

Ole Villund

**Household data in the Norwegian Labour
Force Survey**

© Statistics Norway, January 2011 When using material from this publication, Statistics Norway shall be quoted as the source.	Symbols in tables	Symbol
	Category not applicable	.
	Data not available	..
	Data not yet available	...
	Not for publication	:
	Nil	-
ISBN 978-82-537-8013-9 Printed version	Less than 0.5 of unit employed	0
ISBN 978-82-537-8014-6 Electronic version	Less than 0.05 of unit employed	0.0
ISSN 1891-5906	Provisional or preliminary figure	*
Subject: 06.90	Break in the homogeneity of a vertical series	—
Print: Statistics Norway	Break in the homogeneity of a horizontal series	
	Decimal punctuation mark	.

Preface

This document is a reprint of the final report delivered to EUROSTAT regarding a project partially financed by a grant under the 2008 programme "Pilot studies and quality improvement of the LFS and its modules. Household composition variables."

There is a growing acceptance for the importance of labour market analyses at the household level, complementing traditional statistics at the individual level. The project aimed at improving and documenting the quality of household data in the Norwegian Labour Force Survey.

This technical report is intended for institutions and researchers using data from the Labour Force Survey.

The project is partially financed by EUROSTAT grant no. 10201.2008.001-2008.5

Abstract

This is a technical document describing quality analysis of survey data at the household level. The Norwegian Labour Force Survey uses registered nuclear family as the sampling unit, while actual household is the required unit in data transmitted to EUROSTAT. The purpose of this project was to evaluate how the sampled unit serves as a proxy for the actual household, and what improvements can be made.

Register-based data sources have been linked to survey data at the micro level, in order to compare household identification and classification in detail. Some macro level comparison has been made for the labour market indicator “workless households”.

Some main results and conclusions:

- The response survey sample is biased with respect to household size, mainly due to higher nonresponse rate among one-person-households.
- The main divergence between various data sources for household information is the classification of 1- and 2-person-units.
- Comprehensive register-based household data differs considerably from the sampling unit in the LFS, which is ‘registered nuclear family’. On the other hand, the more comprehensive data require longer production time and thus not considered as an ideal sampling frame.
- The survey-based household identification module consists of a set of questions asked to a subsample. This considerably improves the quality of household size distribution. Specifically, many registered ‘one-person-households’ are identified as defacto couples.

Statistics Norway is in the process of developing new IT-solution for interviewing the defacto household members who are not in the sampling frame, only identified by the survey. Currently, they are identified but not interviewed about their labour market situation.

Contents

Preface	3
Abstract	4
1. Introduction	6
1.1. Background	6
1.2. Project plan and structure of this report.....	6
2. Data sources	7
2.1. LFS sample survey data.....	7
2.2. Registers with household information.....	8
2.3. Register data with employment information	9
2.4. Linking.....	10
3. Comparison of LFS family and register-based household	11
3.1. Coverage and consistency of current data	11
3.2. Improvement by linking data	12
3.3. Sampling and nonsampling errors.....	13
4. Development	17
4.1. Sampling	17
4.2. Identification	19
4.3. Estimation	21
5. Quality improvements	25
5.1. Household identification	25
5.2. Labour market data at the household level	26
5.3. Provisional improvement in data delivery	26
6. Dissemination	27
6.1. Micro data	27
6.2. Publication.....	27
List of tables	28

1. Introduction

1.1. Background

This report from Statistics Norway present a project aimed at improving household data in the LFS (Labour Force Survey). Specific goals were to document the current data quality and to develop methods for improving LFS-based estimates at the household level. A further goal is to transmit data to the EUROSTAT in compliance with EU Council Regulation (EC) 577/98.

The project is partially financed by EUROSTAT grant no. 10201.2008.001-2008.530

At present, Statistics Norway publishes no regular statistics at the household level, based on LFS data. Based on registers, Income statistics for households is published regularly.

Although several quality aspects of the LFS have been investigated, LFS-based household data have until recently been more of an unknown territory.

1.2. Project plan and structure of this report

The overview refers parts of this report to the planned actions and expected results of the project:

Project actions and expected results	Report part
General information	Chapter 1: introduction
	Chapter 2: data documentation
1. Documented knowledge on:	Chapter 3
a. How well the current LFS covers the household concept and statistics.	Section 3.1
b. How much improvement can be gained by using additional data sources.	Section 3.2
2. Development and testing of technical solutions for:	Chapter 4
a. New sampling methods.	Section 4.1
b. New identification methods.	Section 4.2
c. New estimation methods.	Section 4.3
3. Quality improvements	Chapter 5
a. Better household definition and identification.	Section 5.1
b. Better data quality of labour market variables for household members.	Section 5.2
4. A final report on the project work and results.	
5. Compliance with Council Regulation (EC) 577/98, i.e. the delivery of the required variables for all members of the household to EUROSTAT	Chapter 6 Section 6.1
General statistics	Section 6.2

2. Data sources

2.1. LFS sample survey data

Description of the current data

The sampling frame is 'residents between 15 and 74 years old' and the data sources are updated files from administrative registers. More details about register data sources are discussed in the next section.

The LFS sampling design is a stratified one-stage cluster sampling of families. Systematic sampling stratified by regions is used, with some oversampling of population-sparse regions. The sample probability is the same for all families within each stratum. Since the whole family is sampled, each family member within a family has the same sample probability as the family. Thus all families in a region have the same sample probability. This means that a regional subsample is expected to be representative with respect to family size.

The unit 'family' is here defined as register-based nuclear family, composed of family members between 15 and 74 years old. Records for people outside this age group are not included, as we expect very little labour market participation among the very young and old. The current definition of family includes single persons, married couples and their children, registered same-sex couples, and children who live with one parent. The age limitation means that 'children' denote offspring that are at least 15 years old. This means for instance, that a family classified as 'mother and child' can have one, two or more records, depending on the number of children and their age. However, information about all children in each family is obtained by linking LFS data to register data.

Administrative registers is also used in the post-survey process for estimation. Slightly different versions are used for sampling and estimation.

Discussion of quality

One major concern is the correspondence between the international household concept and the current sampling unit in the Norwegian LFS, which is register-based nuclear family unit. Hitherto, it has been more or less taken for granted that the family unit largely overlaps the actual household. Nevertheless, there are known divergences that could affect the quality of LFS-based statistics at the household level. The following list points out several different reasons why LFS response data may not be representative with respect to the labour market situation for households:

1. Coverage error, cases where the sample data fail to identify all and only the household members.
 - a. Under-coverage, people belonging to the actual households but not to the registered nuclear family unit:
 - i. Cohabiting, unmarried couples.
 - ii. Households that consists of two or more family nuclei.
 - iii. Other relatives living in the household.
 - iv. Other, non-related, household members.
 - b. Over-coverage, people who are registered with the nuclear family but are not members of the actual households:
 - i. Students not living at home, but registered at their parents' address.
 - ii. Other cases.
2. Time lag, due to administrative or technical matters.
 - a. People actually moved in or out of the household between sampling and registration.
 - b. People moved in or out of the household between the sampling and interview.

3. Nonresponse bias, that people who answer are not representative of the whole population. The response probability is associated with family size and with employment status.
 - a. Single-person families are underrepresented in the response sample.
 - b. Employed people are overrepresented in the response sample.

Even though this list contains several important and fundamentally different errors, we are not overly pessimistic about the actual quality for our purpose. We focus on whether the LFS can give reasonable information about the labour market situation for households. Details about the number and distribution of household types etc. is already taken care of by register-based statistics.

2.2. Registers with household information

Administrative registers is already used in the ordinary LFS production process, for sampling, interviewing and estimation. In this project we utilize auxiliary information from multiple administrative registers, and for several purposes. We want to check the quality of the survey data, both at a micro level and aggregated figures. Furthermore, we compare properties of different register-based data sources. Finally, we use register data to make alternative estimates, based on various methods and data sources.

Statistics Norway has several sources of register-based micro data with household information. These are the most relevant data sources:

- **Register families:** register-based data used in the LFS production. 'Register families' corresponds mostly to 'formal households'.
- **Formal households:** data used for production of the official Population Statistics on Families and Households.
- **Defacto households:** data used for production of the official Income Statistics for Households.

More on household data for the Income Statistics

The "defacto" household units correspond closest to national and international definitions, but also have the longest production time.

The following describe important points in the production process:

- Data from the central population register is combined with the property register, which contains exact addresses for private residences. By exact address we mean identifying each dwelling also in multi-dwelling buildings.
- The exact address information is used for identifying households as people who share the same dwelling. In particular, it helps identify households that consist of more than one nuclear family and cohabiting unmarried couples.
- Special consideration is given to students with the same registered address as their parents, but possibly living in their own household. Additional data are used for classifying students as a separate household or included in the parent household:
 - The distance between the place of study and parents' home.
 - The distance to the place of work, for students with jobs.
- Even more data sources are used for identifying unmarried couples, who are probably cohabiting:
 - Data on common children.
 - Tax information revealing joint debt or joint holdings.
 - Subsequent data about marriages for "backcasting".
- Data from the welfare services are used for identifying people living in institutions as not belonging to private households.

Discussion of quality There are known to be errors in the address data, and consequently errors in the grouping of households based on this data. One major problem is missing exact dwelling for people living in multi-dwelling houses. People defined as not having a unique address, usually have registered only street name and number of the building. In the current register, exact address for dwellings in a block of flats or apartment building is a code number for each dwelling. The registration of exact dwelling for people living in multi-dwelling houses has taken its time, and is not yet complete. Registration rate is improving, but it remains a problem that data is not missing at random. In January 2005, seven percent of the population did not have a unique exact address; in Oslo the figure was 22 percent. By March 2010 the corresponding figures were 2.95 and 5.79 percent. We already know that in the LFS, Oslo residents have a higher nonresponse rate, and are somewhat undersampled. This only adds to the difficulties for the data quality in this region.

As mentioned, data revision and linking of extra data sources increases the production time for the defacto household data. If we use this register for sampling or estimation, we must balance production time considerations with the fact that time lag introduces some household-identification errors. In addition, the final statistical register made from many different data sources probably inherits some errors from the base registers. For instance, data on joint children and joint debt are only indicators of cohabitation, not exact evidence.

2.3. Register data with employment information

In addition, we utilize registers with employment information at the individual level. There are notable differences in the operational definition of employment between LFS and register, which causes some divergence at the micro level. Employment in the LFS is classified according to the ILO definition, as minimum 1 hour of paid work in the reference week. The bulk of register-based employment data is the central employee register, where the minimum limits for registration are 4 hours of work and at least 6 days of employment. This register is therefore augmented with more administrative data: wage records cover small jobs, and tax records identify self-employed. The properties of the register-based employment data can lead to differences in workless household rate estimates. We use two main register-based data sources in this project:

Employment Register 1 This register is made every month and is linked to the LFS data at the individual level, in the regular production process. It is a full-count based on employee- and tax registers, and contains an employment status for every resident, independent of the LFS survey. The register-based employment status is used as an auxiliary variable for post-stratification estimation.

Employment Register 2 This more comprehensive statistical register is produced once a year, and is used for official statistics and micro data for research purposes. Several more registers are merged and processed than “register 1”. The additional sources are mainly wage records and newer tax data, used for adding and checking records, and adding auxiliary variables. The resulting data source has an employment classification more correlated with the LFS-based employment status, than “Register 1” has.

The main drawbacks are long production time and less frequent updates. Regarding the LFS, this source is mainly used for quality control. Although not a data source used for short-term statistics, this register may be very relevant for making structural statistics for households.

Linking of administrative registers to sample data result in files that have two variables for employment status. Nevertheless, the survey-based employment status is considered the most correct at the micro level, so in data delivery it is the obvious choice.

Table 2.1 shows the association between survey-based and register-based employment at different levels. It should be remarked that systematic misclassification detracts some of the validity of using linear correlation coefficient as a measurement of this association. Nevertheless, all results points to the superiority of register “type 2” at the micro level.

Table 2.1. Association between survey-based employment status and register-based employment status. 2006–2009. Correlation coefficient.

		Employment individual level	Employment family level	Workless family
2006	Register 1	0.73	0.87	0.77
	Register 2	0.79	0.91	0.82
2007	Register 1	0.71	0.86	0.75
	Register 2	0.77	0.91	0.81
2008	Register 1	0.71	0.86	0.77
	Register 2	0.78	0.91	0.81
2009	Register 1	0.74	0.87	0.79
	Register 2	0.80	0.91	0.82

2.4. Linking

Identification of people is consistent between all administrative register data as well as survey data used in this project. Each resident has a unique eleven-digit code number, which is used for identifying the person across all data sources. This makes it possible to link sample data to register data at the individual level, in practice linking records that have identical code numbers. Families and household are usually identified by a reference person. An often used, though not universal, convention is that the oldest person is the reference person.

In the LFS data, families are identified by a code number corresponding to the personal ID of the reference person in each family. Similarly households are identified by a reference person in each household in the household data. Family- and household ID can be used for linking data sources at this level. However, great care should be taken in keeping track of the different identification variables, and at what level data sources are linked. When linking two different household-data sources, one person can seem to belong to two different households. For instance, two persons can be considered single in the LFS (have different family number); while the register identifies the same people as an unmarried couple (have same household number). In this instance the family number is the personal ID of each single person, while the personal ID of the oldest one identify the couple in the household register. Also in other cases the household identification can diverge in the two sources, for instance when changes in the household composition affect who is the oldest member of the household. If we link data at the household level, using household ID, a proportion of records would remain unlinked, corresponding to the people who are grouped and/or classified differently between the data sources. This would require another analytical approach, than we have chosen in this project. Here, we mainly link data at the individual level, not household level. Most figures that are presented are the number of people, or percentages that are relative proportions of people, not households.

People, families or household with missing ID is not linked at all, at the level for which the ID is missing. This is not a major quality problem if we use updated data, as a very small proportion of the sample lack family- or household ID in the base registers.

3. Comparison of LFS family and register-based household

According to the definition in the Household Statistics, a household consists of persons that are permanently resident in the same private dwelling or institution. The Income Statistics define a household as all persons who live permanently in the same dwelling and having common housekeeping, and include only persons in private households. From our perspective, the interesting unit is people who share housework and household expenses, and other relations that affect their labour market situation.

On the other hand, the registered nuclear family unit is a convenient sampling unit. It is unambiguously defined, readily available from register and has great benefits when it comes to contacting and interviewing. This family unit has come to be regarded as a reasonable proxy for actual household unit. In this chapter we show actually how the LFS family covers the household, at the micro level. In order to achieve this we have linked survey data to the household data, at the individual level.

3.1. Coverage and consistency of current data

Size We compare unit size between data sources in the following way. First, we count the number of people between 15 and 74 years old that has the same family ID in the sample, and group them. This number we call 'family size' and link this property to each individual. Then we count the number of people between 15 and 74 years old that has the same household ID in the register, and group them. This number we call 'household size' and link this property to each individual. Finally, we link these two properties for every person in the sample. '

Table 3.1 show the cross classification of size, by the survey family and the register household. The percentages along the diagonal represent the consistency or 'agreement' between the classifications at the micro level. Percentages outside the diagonal represent disagreement between the two definitions. Overall, the consistency can be said to be reasonable high, between around 60 percent and up to well over 80 percent. Notice that the disagreeing percentages are not symmetrical around the diagonal. Many more people are classified in a larger household than family, than vice versa.

Table 3.1. Family size by household size. LFS 2009. People 15–74 years old. Per cent

	Percent relative to family size				
	Total households	1	2	3	4 or more
Total families	100.00	20.99	47.02	19.53	12.46
1	100.00	52.08	39.65	6.12	2.16
2	100.00	2.93	84.59	10.58	1.89
3	100.00	6.83	14.38	66.01	12.78
4 or more	100.00	11.92	5.36	18.12	64.60
	Percent relative to household size				
	Total households	1	2	3	4 or more
Total families	100.00	100.00	100.00	100.00	100.00
1	32.88	81.57	27.73	10.30	5.69
2	36.42	5.09	65.53	19.73	5.54
3	16.92	5.51	5.18	57.19	17.35
4 or more	13.78	7.82	1.57	12.79	71.42

Type The register data contains detailed classification of family types and household types, according to the number of adults, their age and marital status; and number and age of children. We link this information at the individual level, to the number of people in each family. From the resulting data we present a cross classification of family type (based on register) and family size (based on the survey). This reflects the consistency at the micro level, and gives some insight into the disagreement presented in the previous table.

Table 3.2 shows the percentages of each family type by family size, and vice versa. Table 3.3 gives analogous results for household type. Note that there is not a simple “diagonal agreement” interpretation in these tables. The most important results regard smaller units. Around 2/3 of people between 15 and 74 years old live in families or household with only one or two people in this age group. Over 80 percent of those classified in the register as one-person family or household, are identified as one-person-family in the LFS. However, about half of the one-person-families in the LFS are classified as other family or household types. One of the major shortcomings of the LFS family unit is the failure to identify cohabiting unmarried couples.

Table 3.2. Family size by family type. LFS 2009. People 15–74 years old. Percent.

	Percent relative to register family type				
	Total	1	2	3	4 or more
Total	100.0	32.9	36.4	16.9	13.8
One-person family	100.0	80.5	5.2	5.9	8.4
Couple with small children	100.0	37.6	51.5	7.6	3.4
Couple with older children	100.0	11.7	24.1	28.6	35.6
Lone parent with children	100.0	37.4	40.1	20.5	2.0
Married couple without children	100.0	5.0	84.3	8.5	2.2
Cohabiting couple without children	100.0	90.7	3.6	3.3	2.4
Parent or couple with adult offspring	100.0	6.8	24.5	45.1	23.6
	Percent relative to survey family size				
	Total	1	2	3	4 or more
Total	100.0	100.0	100.0	100.0	100.0
One-person family	20.0	49.0	2.9	7.0	12.1
Couple with small children	14.1	16.1	19.9	6.4	3.5
Couple with older children	23.3	8.3	15.5	39.4	60.3
Lone parent with children	5.1	5.8	5.6	6.1	0.7
Married couple without children	20.8	3.2	48.2	10.4	3.4
Cohabiting couple without children	5.6	15.4	0.6	1.1	1.0
Parent or couple with adult offspring	11.1	2.3	7.5	29.6	19.1

Table 3.3. Family size by household type. LFS 2009. People 15–74 years old. Percent.

	Percent relative to register household type				
	Total	1	2	3	4 or more
Total	100.0	32.9	36.4	16.9	13.8
One-person-household	100.0	79.8	5.2	6.2	8.8
Couple without children	100.0	23.5	67.1	7.1	2.3
Couple with small children	100.0	37.9	51.3	7.5	3.3
Couple with older children	100.0	11.7	23.9	28.5	35.9
Lone parent with children	100.0	36.8	40.6	21.2	1.4
Parent or couple with adult offspring	100.0	6.7	24.4	45.3	23.7
Multifamily household without children	100.0	50.6	31.2	11.5	6.8
Multifamily household with small children	100.0	32.7	42.0	21.6	3.7
Multifamily household with older children	100.0	31.6	31.6	16.5	20.4
	Percent relative to survey family size				
	Total	1	2	3	4 or more
Total	100.0	100.0	100.0	100.0	100.0
One-person-household	18.5	44.8	2.6	6.8	11.9
Couple without children	25.5	18.2	46.9	10.7	4.3
Couple with small children	13.6	15.6	19.1	6.0	3.2
Couple with older children	22.7	8.1	14.9	38.2	59.1
Lone parent with children	4.7	5.3	5.3	5.9	0.5
Parent or couple with adult offspring	10.6	2.2	7.1	28.4	18.3
Multifamily household without children	2.1	3.2	1.8	1.4	1.0
Multifamily household with small children	1.2	1.2	1.4	1.5	0.3
Multifamily household with older children	1.0	1.0	0.9	1.0	1.5

3.2. Improvement by linking data

We want to assess the improvement we can gain by linking register data to the LFS. For a simple but effective overview, we compare the distribution of family size with household size in the sample and in the total register. Table 3.4 shows first the three different distributions, the figures represent the proportion of people that live in families or households of each size. The second part of the table shows the difference between the full count and the two sample figures. The register-based full count is the benchmark which we use to assess the data quality.

The results suggest a marked improvement for households with one or two people. Using the ordinary LFS family-ID will result in a considerably overestimation of one-person households. Closely connected with this is the near equal underestimation of two-people households.

There are also some small differences between sample figures based on household-ID and full counts based on the same ID. This outcome, in addition to more details on register data, will be explained and analyzed further in the next subsection.

Table 3.4. Family/household size by data source. LFS and Register 2009. People 15–74 years old.

	Percent of people by family/household size		
	Sample families	Sample households	Full count of households
Total	100.0	100.0	100.0
1	35.4	23.3	24.7
2	34.5	46.1	45.2
3	16.4	18.9	18.0
4 or more	13.7	11.8	12.2
	Percentage point difference from full count		
	Sample families	Sample households	Full count of households
Total	0.0	0.0	0.0
1	10.7	-1.4	0.0
2	-10.7	0.9	0.0
3	-1.6	0.9	0.0
4 or more	1.5	-0.4	0.0

3.3. Sampling and nonsampling errors

In this section, we look at some more details: the distribution of household types, assess the effect of sampling design and nonresponse, and compare two different registers containing household data. The results are presented as distribution figures, measured in percentage of individuals in each household type. In practice, the figures are calculated from the number of people belonging to a household type divided by the total number of people. Difference between percentages for the total sample and response sample indicates nonresponse bias with respect to household type.

Using formal household data

The “formal household” (family) data cover most of the one-family households, and this data source is usually the most updated and readily available.

In Table 3.5 we compare the distribution of people by family type in the LFS response sample with the total sample. The results reveal a nonresponse bias that is fairly consistent over the four-year period. One-person families are under-represented by about 2.7 percentage points (average difference between “response” and “total”). Married couples with children are overrepresented by around 2.1 percentage points. Married couples without children are overrepresented by around 1 percentage points.

We observe also some difference between “total sample” and “register”. This difference is due to some systematic effect from the sampling design, in addition to random events. The effect is smaller than the nonresponse error, but has the same direction for the most affected family types. As a consequence, the divergence between the response sample and the register is greater than between the response sample and the total sample.

Table 3.5. Distribution of 'formal household' type. Response-, total sample and register. LFS 2006–2009. People 15–74 years old. Percent

	Response	Total	Register
2006			
Total	100.0	100.0	100.0
1 One-person-family	21.6	23.9	24.9
2 Married couples without children	17.7	16.8	17.1
3 Married couples with children	42.7	40.9	39.3
4 Mother with child	8.1	8.4	8.5
5 Father with child	2.1	2.1	2.3
6 Cohabiting couples with common children	7.9	7.9	7.8
7 Registered partnerships without/with children	0.1	0.1	0.1
2007			
Total	100.0	100.0	100.0
1 One-person-family	22.3	24.8	25.5
2 Married couples without children	17.5	16.6	17.1
3 Married couples with children	41.0	39.2	38.4
4 Mother with child	8.5	8.7	8.6
5 Father with child	2.4	2.4	2.4
6 Cohabiting couples with common children	8.2	8.2	8.1
7 Registered partnerships without/with children	0.2	0.2	0.1
2008			
Total	100.0	100.0	100.0
1 One-person-family	22.0	25.1	26.2
2 Married couples without children	17.9	16.7	17.2
3 Married couples with children	41.8	39.2	37.4
4 Mother with child	7.6	8.1	8.5
5 Father with child	2.3	2.4	2.4
6 Cohabiting couples with common children	8.3	8.4	8.3
7 Registered partnerships without/with children	0.1	0.1	0.1
2009			
Total	100.0	100.0	100.0
1 One-person-family	22.6	25.5	26.6
2 Married couples without children	18.7	17.5	17.1
3 Married couples with children	40.2	38.0	36.9
4 Mother with child	8.0	8.3	8.5
5 Father with child	2.2	2.3	2.4
6 Cohabiting couples with common children	8.3	8.3	8.4
7 Registered partnerships without/with children	0.1	0.1	0.1

Using defacto household data

Again we compare survey- and register data, now using the ostensibly 'best' register type. In addition we compare two register data sources. The register figures are full-count in the sense that they cover the population between 15 and 74 years old, for comparison with the LFS sample restricted by the same age limits.

First we compare the response sample with the total sample, and assess the effect of nonresponse bias. Table 3.6 show the distribution of household categories in percent. The classification of households used here is a slightly aggregated version. The main difference from the family classification is the several types of households with more than one family.

The results show that one-person-households are underrepresented in the response sample by about 2 percentage points. Couples with no children living at home, are overrepresented by close to 1 percentage point. Couples with children between 6 and 17 years old, are overrepresented by around 1.3 percentage point. Other groups are either small or show little differences. Households with more than one family cover less than 5 percent of the total population. Multifamily households seem to be fairly well represented in the response sample, compared with the total sample and the register.

Table 3.6. Distribution of 'defacto'-household type. Response-, total sample and register. LFS 2006–2009. People 15–74 years old. Per cent

	Response	Total	Register
2006			
Total	100.0	100.0	100.0
Living alone	17.6	19.4	20.0
Couple without children	24.1	23.4	23.9
Couple with small children	14.1	14.0	14.0
Couple with older children	23.1	22.1	21.5
Lone parent with children	4.8	5.0	5.0
Parent or couple with adult offspring	11.8	11.6	11.1
Multifamily household without children	2.4	2.5	2.5
Multifamily household with small children	1.0	1.1	1.2
Multifamily household with older children	0.9	1.0	1.0
2007			
Total	100.0	100.0	100.0
Living alone	18.1	19.9	20.4
Couple without children	24.3	23.7	24.0
Couple with small children	13.8	13.6	13.8
Couple with older children	23.7	22.6	21.5
Lone parent with children	5.1	5.3	5.0
Parent or couple with adult offspring	10.8	10.6	10.5
Multifamily household without children	2.1	2.3	2.5
Multifamily household with small children	1.2	1.2	1.3
Multifamily household with older children	0.9	0.9	1.0
2008			
Total	100.0	100.0	100.0
Living alone	17.9	20.1	20.8
Couple without children	24.9	24.0	24.1
Couple with small children	14.2	13.9	13.6
Couple with older children	23.5	22.0	21.2
Lone parent with children	4.3	4.8	5.1
Parent or couple with adult offspring	11.0	10.6	10.4
Multifamily household without children	2.2	2.4	2.5
Multifamily household with small children	1.1	1.1	1.3
Multifamily household with older children	1.0	1.0	1.0
2009			
Total	100.0	100.0	100.0
Living alone	18.5	20.5	21.1
Couple without children	25.5	24.5	24.1
Couple with small children	13.6	13.6	13.5
Couple with older children	22.8	21.4	20.9
Lone parent with children	4.7	5.0	5.0
Parent or couple with adult offspring	10.6	10.3	10.5
Multifamily household without children	2.1	2.4	2.6
Multifamily household with small children	1.2	1.2	1.4
Multifamily household with older children	1.0	1.0	1.1

One of the most prominent differences between the two register-based data sources is the distribution of one-person units. Around 25 percent of the population of formal family units consists of one person; while the corresponding figure for defacto households is around 20 percent. Conversely, couples without children comprise nearly 24 percent in the defacto-data and just over 17 percent in the formal data. As mentioned in chapter 2, one of the main features of the defacto-households is the reclassification of singles into cohabiting couples.

We now turn to a direct comparison of the classification at the micro level. The two register-based data sources are linked at the individual level and table 3-7 show the cross-classification of family type and household type. Nearly 10 percent of those classified as one-person-families, are classified as other household types (in addition to unclassified due to non-linking of the employment register). Over 60 percent of multifamily households with no children are classified as one-person-families. For lone-parents there is some slight divergence, and for other types the classifications are very close. In all, the use of the defacto-household data will have most impact on statistics for one-person-units.

Table 3.7. Classification of family type and household type. Register data 2009. People 15–74 years old. Percent.

	Percent of family type							
	Total	One-person family	Couple with small children	Couple with older children	Lone parent with children	Married couple without children	Cohabiting couple without children	Parent or couple with adult offspring
Total	100.0	23.2	14.0	21.5	5.4	18.8	6.1	11.0
One-person-house-hold	100.0	100.0
Couple without children	100.0	75.1	24.9	.
Couple with small children	100.0	.	100.0
Couple with older children	100.0	.	.	100.0
Lone parent with children	100.0	.	.	.	100.0	.	.	.
Parent or couple with adult offspring ...	100.0	100.0
Multifamily household without children	100.0	60.8	.	.	.	21.2	2.8	15.2
Multifamily household with small children	100.0	13.6	41.6	8.8	11.7	13.8	0.8	9.6
Multifamily household with older children	100.0	17.7	.	50.2	21.0	8.2	0.4	2.6

	Percent of household type							
	Total	One-person family	Couple with small children	Couple with older children	Lone parent with children	Married couple without children	Cohabiting couple without children	Parent or couple with adult offspring
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
One-person-household	21.0	90.3
Couple without children	24.0	95.6	98.5	.
Couple with small children	13.5	.	96.0
Couple with older children	20.8	.	.	97.0
Lone parent with children	5.0	.	.	.	92.9	.	.	.
Parent or couple with adult offspring ...	10.4	95.0
Multifamily household without children	2.6	6.7	.	.	.	2.9	1.2	3.6
Multifamily household with small children	1.4	0.8	4.0	0.6	2.9	1.0	0.2	1.2
Multifamily household with older children	1.1	0.8	.	2.5	4.1	0.5	0.1	0.3

Household size Comparing the distribution of families and households between register and survey is not straightforward since the classification of families and households are different and there is no independent classification based on the survey. In order to make a comparison of all the data sources, we compare distribution of unit size, not type. Size is defined as the number of people between 15 and 74 years old having the same unit identification.

Table 3.8 present percentages calculated as the number people in specific unit size divided by the total number of people in the age group in the data source. The register data source here is the “defacto household register”, since we consider this to have the most correct household size.

Results:

- Household with one person are:
 - Underrepresented in response sample compared with the total sample.
 - Considerably overrepresented in the total sample compared with the household register.
 - Overrepresented in response sample compared with the household register.
- Household with two people are:
 - Overrepresented in response sample compared with the total sample.
 - Considerably underrepresented in the total sample compared with the household register.
 - Underrepresented in response sample compared with the household register.
- For other households the effects are similar, but relatively small.

Conclusions:

- The overrepresentation and underrepresentation are opposite in households with one and two people.
- The smaller households constitute the biggest problem.
- For households with one or two people, the nonresponse error counteracts some of the effect of the sampling design. This is not to say that the resulting quality is especially promising, as the standard error increases as the effective sample size decreases. In addition it may result in bias with respect to other important characteristics than we have had the opportunity to investigate.

Table 3.8. Household size by time and data source. People 15–74 years old. Percent. 2006–2009

	Response sample	Total sample	SAMPLE			FULL COUNT		
			Register family	Formal household	Defacto household	Register family	Formal household	Defacto household
2006								
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1	31.5	33.0	36.9	38.0	37.8	36.9	29.1	23.8
2	35.1	33.9	34.9	36.0	36.1	35.2	38.3	45.3
3	18.5	17.8	16.5	16.6	16.6	15.7	17.4	18.7
4 or more	14.9	15.3	11.6	9.5	9.5	12.3	15.2	12.2
2007								
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1	32.2	33.9	38.1	38.7	38.5	37.7	29.6	24.1
2	35.4	33.9	34.3	36.0	36.0	35.1	38.2	45.4
3	17.9	17.4	15.8	15.9	16.1	15.4	17.2	18.4
4 or more	14.5	14.9	11.8	9.4	9.4	11.9	15.0	12.1
2008								
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1	31.9	34.6	38.4	39.4	39.3	38.2	30.2	24.5
2	36.0	34.1	34.5	36.0	36.0	34.9	38.3	45.3
3	18.1	17.1	15.8	15.2	15.4	15.1	16.9	18.2
4 or more	14.1	14.1	11.3	9.3	9.4	11.8	14.7	12.1
2009								
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1	32.9	35.4	38.8	40.0	39.9	38.5	30.6	24.7
2	36.4	34.5	35.0	36.4	36.4	34.9	38.2	45.2
3	16.9	16.4	15.0	15.0	15.2	14.9	16.6	18.0
4 or more	13.8	13.7	11.2	8.6	8.6	11.7	14.7	12.2

4. Development

4.1. Sampling

Firstly, we discuss the data source used for sampling and sampling procedure. Currently, the ordinary LFS sampling procedure uses the most updated file from the Central Population Register, governed by the Tax Administration. Statistics Norway regularly receives a copy of this register, and a new file is used at every sampling. Information that is mandatory to report, and affect the household composition, include births, marriages, deaths, immigration, moving and emigration. We assume that the data quality generally is very high. Records of births, marriages and deaths are transmitted from official sources and probably have the best quality. Illegal immigration is estimated to be very small. Delayed or missing reports on moving and emigration are probably the most important source of error.

The LFS sampling design is somewhat unusual in that we call it cluster sampling, but has only one stage. The sampling unit is a cluster and we sample the whole cluster, and do not go through the usual second stage of sampling within clusters. The result is a sample of complete families. The sampling is stratified by geographical regions only, and the sample probability is the same for all families

within each stratum. Within each region, a subsample is expected to be representative with respect to family size.

In all, to the extent that family is a good proxy for household, the data source and the sampling procedure should insure high data quality. However, our results have revealed notable divergence between the LFS family definition and household data, for major groups. In principle, errors should be avoided as early as possible in any survey, and especially in a long and complex production process such as the LFS. Nevertheless, changing the sampling frame data source or sampling design could introduce other types of errors, and possibly more serious errors.

If the defacto-household data were to be used in sampling, time lag could be a major source of error. Though many households are stable with respect to identification and classification, the time-consuming production process and infrequent updates could affect the quality. Since this production depends on several other data sources, there is probably little to be done about the production of this data in itself. A possible compromise could be to amalgamate household data with more the updated family file from the Central Population Register. The resulting heterogeneous data source should be thoroughly scrutinized before we change the sampling process.

To get some idea about the effect of time lag, we link household register data at the individual level for two consecutive years. This enables us to calculate how many people change their household ID and/or household type.

Table 4.1 shows the stability and flows between household types, measured as percent of people between 15 and 74 years old.

Table 4.1. Household type 2009 by household type 2008. People 15–74 years old. Per cent

2008/2009	Total	One-person household	Couple without children	Couple with small children	Couple with older children	Lone parent with children	Parent or couple with adult offspring	Multi-family household without children	Multi-family household with small children	Multi-family household with older children
Total	100.0	21.3	24.4	13.5	20.0	4.8	10.6	2.6	1.4	1.0
One-person-household	100.0	86.4	6.2	1.2	1.2	0.6	2.6	1.1	0.3	0.2
Couple without children	100.0	2.8	89.7	4.3	0.2	0.1	1.8	0.8	0.3	0.1
Couple with small children	100.0	1.3	0.2	85.2	10.5	1.7	0.1	0.1	0.7	0.1
Couple with older children	100.0	3.1	0.8	1.2	84.0	1.8	7.8	0.3	0.3	0.7
Lone parent with children	100.0	3.8	0.7	2.8	5.7	76.3	8.6	0.5	0.6	0.8
Parent or couple with adult offspring	100.0	11.0	13.4	0.5	0.3	0.2	72.2	1.6	0.7	0.1
Multifamily household without children	100.0	12.4	8.6	1.2	0.6	0.3	2.9	70.5	2.5	0.6
Multifamily household with small children	100.0	3.9	4.0	9.5	2.6	2.9	2.2	1.5	67.2	6.1
Multifamily household with older children	100.0	4.9	2.8	0.7	8.6	4.4	2.0	8.7	2.8	64.9

Table 4.2 illustrates the association between changes in household identification and changes in household types. The figures are calculated from people who are registered, and between 15 and 74 years old, in both 2008 and 2009. Overall, most people have the same household identification (92 percent) and are in the same household type (nearly 80 percent). Nearly 98 percent who changed household ID also change household type. 14 percent remain in the same household, while the household changes type.

Table 4.2. Changes in household type by changes in household identification. People 15–74 years old. Per cent. Register 2008–2009.

	Panel	Changed type	Same type
Panel	100,0	20,4	79,6
Changed ID	100,0	97,9	2,1
Same ID	100,0	14,1	85,9
	Panel	Changed type	Same type
Panel	100,0	100,0	100,0
Changed ID	7,6	36,3	0,2
Same ID	92,4	63,7	99,8

4.2. Identification

It has of course been known from the start that there was some divergence between registered nuclear family and actual household. As a result of growing interest about the labour market situation for households, the LFS protocol was redesigned in 2006 in order to identify the actual household. Several questions were included to identify household members who are not family members, and vice versa. These extra questions are asked only at the first and last interview, of eight, thus the subsample comprise about ¼ of the total sample. The following subsections describe the identification process, and discuss challenges and possible developments.

Protocol First, the names of all registered family members are read out loud. Then the interviewer asks explicitly for each member (as each name is read out loud again) whether he or she lives in the dwelling. This makes it easy (at least in principle) to remove people who do not belong to the household, as they are already registered and identified.

The next step is designed to record actual household members that are not part of the registered nuclear family. To identify them, name and birth date is recorded. In order to classify them as part of the household, there are questions about the connection and/or relations between the sampled subject and each of the extra household members. For instance cohabiting couples are regarded to belong to the same household, while lodgers are not if they have separate entrance to their part of the dwelling.

Discussion The identification process of the additional household members has met with difficulties due to the response and the technical solution for data collection.

During the construction process of the questionnaire, it was assumed that a substantial proportion would refuse or not be able to provide the official personal ID for other household members. As a consequence, only name and birth date is recorded. Still, a number of people refuse or are unable to provide names and/or birth dates. Even with perfect information on name and birth date, the survey data can't be linked directly to the register-based data. Due to limited resources, the cross-linking of names and birth dates with personal ID has yet to be performed. As there is no time limit for using these data for research purposes, this linking process can be performed at a later stage.

Furthermore, both the identification process and interview of household members are hampered by the current technical solution. In practice, each record in the LFS data corresponds to an interview. The data for removing or adding household members are organized as groups of extra variables linked to the reference person. This makes it more cumbersome to deal with this data, and limits the number of household members to a predetermined number of variables.

In addition to identifying actual household members, it was also the purpose to interview and record information from the extra members (people that was not sampled). However, due to technical problems, these additional household members are not interviewed. Currently, we neither add records nor impute labour

market characteristics for these additional household members. Consequently, no information about them is included in the data transmission to EUROSTAT. A new IT-system for interviewing is being developed, which we hope will resolve most of the problems we have mentioned.

For the time being, we use a simplified procedure to impute household identification and individual attributes pertaining to the household, such as the relations between its members. This simplified procedure uses only sample data containing age, gender and marital status. As a result of this project, we propose to implement a more comprehensive procedure linking register-based household data to the LFS data, and delivering data with better quality at the household level. This is meant to be a temporary solution until the new IT-system is ready.

Results

We present some preliminary results, based on the surveyed households. Table 4.3 show the cross-classification of the actual household size based on the interview against the family size based on register data. Both measurements are people between 15 and 74 years olds. It is clear that the main improvement is the identification of defacto couples that are registered as one-person families. On average just 40 percent of registered one-person families are in fact living alone, according to this survey data. Table 4.4 shows analogous comparison with the register-based “defacto households”. From this data over 20 percent of registered one-person families respond that they actually live in larger households, mainly couples.

One should bear in mind that one-person families are underrepresented in response data, before drawing far-reaching conclusions from this. Also the comparison with register-data is slightly disfavoured by time lag and other technical problems. Nevertheless, the results indicate that substantially improvement can be had from survey-based identification of actual households. Especially the correct classification of one-person families and couples seem to gain some new information from the respondents, which are not available from the other data sources.

Table 4.3. Interview-based household size, by sampled family size. LFS subsample. Average 2006-2009. Percent.

	Surveyed household					
	Total	1 person	2 people	3 people	4 people	5 or more
Sampled family						
Total	100.0	17.2	55.9	17.2	7.9	1.9
1 person	100.0	37.9	54.7	5.7	1.5	0.3
2 people	100.0	2.4	86.7	9.6	1.2	0.2
3 people	100.0	0.8	15.9	71.7	10.8	0.8
4 people	100.0	0.3	6.4	21.4	67.2	4.7
5 or more	100.0	0.1	3.3	10.9	20.0	65.6
	Surveyed household					
	Total	1 person	2 people	3 people	4 people	5 or more
Sampled family						
Total	100.0	100.0	100.0	100.0	100.0	100.0
1 person	42.7	94.4	41.7	14.1	8.1	6.8
2 people	34.5	4.9	53.5	19.2	5.2	2.9
3 people	13.5	0.6	3.8	56.2	18.4	5.7
4 people	7.4	0.2	0.9	9.3	63.5	18.5
5 or more	1.9	0.0	0.1	1.2	4.8	66.2

Table 4.4. Interview-based household size, by register-based “defacto” household size. LFS subsample. Average 2006-2009. Percent.

	Surveyed household					
	Total	1 person	2 people	3 people	4 people	5 or more
Defacto reg.						
Total	100.00	17.06	55.94	17.21	7.91	1.88
1 person	100.00	77.55	19.48	2.09	0.62	0.26
2 people	100.00	8.07	86.92	4.05	0.86	0.10
3 people	100.00	3.94	21.82	64.46	8.67	1.11
4 people	100.00	2.23	8.67	24.38	58.63	6.09
5 or more	100.00	2.86	8.86	10.16	23.89	54.23

	Surveyed household					
	Total	1 person	2 people	3 people	4 people	5 or more
Defacto reg.						
Total	100.00	100.00	100.00	100.00	100.00	100.00
1 person	14.97	68.02	5.21	1.81	1.17	2.08
2 people	55.13	26.08	85.66	12.97	5.98	2.90
3 people	19.07	4.41	7.44	71.44	20.90	11.27
4 people	8.93	1.17	1.38	12.65	66.20	28.91
5 or more	1.90	0.32	0.30	1.12	5.75	54.84

4.3. Estimation

In order to compare alternative estimation procedures, we want to construct a variable of interest that is easy to understand, is of social importance and potentially a candidate for new official statistics. We have therefore chosen ‘workless households’ as the variable of interest throughout this project.

We define a workless household as a household where no one is employed. Workless households are not homes where everybody is unemployed. Those not employed may be unemployed or ‘outside workforce’, aka economically inactive. The indicator corresponds closely to the current EUROSTAT indicator ‘jobless households’, though its definition excludes 18-24 years old students. The reason why we do not exclude young students in this particular analysis is that we do not know the student status of the nonresponse group.

First, we compare estimates of workless household rate based on survey- and register-data of several types. Table 4.5 show five estimates based on the LFS sample survey data linked to employment registers and register-based household classification. The register-based employment status can be linked for every individual, including those who do not respond. This fact means we can assess the effect of nonresponse bias. We calculate one register-based estimate from the total sample and one register-based estimate from the response sample. The difference between these two estimates, both based on the register status, reflects the nonresponse error.

Second, we also compare register-based estimates based on different data sources with the survey-based. This is done in order to assess the variation in quality in the register data themselves, refer to subsection 2.3 for more details on the employment registers.

To sum up: first we use the register-based estimates in order to assess the survey quality, specifically the effect of nonresponse bias. Then we use the survey-based estimate in order to compare the quality of the registers. That means we do not assume the absolute level of workless households in any register is the correct one. On the contrary, we assume that the survey-based employment status is the best of all three. From this premise, and given a large and representative sample, the survey-based indicator for workless households would be best.

Main results (average percentage point differences):

- The response sample underestimates the rate of workless households, for all household types. The overall effect is between 1.1 and 1.3 percentage point.

- The effect is greatest for lone parents with children, with over 2.4 percentage point underestimation.
- For household with more than one family, there is also considerably underestimation, but the results vary more.

The household types with the highest rate of workless households are one-person-households and household with more than one family but no children. These groups may have a relatively large proportion of young people attending education. A high rate of workless households is then not alarming.

Table 4.5. Workless households, by household type. Sample data with survey-employment and register-employment. Per cent. 2006–2009

	Total sample		Survey	Response sample		
	Register1	Register2		Register1	Register2	Survey
2006						
Total	20.75	16.17	16.14	19.42	14.91	16.14
One-person-household	41.68	33.53	34.73	40.26	32.29	34.73
Couple without children	28.52	23.32	25.47	28.03	22.78	25.47
Couple with small children	11.83	7.85	6.17	10.18	6.07	6.17
Couple with older children	5.03	2.98	2.21	3.97	2.23	2.21
Lone parent with children	23.59	19.06	16.6	21.43	16.4	16.6
Parent or couple with adult offspring	6.59	4.54	4.29	6.38	4.04	4.29
Multifamily household without children ..	34.67	27.5	27.79	32.25	25.35	27.79
Multifamily household with small children	23.13	15.3	13.46	22.12	12.98	13.46
Multifamily household with older children	18.06	13.66	16.58	18.18	14.44	16.58
2007						
Total	20.83	15.69	15.76	19.24	14.31	15.76
One-person-household	41.24	32.19	34.82	40.63	31.8	34.82
Couple without children	28.12	23.3	25.39	26.78	21.93	25.39
Couple with small children	11.19	6.24	4.53	9.77	5.06	4.53
Couple with older children	5.29	3	2.31	4.59	2.17	2.31
Lone parent with children	21.32	14.74	13.95	18.89	13.16	13.95
Parent or couple with adult offspring	7.02	4.41	3.36	5.37	3.07	3.36
Multifamily household without children ..	40.22	28.55	27.06	36.47	26.38	27.06
Multifamily household with small children	22.58	17.56	10.78	20.69	12.93	10.78
Multifamily household with older children	20.18	13.9	13.83	17.02	12.23	13.83
2008						
Total	20.2	15.74	16.45	18.99	14.66	16.45
One-person-household	40.7	32.89	36.2	40.7	32.76	36.2
Couple without children	28	23.05	26.51	27.25	22.43	26.51
Couple with small children	9.73	5.88	4.63	9.03	5.01	4.63
Couple with older children	5.01	2.97	2.49	4.46	2.4	2.49
Lone parent with children	19.69	14.29	13.26	16.61	11.3	13.26
Parent or couple with adult offspring	6.74	4.77	4.32	5.46	4.09	4.32
Multifamily household without children ..	31.04	25.57	25.45	29.69	24.78	25.45
Multifamily household with small children	23.25	15.13	14.22	19.91	10.9	14.22
Multifamily household with older children	17.96	15.51	15.59	18.28	15.59	15.59
2009						
Total	20.92	17.38	17.84	19.87	16.49	17.84
One-person-household	42.41	35.78	37.8	42.08	35.43	37.8
Couple without children	27.79	23.87	27.85	27.68	23.82	27.85
Couple with small children	10.69	7.35	5.03	9.47	5.94	5.03
Couple with older children	4.74	3.44	2.73	3.9	3.02	2.73
Lone parent with children	19.95	17.17	14.03	16.58	14.77	14.03
Parent or couple with adult offspring	7.13	5.33	5.14	7.38	5.24	5.14
Multifamily household without children ..	31.21	27.66	29.06	29.54	26.15	29.06
Multifamily household with small children	23.21	17.75	17.65	20.17	15.13	17.65
Multifamily household with older children	17.09	16.67	17.91	17.91	17.41	17.91

Adjustment The previous sections should have made it clear that nonresponse error is a major factor to the quality of LFS-based estimates. In order to demonstrate an adjustment for the nonresponse error, we use the register data to make post-stratified estimates. This method is chosen because it is relatively simple to implement and comparable to the ordinary estimation process. In the ordinary weighting procedure, age, gender, and register-based employment status is used to define post-strata.

Here, we add household type to the definition of strata. In order to get reasonable stratum sizes, aggregation or some other limitation is necessary. Employment and age are associated, and we normally wish to control for age in some way. The youngest and oldest have low employment rate, for natural reasons, and we have chosen to simply exclude some age groups in this part of the analysis. It could be argued that the age limitation makes the indicator a more reasonable measurement of unwanted worklessness.

The presented figures are percentage of people 18–59 years old living in households where no one works. Table 4.6 compare three figures for workless households: sample (unweighted) percentage, post-stratification (I) using employment status and gender, and post-stratification (II) that include also household type.

Some observations:

- The total workless household rate is adjusted upwards, by both methods.
- For one-person-household there is little added effect of introducing household type.
- For couples, the household type method adjusts estimates slightly lower than the simpler method.
- For multifamily-households there is some extra adjustment upwards by introducing household type.

Some conclusions:

- It is plausible that total workless household rate is underestimated because of nonresponse error.
- We assume the weighted estimates overall give a more accurate picture of the situation for workless households, than unweighted.
- For one-person-household most of the bias is already adjusted by using employment status and demographic variables.
- Estimates for other household types will benefit most from adding household type to a post-stratification scheme.

The individual weights calculated in the ordinary LFS production, are more like the ‘pst I’, and give similar results. That means some adjustment “in the right direction” will be achieved simply by using these weights.

Table 4.6. Workless household rate, by household type. People 18–59 years old. Per cent. 2006–2009.

	Sample	PST I	PST II
2006			
Total	9.8	10.9	11.4
One-person-household	24.3	26.9	26.8
Couple without children	8.8	9.8	9.5
Couple with small children	6.5	7.2	7.0
Couple with older children	2.8	3.2	3.1
Lone parent with children	17.1	18.8	19.2
Parent or couple with adult offspring	4.3	4.8	4.9
Multifamily household without children	18.4	20.8	20.1
Multifamily household with small children	11.7	12.8	12.7
Multifamily household with older children	12.9	14.0	13.7
2007			
Total	8.9	10.1	10.6
One-person-household	23.3	25.9	25.6
Couple without children	8.3	9.4	9.0
Couple with small children	4.6	5.2	5.4
Couple with older children	2.7	3.1	2.9
Lone parent with children	14.6	16.4	17.0
Parent or couple with adult offspring	3.8	4.2	4.3
Multifamily household without children	18.1	20.7	20.1
Multifamily household with small children	7.4	8.5	8.9
Multifamily household with older children	9.0	10.3	10.5
2008			
Total	9.1	10.3	11.2
One-person-household	24.4	27.2	27.0
Couple without children	8.4	9.6	9.1
Couple with small children	4.7	5.3	5.2
Couple with older children	2.7	3.0	2.8
Lone parent with children	14.7	16.4	18.6
Parent or couple with adult offspring	4.9	5.5	5.6
Multifamily household without children	14.4	16.5	16.6
Multifamily household with small children	11.1	12.2	12.7
Multifamily household with older children	8.5	9.5	10.0
2009			
Total	10.3	11.6	12.2
One-person-household	26.4	29.1	28.5
Couple without children	9.2	10.4	9.8
Couple with small children	5.0	5.6	5.9
Couple with older children	3.3	3.8	3.6
Lone parent with children	15.5	17.2	16.8
Parent or couple with adult offspring	5.4	6.2	6.4
Multifamily household without children	20.4	22.8	21.7
Multifamily household with small children	15.1	16.6	16.7
Multifamily household with older children	9.9	11.2	11.0

Comment We have used post-stratification because it is a simple and straightforward method that can be used with the available data. In general, post-stratification have limitations such as minimal stratum sizes. This means that the number of useful variables and their aggregation level must be balanced. If we for instance impose stratification with detailed household type, additional variables must be fewer or more aggregated. For other labour market analyses at the household level, alternative methods should be considered.

5. Quality improvements

In this chapter, we sum up and explain in more detail the various improvement proposals based on the results of this project as well as knowledge about the data sources. Throughout this chapter, by ‘extra household members’ we mean actual household members who are not sampled.

5.1. Household identification

We recommend a step-wise approach for optimal data quality from the household identification process. Furthermore, we suggest integration of several data sources.

The following list sums up developments for better household identification.

1. **Sampling frames** should be selected from the most comprehensive data sources, but also the most updated data sources. The household data used for income statistics is the most comprehensive in terms of household definition. On the other hand, the nuclear family data is the most frequently updated. Any data collection used for sampling should be updated with recent data about marriages, deaths, moving and immigration. Using year-old household data would result in sampling people that are not in the actual population. Nevertheless, it could improve efficiency to sample additional (“non-nuclear”) household members by using integrated household data. Identifying them through an interview process takes time, and there is a secondary identification process in order to link data from registers using the personal ID.

Before revising the LFS sampling plan we should settle two major concerns. Firstly, consequences for other labour market statistics must be investigated closer. Secondly, there are practical consequences in terms of IT-solutions etc.

2. **Questionnaire and protocol** are of course important in the collection of data, and as we develop new practical solutions, these factors should be evaluated and to consider possible developments. It could be wise to evaluate the questionnaire and protocol before implementing a major overhaul of the IT-system. We know there are people unwilling or unable to provide an exact identification of extra household members. Names, as well as birth dates are missing to some degree. Development should include improving motivation for providing correct and complete information. Furthermore, though we expect a limited number of people to report the exact personal ID-number, we could provide the possibility by including a voluntary question and allocate space for the response. Another possibility is to pre-link register-based household information, for use as auxiliary information during the interview. A checklist, containing correctly spelled names and personal ID available for the interviewer, could speed up the identification process.

3. **Secondary identification** process in order to link data to the personal ID is done by combining birth date and name, the latter by algorithms utilizing text-functions the resolves spelling variations etc. We have at the moment no suggestions for improvement of this stage, other than reducing the volume as much as possible by the improvements mentioned in point 1 and 2.

4. **Organization of data** should be improved; we mention here only the logical system, not suggesting a technical solution. Currently, the household members are represented by several variables in the same record as the reference person. These variables are for identification- and technical purposes. For the extra household members, no information about labour market characteristics is surveyed. As the number of variables is predetermined, this represents a limitation to the household size, as well

as redundant space for the smaller households. Most important, it has made it impracticable under the current technical solution to survey the extra members.

All actual household members, both sampled and survey-based, should be represented by a record for each person. Similar to an ordinary relational database, a separate data matrix representing household members could be linked to reference person by the personal ID. This would make it more convenient to perform the interview and record the necessary labour market characteristics, as well as make it possible to represent an arbitrary household size.

5. **Post-survey** possibilities include linking the survey-based identification to register-based identification, when more comprehensive household data becomes available that was not available when sampling. Since this takes time, it is most relevant for quality analysis and improvement in the long run.

5.2. Labour market data at the household level

In this project we have chosen employment as an example of labour market attribute and workless household rate as variable of interest. There are of course other characteristics of a household that are important from a labour market perspective, such as allocation of working hours, sharing of income and expenses, unemployment, welfare, decisions about retirement. When suggesting actions for improving labour market data at the household level, we cannot foresee the consequence for all kinds of characteristics. But we assume that actions that can be shown to improve for instance employment data, also improve the overall data quality for labour market characteristics. The following list sums up developments for better labour market data at the household level.

1. **Sampling.** We will study results from other countries about oversampling of certain groups. Oversampling, for instance single-person households, or people registered as not-employed, should be considered.
2. **Response.** Actions for increasing response quality must continue to be developed and implemented. This should be directed at specific groups in order to reduce nonresponse bias. A general increase in response rate may not improve the overall quality.
3. **Processing.** Auxiliary data with optimal quality should be used for the post-survey process. At present we link a simplified short-term employment register. With some adjustment in timing of the base registers, we may achieve a higher correlation between the employment variables. We should also reconsider the use of register-based unemployment information at the micro level.
4. **Estimation.** We propose a separate estimation procedure for statistics at the household level. This entails separate weights to be used for the household variables. The more comprehensive employment register can be used for making structural household statistics, when the long production time and infrequent updates of this data source can be acceptable.

5.3. Provisional improvement in data delivery

Until new IT-solutions can be implemented, some improvements can be made in the micro data with household information.

1. Information about the roles and relations of the household members can be based on linking the survey data to better register-data. Currently this information is imputed, based only on the LFS sample frame.

2. Records for household members younger than 15 or older than 74 years can be added, so that complete data for registered nuclear families can be delivered.

We have decided against a procedure for adding records for “extra household members” based only on register household data and imputation of labour market attributes.

6. Dissemination

6.1. Micro data

Norwegian LFS data is routinely transmitted to the EUROSTAT every quarter. This data contain household data based on the simplified imputation process described earlier.

Until we have a working IT-solution for survey-based household identification and interviewing, linked register-based household information will play an important part in the LFS.

Data based on the provisional procedure, should be possible to start with the first delivery within 12 weeks after 1st quarter 2011.

The delivery date for data based on the new IT-solution will take longer, as we expect the development process to be finished within the first half of 2011. After the IT-development itself some technical implementation and change of production routines will most likely be required. As stated in the Grant agreement, we expect the regular delivery of the data according to the new method to start from third quarter of 2011.

6.2. Publication

We have at the moment no plans for new official labour market statistics at the household level at a regular basis. With the growing interest for the labour market situation for households, analyses and reports about for instance workless households will likely be made at an ad-hoc basis.

Register-based full-counts can give insight into the situation for social subgroups and details about regional differences, and thus complement survey-based analyses. However, the LFS contains many labour market variables not available from other sources and will be important in monitoring the labour market situation for households.

List of tables

Table 2.1.	Association between survey-based employment status and register-based employment status. 2006–2009. Correlation coefficient.	10
Table 3.1.	Family size by household size. LFS 2009. People 15–74 years old. Per cent.....	11
Table 3.2.	Family size by family type. LFS 2009. People 15–74 years old. Percent.	12
Table 3.3.	Family size by household type. LFS 2009. People 15–74 years old. Percent.	12
Table 3.4.	Family/household size by data source. LFS and Register 2009. People 15–74 years old.	13
Table 3.5.	Distribution of 'formal household' type. Response-, total sample and register. LFS 2006–2009. People 15–74 years old. Percent	14
Table 3.6.	Distribution of 'defacto'-household type. Response-, total sample and register. LFS 2006–2009. People 15–74 years old. Per cent	15
Table 3.7.	Classification of family type and household type. Register data 2009. People 15–74 years old. Percent.....	16
Table 3.8.	Household size by time and data source. People 15–74 years old. Percent. 2006–2009	17
Table 4.1.	Household type 2009 by household type 2008. People 15–74 years old. Per cent	18
Table 4.2.	Changes in household type by changes in household identification. People 15–74 years old. Per cent. Register 2008–2009.	19
Table 4.3.	Interview-based household size, by sampled family size. LFS subsample. Average 2006-2009. Percent.....	20
Table 4.4.	Interview-based household size, by register-based "defacto" household size. LFS subsample. Average 2006-2009. Percent.....	21
Table 4.5.	Workless households, by household type. Sample data with survey-employment and register-employment. Per cent. 2006–2009	22
Table 4.6.	Workless household rate, by household type. People 18–59 years old. Per cent. 2006–2009	24