$F_h$ : it is the weight for the stratum h

$n_h$ : it is the number of persons in the sections of the sample in stratum h.

$y_{hi}$ : it is the value of the characteristic researched in person i-th, of stratum h.

Disability is an atypical phenomenon than affects to a small percentage of the population and the EAPS is a survey designed to obtain labour market results, but not disability figures. So it is expected that the sample of persons with disability obtained from the integration of the SDPD with EAPS would be not enough.

For deducing if the sample size is enough, it is compared with a sub-scope from the EAPS for which quarterly results are delivered. The reference sub-scope that it is chosen is the “region”, that in Spain it is defined by autonomous community:

Sampling size (n) obtained to estimate the total population (N)

QUARTERLY SAMPLE SIZE IN EAPS BY REGIONS

Regions	n	n/N
<b>Total (Spain)</b>	<b>141.118</b>	<b>0,45%</b>
Andalucía	24.821	0,45%
Aragón	6.305	0,73%
Asturias (Principado de)	4.150	0,57%
Balears (Illes)	3.384	0,46%
Canarias	7.316	0,50%
Cantabria	3.825	0,97%
Castilla y León	14.303	0,86%
Castilla - La Mancha	10.216	0,77%
Cataluña	14.644	0,29%
Comunitat Valenciana	11.715	0,35%
Extremadura	5.729	0,80%
Galicia	9.686	0,52%
Madrid (Comunidad de)	7.355	0,17%
Murcia (Región de)	4.486	0,46%
Navarra (Comunidad Foral de)	3.134	0,76%
País Vasco	6.586	0,45%
Rioja (La)	2.465	1,16%
Ceuta (Ciudad Autónoma de)	552	1,14%
Melilla (Ciudad Autónoma de)	446	0,97%

Persons with disabilities  
in QUARTERLY EAPS sample:

$$n_D = 3.000$$

$$n_D / N_D = 0,33\%$$

As shown in the figures, there are some regions whose estimations are quarterly provided and that have similar sample size to the persons with disability.

ANNUAL SAMPLE SIZE IN EAPS BY REGIONS

Regions	n	n/N
<b>Total (Spain)</b>	<b>141.118</b>	<b>0,45%</b>
Andalucía	24.821	0,45%
Aragón	6.305	0,73%
Asturias (Principado de)	4.150	0,57%
Balears (Illes)	3.384	0,46%
Canarias	7.316	0,50%
Cantabria	3.825	0,97%
Castilla y León	14.303	0,86%
Castilla - La Mancha	10.216	0,77%
Cataluña	14.644	0,29%
Comunitat Valenciana	11.715	0,35%
Extremadura	5.729	0,80%
Galicia	9.686	0,52%
Madrid (Comunidad de)	7.355	0,17%
Murcia (Región de)	4.486	0,46%
Navarra (Comunidad Foral de)	3.134	0,76%
País Vasco	6.586	0,45%
Rioja (La)	2.465	1,16%
Ceuta (Ciudad Autónoma de)	552	1,14%
Melilla (Ciudad Autónoma de)	446	0,97%

Persons with disabilities  
in ANNUAL EAPS sample:

$$n_U = 5.000 \quad \text{UNION= different units}$$

$$n_U / N_D = 0,56\%$$

$$n_S = 12.000 \quad \text{SUM= sum of interviews}$$

$$n_S / N_D = 1,38\%$$

In this way, annual data will provide more reliable estimations.

So the estimator used is the annual average calculated as average of the quarterly estimators:

$$\hat{Y} = 1/4 \sum_{i=1}^4 \sum_{hi=1}^{ni} F_{hi} Y_{hi} = \sum_{h=1}^{nu} (F_{h1} Y_{h1} + F_{h2} Y_{h2} + F_{h3} Y_{h3} + F_{h4} Y_{h4}) / 4 = \sum_{h=1}^{n1+\dots+n4} F_h / 4 Y_h$$

where:

$n_i$ = sample size in the quarter  $i$ , for  $i=1$  to  $4$

$n_U$ = total number of different units in the sample=union of the four quarterly samples

$F_{hi}$ = original weight of the unit  $h$ , if the unit  $h$  is interviewed in the quarter  $i$   
 0 if the unit  $h$  is not interviewed in the quarter  $i$

$Y_{hi}$ = value of the variable  $Y$  for the unit  $h$ , if the unit  $h$  is interviewed in the quarter  $i$   
 0 if the unit  $h$  is not interviewed in the quarter  $i$

$j$ = re-numeration of the units, from 1 to  $n1+\dots+n4$

With this method, really there are four scopes, one for each quarter:

SDPD registers with age [16,64]		Persons with disabilities		
by quarters		EAPS sample		
<b>T1</b>	883.010	<b>T1</b>	3.155	
<b>T2</b>	876.038	<b>T2</b>	3.047	
<b>T3</b>	869.727	<b>T3</b>	2.966	
<b>T4</b>	863.421	<b>T4</b>	2.944	
Reference population		=> AVERAGE SIZE=		3.028
		UNION=		5.056 = $n_h$
		SUM=		12.112= $n1+\dots+n4$

The UNION gives the total number of interviewed units different (common units in different quarters are taken into account once) the estimator can be written as a lineal combination of the responses for each quarter.

The SUM gives the total number of interviews, independently of the units (the unit can have from 1 to 4 interviews in a year), the estimator can be written as a Horvitz-Thomson estimator whose weights are equal to the original ones divided by four.

### 3.3.2. FINAL WEIGHTS:

The survey process includes a calibration<sup>1</sup> of the design factors using the following auxiliary variables inside each autonomous community:

- X1: Population aged 16 years and over by age groups and sex.
- X2: Population aged 16 years and over by autonomous community and nationality, Spanish or foreign.
- X3: Population aged 16 years and over by province

<sup>1</sup> The calibration is made with the CALMAR (CALage sur MARGes) software, programmed by INSEE (Institut National de la Statistique et des Études Économiques) de France.

In the case of estimations for persons with disabilities:

- Due to the population with disabilities represents a small part from the total population (5,1% in the case of Spanish population with age [16,64]).
- As disability is not an objective in the design of EAPS, it is expected that the subsample of persons with disability in EAPS would produce subestimations in the results.

**NUMBER OF PERSONS WITH DISABILITIES-FIRST TEST ESTIMATION**

SEX	Dis.Pers. ESTIMATIONS	SDPD (1)	DIFFERENCES
TOTAL	717.221	960.403	-25,3%
MEN	406.659	540.294	-24,7%
WOMEN	310.562	420.109	-26,1%

(1) SDPD size:  
initial cleaning state

Hence, for obtaining data for persons with disability, the EAPS ones are calculated again by applying new auxiliary variables to reweight and adjust the survey estimates to the information from the SDPD in each autonomous community.

As EAPS results have already been delivered and disclosed, it cannot be obtained a different number of employed/unemployed Spanish people referred to the same period. Therefore it is also needed to adjust the survey estimations to the previously disclosed results.

So finally, the EAPS reweighting procedure is recalculated including now inside each autonomous community (in the same way that EAPS):

a. Adjustment to the same variables than the original EAPS

- X1. Population by province
- X2. Population by age & sex groups
- X3. Population by nationality, Spanish or foreign

b. Adjustment to the main EAPS results (quarterly)

- E1. Number of employed by sex
- E2. Number of unemployed by sex
- E3. Number of inactive by sex
- E4. Number of households

c. Adjustment to the principal SDPD information

- B1. Disability population by sex
- B2. Disability population by age
- B3. Disability population by impairments
- B4. Disability population by severity

As a consequence of this calibration, the average weight for persons with disabilities grows up, while the average weight for persons without disabilities lightly decreases.

## AVERAGE EAPS WEIGHT

Average WEIGHTS	Original WEIGHTS	Final WEIGHTS
TOTAL	280,5745	280,5714
Persons without disabilities	281,9268	280,5041
Persons with disabilities	234,5017	282,8635

#### 4. RESULTS. COMPARISON WITH OTHER SOURCES

The “Employment of the persons with disability” (EPD) results can be compared with the Disabilities, Independence and Dependency Situations Survey (DIDSS-2008) ones.

## PERSONS WITH DISABILITIES PERCENTAGE OF THE [16,64] POPULATION

	EPD-2008	DIDSS-2008
<b>Total</b>	2,8%	4,8%
<b>With legal disability certificate</b>		4,3%

## PERSONS WITH DISABILITIES AND ACTIVITY

			EPD-2008	DIDSS-2008
<b>Total population</b>	<b>Total</b>	<b>Total</b>	<b>873,3</b>	<b>1.322,2</b>
	<b>Active</b>	Total	292,3	408,0
		Employed	244,6	331,5
		Unemployed	47,7	76,5
	<b>Inactive</b>	Inactive	581,0	914,0
	<b>Does not Known</b>	Does not Known	0,0	0,3

The main differences between both surveys are:

- c) The objective population that in the DIDSS is the population with limitations of activity and with restrictions of participation in the everyday situations, while in the EPD it is the persons with legal disability certificate
- d) The way of collecting the information, because the DIDSS collects them from the auto-declaration of the persons, who have answered both if they have limitations or restrictions and disability certificate., while the SDPD are collected from the official database of disability certificates. The autodeclaration always overestimates the real results because there are persons declaring certificates only for having been evaluated (though they haven't got more than 33% of disability degree), or only for having some kind of physical problem (and they don't have official certificate)...
- e) The different period of collection: the complete year 2008 in EPD, from 2007-November to 2008-February in DIDSS

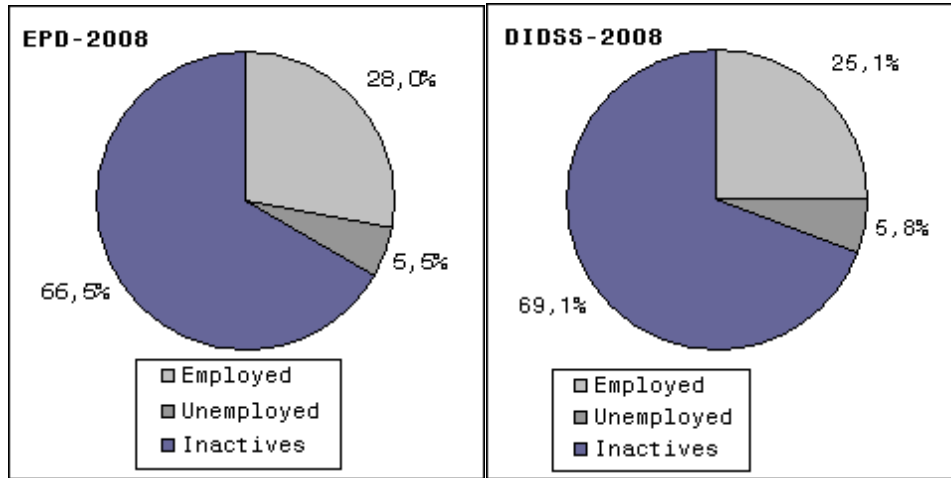
In summary, different objectives, different samples and different periods of collection.

As a consequence, the DIDSS shows greater figures on disability than the EPD

But, without forgetting the differences in definitions and scopes, and taking into account that the main objective of the EPD is to provide information about the activity of the



persons with disability, the correct point of view to compare the EPD with DIDSS is in percentage terms.



The comparison between the percentages of persons with disability who are employed, unemployed or inactive shows similar results in both surveys.

## 5. CONCLUSIONS

The Employment of the persons with disabilities is a survey that has been carried out through a low priced and efficient method.

It is a model of utilization of administrative sources that provides reliable and periodic figures on an preferential variable that is object of social and labour policies.