

# The Norwegian Labour Force Survey sampling design

A revision of the sampling plan for the 2021 NLFS

TALL

SOM FORTELLER

NOTATER / DOCUMENTS

2022/34

Susie Jentoft

In the series Documents, documentation, method descriptions, model descriptions and standards are published.

© Statistics Norway

Published: 28 November 2022

ISBN 978-82-587-1619-5 (electronic)

ISSN 2535-7271 (electronic)

<b>Symbols in tables</b>	<b>Symbol</b>
<b>Category not applicable</b> Figures do not exist at this time, because the category was not in use when the figures were collected.	.
<b>Not available</b> Figures have not been entered into our databases or are too unreliable to be published.	..
<b>Confidential</b> Figures are not published to avoid identifying persons or companies.	:
<b>Decimal punctuation mark</b>	.

## **Preface**

This document describes the 2021 sampling plan for the Norwegian Labour Force Survey. Work on the new sampling design started in 2017 as a collaboration between the Division for methods and the Division for labour market and wage statistics. The background for much of the work in this document is a result of a project on changing the sampling unit, completed with the support of Eurostat under Grant agreement 07131.2017.003-2017.597.

Many thanks go to all those in both the Division for methods and the Division for labour market and wage statistics whom contributed to the development and implementation of this new sampling plan.

Statistisk sentralbyrå, 30 October 2022

Arvid Olav Lysø

## **Abstract**

In 2021, many changes were implemented in the Norwegian Labour Force Survey (NLFS). The sampling plan was changed to better reflect the main output goals for the survey. It included moving to an individual sampling unit, allowing new stratification variables and allocations. An optimal sample allocation technique was used, together with new register employment status stratification, resulting in over-sampling of unemployed persons. This leads to better precision in estimates of the unemployment rate, a key output indicator for the survey. A new age-group (75-89 years) was included in the population definition and is sampled from in a separate stratification process. The overall sample size was adjusted down to allocate some resources to a sub-sample wave of all household members. This allows required data deliveries of household statistics not able to be met earlier. A supplementation plan was established to ensure all age-groups are represented in the survey in each wave.

# Contents

<b>Preface .....</b>	<b>3</b>
<b>Abstract.....</b>	<b>4</b>
<b>1. Introduction.....</b>	<b>6</b>
1.1. NFLS survey design history .....	6
1.2. Previous sampling technical process .....	7
1.3. Goals for the 2021 NLFS sampling design.....	8
<b>2. Description of NLFS 2021 changes .....</b>	<b>9</b>
2.1. Population definition changes .....	9
2.2. Sampling unit .....	9
2.3. Regional changes .....	9
2.4. Proxy interviews.....	10
<b>3. Sampling design for 2021 NLFS .....</b>	<b>11</b>
3.1. Sampling frame.....	11
3.2. Sample size .....	11
3.3. Stratification .....	11
3.4. Sample allocation .....	12
3.5. Supplement of 15-year olds .....	16
3.6. Sample of older persons (74-89 years).....	17
3.7. Calculation of selection probabilities.....	17
3.8. Sub-sampling.....	18
<b>4. Conclusion.....</b>	<b>19</b>
<b>References .....</b>	<b>20</b>
<b>Appendix A: Sampling ratio for NLFS 2021 .....</b>	<b>22</b>

# 1. Introduction

In 2021, the Norwegian Labour Force Survey (NLFS) underwent several key changes due to new European regulations (Eurostat, 2022). The questionnaire was revised and updated. The target age group for the survey expanded to include those between 75–89 years. A broader household definition than previously used was required and the sampling unit changed from family to person. These changes along with the goal of improving precision in the main indicators, contributed to the need for a revision of the NLFS sampling design.

In 2017, a project funded by Eurostat, investigated the feasibility of moving the sampling unit for the NLFS from a family to a person unit (Hamre & Jentoft, 2019). As part of this project, alternative sampling allocations and size were investigated. This provided the foundation for the work presented in this document. The work was built on and implemented in the first quarter of 2021.

## 1.1. NLFS survey design history

The Norwegian Labour Force Survey (LFS) has been running as a quarterly survey since 1972. The original sampling plan started as a stratified, two-stage cluster design, selecting primary sample units first, followed by a selection of family units. There were 102 geographical strata made of which 24 consisted of only one municipality with selection probability equal to 1 (Thomsen & Rideng, 1974). This reflected a time when the data collection was done using face-to-face interviews (PAPI) by a team of interviewers spread around the country. It was important to cluster the sampling units to reduce travel costs for the interviewer. The sampling changed to a stratified, single-stage cluster design in 1996 (Statistics Norway, 2003). At this time, interviews were moved to a computer-assisted telephone interview (CATI) collection method. This is the current data collection and gradually interviewers have moved to a centralised interview team. A family unit was the selection unit until 2020.

The LFS has a rotating panel design where the same selected people are requested to respond over several quarters. This improves the survey estimates for changes between quarters. The rotation system was previously a 2-2-2 rotation where participants respond for 2 quarters followed by a two-quarters break and then another 2 quarters participating. This was changed in 1996 to request participants respond in 8 consecutive quarters to further improve the accuracy of quarterly changes. Each quarter,  $1/8^{\text{th}}$  of the sample was rotated out and replaced with a new group, equating to approximately 3000 people. This resulted in an overlap of  $7/8^{\text{th}}$  between quarters and provided for good quarterly change statistics.

LFS became a continuous survey from 1996, covering all weeks in the year instead of only one week per month or quarter. Each quarter consists of 13 weeks, and the sample is divided evenly over the period with fixed reference weeks. The sampling frame for the NLFS between 2006 and 2021 was based on all Norwegian residents aged 15-74 years in the Norwegian Central Population Register. Prior to 2006, 15-year olds were excluded from the sampling frame. Statistics Norway receives an updated copy of the Norwegian Central Population register every evening and the quality of the register is very high.

The sample size started as 6 000 family units, or approximately 12 000 people. This has increased over the years up until 1990 where it was around 24 000 people per quarter and where it currently stands. Prior to 2021, the sample was supplemented with 15-year olds within sampled families, if a family member turned 15 years within the two-year survey period.

Between 1996-2020 the NLFS used a stratified, one-stage cluster design. Individuals were clustered into nuclear family units; people living at the same registered address whom were married or registered partners (same-sex couples), along with their child/children or parent(s). All others were

considered single-person family units (Villund, 2010). The sample was stratified using geographical areas to ensure adequate regional estimates. Detailed regional estimates for employment were an important output from the NLFS previously, however, are less important today due to good register statistics. Table 1.1 gives an overview of the sampling probabilities proportional to the population distribution within each region/stratum for 2019 (Bø & Håland, 2015). Due to changes in the regions in 2020, there were some small changes implemented not shown in this table. In 2018, Nord-Trøndelag and Sør-Trøndelag were merged into 1 region. Prior to this, Nord-Trøndelag was oversampled with a 1.151 rate.

**Table 1.1. Proportional selection probabilities for stratification by region (per 01.01.2019)**

Region	Selection probability
Østfold	1
Akershus	0.871
Oslo	0.871
Hedmark	1
Oppland	1
Buskerud	1
Vestfold	1
Telemark	1
Aust-Agder	1.484
Vest Agder	1
Rogaland	1
Hordaland	0.971
Sogn & Fjordane	1.398
Møre & Romsdal	1
Nordland	1
Troms	1
Finnmark	1.851
Trøndelag	1

Source: Statistics Norway and Bø & Håland, 2015

Until 2021, the sampling design, sample size, and rotation system was virtually unchanged since 1996. There were, however, several documented studies identifying problems in its methods and providing specifications for improvements (Thomsen & Zhang, 2001; Solheim & Zhang, 2001; Vedø & Zhang, 2004; Zhang L.-C. , 2008; Vedø & Rafat, 2003). These were addressed and implemented as part of the 2021 changes.

## 1.2. Previous sampling technical process

Sampling was previously done within Statistics Norway application TREKKBAS. This system is used for sampling most of Statistics Norway's person/family surveys. The advantages of using this system are that all surveys are (theoretically) co-ordinated and it was easy to use. However, there are several technical limitations in this system:

- It does not handle other stratification variables than region.
- It does not handle other sampling units other than nuclear family or person.
- It does not save the actual sample selection probability for each stratum.
- It does not handle real random sampling (only systematic sampling with assumed random pre-sorting), which causes unnecessary varying variance according to Vedø & Zhang (2004) and Zhang (2008).
- It uses a reference person within each family unit as the oldest male. This leads to undesired under-sampling of older females.

These limitations stagnated progress on developing a new plan until 2021. From this time, a new sampling system has been established in the program R and is able to solve these earlier problems.

### **1.3. Goals for the 2021 NLFS sampling design**

The main indicator of interest produced from the NLFS is now the unemployment rate. Norway produces good quality statistics on employment based on register data. Unemployment is still difficult to estimate with good precision based on registers due to coverage errors and definition differences. The main goal was therefore to create a sampling design that focused on improving the precision of the unemployment rate.

Additional to national interests for the NLFS, precision requirements are specified by EU regulations (Regulation (EU) 2019/1700 of the European Parliament and of the Council, 2019). These relate to the employment and unemployment rate at a national and regional (NUTS 2) level. The new sampling design also needed to consider these requirements.

This document describes the new sampling design and gives explanations for some of the choices made. The document is organised into three parts. Chapter 2 describes some of the main changes to the 2021 NLFS in general terms. Chapter 3 describes the new sampling plan in detail. Finally, chapter 4 provides some concluding remarks.

## 2. Description of NLFS 2021 changes

The following sections describe some of the changes to the NLFS in 2021 that affect the sampling design. The NLFS questionnaire changed significantly but did not directly affect the sampling design and is therefore not discussed further. A new estimation method was implemented in 2018 and was not significantly changed in the main 2021 NLFS (Oguz-Alper, 2018). Changes were made to the estimation of household weights from 2021 but will be covered in a future document.

### 2.1. Population definition changes

From 2021, people from 15-89 years are included in the NLFS population definition. This is an expansion of the previous population definition of 15-75 years and now includes those between 75 and 89 years. The oldest age-group are generally a more homogenous group with regards to labour force status than the rest of the population, with the majority being retired. This group is easy to identify in the sampling frame as age is available in the Norwegian Central Population Register. The older population are not a specific target group for the main indicators in the national NLFS and it was therefore desirable to establish a sampling plan to under-sample this group.

Additionally, the population definition states that only those living in private-households should be included. This had previously been achieved through a cleaning process after units were selected. From 2021, a new household register allowed for the sampling frame to better target this definition through excluding many of those living in institutions. This is described further in chapter 3.1.

### 2.2. Sampling unit

In 2019, Statistics Norway completed a report on NLFS sampling unit (Hamre & Jentoft, 2019). This was part of a Eurostat grant project and detailed the advantages and disadvantages of moving away from sampling a family unit in favour of selecting persons instead. This was implemented in 2021.

Several earlier studies have investigated the effects of a family sampling unit (Villund, 2010; Vedø & Rafat, 2003; Hagesæther & Zhang, 2007; Hagesæther & Zhang, 2009). One consideration is the potentially higher variance and bias of the indicators, due to the clustering effect of participants in the same family (Vedø & Rafat, 2003). However, this may not be a significant problem given good calibration methods targeting the individuals (Hagesæther and Zhang, 2009).

The use of good stratification variables is important in an effective sampling design. Moving to a person sampling unit allows for more specific stratification, for example targeting those with higher probabilities of unemployment. Oversampling of these groups can lead to considerable precision improvements and was a major consideration in the decision to change the sampling unit (Hamre & Jentoft, 2019).

### 2.3. Regional changes

Earlier sampling plans have included region in the stratification design. This is mainly because there have been both national and international interest in labour market statistics with a good precision at a regional level. Changes to the administrative regions therefore plays an impact on the sampling design. In 2020, the regions of Norway (NUTS 2) changed and were reduced from 7 to 6 regions (Statistics Norway, 2022). NUTS (Nomenclature of territorial units for statistics) are the standard regions used throughout Europe which are organised in a hierarchical system. There are Eurostat requirements for precision levels of the unemployment rate within these NUTS 2 regions. However, in general, there are not strong differences for employment/unemployment rates between the regions.

## 2.4. Proxy interviews

Prior to 2021, proxy interviews were used in the NLFS. This is when a close family member (parent or spouse) responds on behalf of another individual. They were used to increase the response rate for hard-to-reach individuals and to save contact and interview time. Previous studies have shown that proxy interviews likely improve the quality of indicators when good auxiliary variables are lacking for adjustment (Zhang, Thomsen, & Kleven, 2013). However, due to the detailed nature of some questions in the NLFS, it may also lead to bias in certain situations (Jentoft & Mevik, 2014; Li, 2018).

In 2021, proxy interviews were discontinued because of the change to a person sampling unit. All interviews for the main NLFS are now performed directly. This has potentially contributed to an increase, of around 5 percent points, in the non-response rate, however, there are other contributing factors, discussed further in section 0. Prior to 2021, around 15 percent of interviews were performed by proxy interviews, including a high number of students (Hamre & Jentoft, 2019).

### 3. Sampling design for 2021 NLFS

From 2021, a sampling frame of individuals living in private households is used to draw a random sample of persons. The overall sample size remains unchanged, however, new stratification and allocation have been implemented to improve the precision of statistics on unemployment. The survey continues to be run as a quarterly survey with an 8 waves rotation. Each person selected to the survey is invited to participate in the survey every quarter for two years. The sample is spread evenly over the quarter (13 weeks) giving a continuous data collection process. A set of core questions are asked in every wave of every quarter, while annual, biannual and ad. hoc. modules are asked in waves 2 and 6 only. An additional sub-sample of all adult household members from wave 2 are also interviewed. While there is ongoing research on moving to a CAWI, the NLFS is currently still run as a CATI survey. This chapter provides further details on the new sampling plan, implemented in 2021.

#### 3.1. Sampling frame

The Norwegian Central Population Register continues to be the main sampling frame for the NLFS. This covers all persons that have lived or intend to live in Norway for 6 months or more. Those living in both private and institutions are included in this register.

From 2021, a new household register was established which combines many administrative data sources. This includes checking for shared loans or property, shared moving dates, and other family relations (for example siblings). It also only includes households living in private homes and therefore aligns better with the Eurostat definition of the target population for NLFS. Therefore, this register is used to filter the Central population register to create the sampling frame.

#### 3.2. Sample size

The overall sample size for the NLFS remained approximately unchanged in 2021 with around 24 000 invited to participate each quarter.

Due to sub-sampling of additional household members in one wave, the sample size for the main NLFS was reduced to allow resources for the additional interviews (see section 3.8 for more details on sub-sampling). Earlier investigation, gave an average household size of around 2.0 persons, meaning for every person in the main survey, one extra household member would need to be interviewed. A sample size of 21000 was allocated to the main NLFS leaving 3000 reserved for additional household interviews (in 1/8 of the interviews throughout a year). A small number of 15-years olds were sampled in addition to this and are described in section 3.5.

#### 3.3. Stratification

Stratification of the population before sampling allows us to create more homogenous groups to select from. By over- and under-sampling persons in certain strata we can create statistics with better precision. For the NLFS, the main indicator is the unemployment rate, so stratification that correlates with this will allow for more control of the precision level in the statistics. Three variables were selected for stratification of the 2021 NLFS: *region*, *age-groups* and *register employment status*.

Region was selected to allow control and improvement for the precision of the smaller regions. Norway is currently divided into 6 NUTS 2 areas which are used in the sampling, estimation and publishing of statistics: Oslo og Viken, Innlandet, Agder og Sør-Østlandet, Vestlandet, Trøndelag, and Nord-Norge (Statistics Norway, 2022).

Connections to the labour market change considerably with age. Therefore, stratification by age can allow control over the precision of labour market statistics. Five age-groups were chosen: 15-24 years, 25-54 years, 55-66 years, 67-74 years and 75-89 years. The oldest age-group is discussed further in section 3.6.

In 2015, a new monthly register of employment was established (A-ordningen) as a collaboration between Statistics Norway, The Norwegian Tax Administration and the Norwegian Labour and Welfare Administration. Employers are required to report details on hours worked and pay for all employees in Norway. Additionally, the register contains information on those registered as unemployed, those self-employed, students and individuals conscripted into compulsory military service (Statistics Norway, 2020). This register provided the final stratification variable, with three groups formed: employees (wage-earners), unemployed, other (student, unemployed, self-employed).

#### Stratification variables

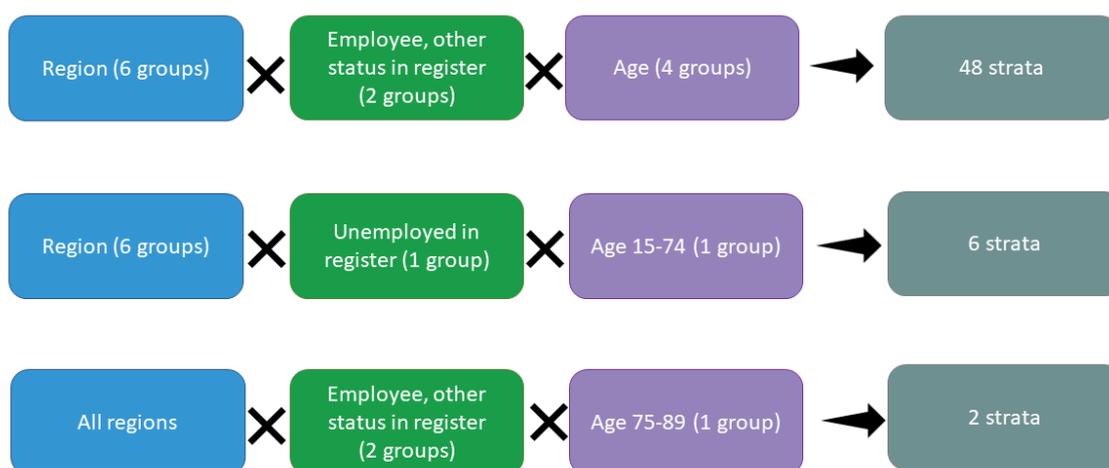
Region: Oslo og Viken, Innlandet, Agder og Sør-Østlandet, Vestlandet, Trøndelag, and Nord-Norge

Age: 15-24 years, 25-54 years, 55-66 years, 67-74 years and 75-89 years

Register employment status: employee, unemployed, other

The final stratification groups are combinations of crossing these 3 variables. Those who are unemployed in the register are a small group and are therefore not divided into age groups. This gives 6 strata of register-unemployed persons, one for each NUTS 2 region. The oldest population group (75-89 years) are only divided into employees and other for labour status with no regional stratification giving 2 strata for this group. Figure 3.1 gives an overview of the crossing of the variables to create the 56 strata. A list of all strata with their sampling ratio is given in Appendix A.

**Figure 3.1** Diagram showing the various crossing of variables to produce the 56 strata



Source: Statistics Norway

### 3.4. Sample allocation

The sample was first allocated using Bethel's multivariate optimal sampling allocation (Bethel, 1989) to fulfil the Eurostat precision requirements (Eurostat, 2018). Secondly, any remaining sample was allocated to optimise the unemployment precision at a national level. Multivariate optimal allocation is a development of Neyman's univariate allocation that oversamples strata with higher variances. The optimisation was implemented using the R-package *MAUSS-R*. More details on the theory and implementation can be found in Barcaroli et. al. (2015).

### Precision requirements

Precision requirements were based on the provisional agreement of the proposed common framework for European statistics relating for persons and households, based on data at individual level collected from samples (Eurostat, 2018). There are two requirements at a national level, one for the estimated quarterly unemployment-to-population rate and one for the estimated quarterly employment-to-population rate. There is one precision requirement at a domain level, NUTS 2 regions, for the estimated quarterly unemployment-to-population rate. The standard error of a particular estimate,  $\widehat{SE}(\hat{p})$ , shall not be larger than

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{f(N)}}$$

where for the two national requirements

$$f(N) = 7800\sqrt{N} - 4500$$

and  $N$  is the population aged 15-74 years in millions. For the domain level requirement

$$f(N_r) = \begin{cases} 1300, & \text{if } N_r \geq 0.3 \text{ million} \\ \frac{1300}{0.3}, & \text{if } N_r < 0.3 \text{ million} \end{cases}$$

where  $N_r$  is the 15-74 years population in NUTS 2 region,  $r$ , in millions.

The requirements were calculated using a  $\hat{p} = 0.6$  for employment-to-population rate and  $\hat{p} = 0.05$  for unemployment-to-population rate. All NUTS 2 regions have more than the 300 thousand persons boundary besides Innlandet. The 15-74 years population is around 290 thousand persons, so the precision error can be relaxed there. However, given it is very close to the boundary we have used a value of  $f(N_r) = 1300$  for this region as well to allow for an increase in population in the coming years.

Optimal allocation requires defining the coefficient of variation (CV) rather than the standard error directly. This is defined as

$$CV = \frac{\widehat{SE}(\hat{p})}{\hat{p}} \cdot 100$$

The CV requirements are given in Table 3.1. These requirements allow for an employment percentage drop to 60 percent, and an unemployment rate of 5 percent in the population, or up to 7 percent of the labour force (based on a 70 percent labour force in the population).

**Table 3.1 CV requirement by IESS-FR for NLFS**

	CV requirement	
	National	NUTS 2
Employment	0.007	-
Unemployment (in population)	0.04	0.12

Source: Statistics Norway & Council of the European Union, 2019

### Comparison of allocation schemes

Optimizing the allocation is done to improve precision in the estimates. Figures 3.1 and 3.2 show the allocation of the new sample (called optimal allocation) compared to the prior allocation, and proportional allocation. Proportional allocation is a popular method of distributing a sample where

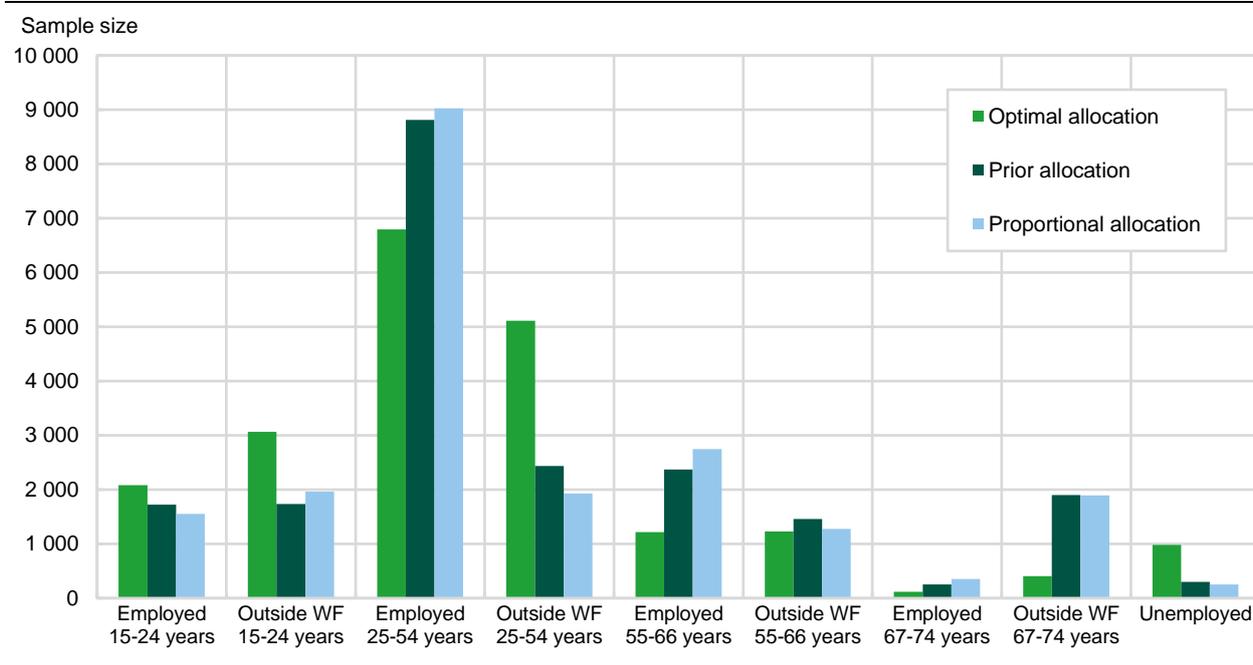
the sample size in each stratum is proportional to the population size. Each strata sample size is calculated as

$$n_h = \frac{N_h}{N} \cdot n$$

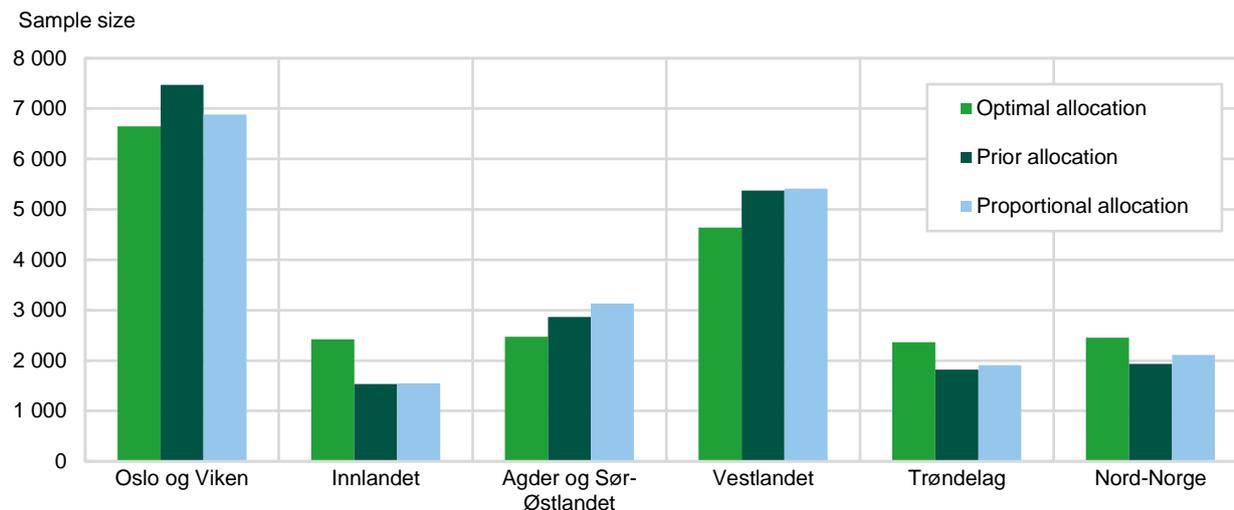
where  $n$  is the total sample size for the survey and  $n_h$  and  $N_h$  are the sample and population sizes in stratum  $h$ .

We see that the new sample allocation selects more register unemployed and persons outside the labour force compared to the other sampling allocations. This is intuitive as these groups are likely to have a higher percent of unemployed persons in them and increasing the numbers will improve the accuracy. We also see an increase in the number of selected persons from Innlandet which is due to the NUTS 2 employment requirement.

**Figure 3.2. Comparison of allocation schemes for age-groups and register employment status. Employed refers to the strata of those that are employed (receiving wages) in the A-ordningen register. Outside WF refers to strata of those that are not receiving wages and not unemployed in the A-ordningen register**



Source: Statistics Norway

**Figure 3.3. Comparison of allocation schemes for NUTS 2 regions**

Source: Statistics Norway

### Variance calculations for key indicators

The coefficient of variation was estimated for five key indicators using a simplified variance estimation. Variance was calculated using the sampling stratification strata ( $h$ ) where the weights of the stratum,  $w_h$ , are

$$w_h = \frac{N_h}{N}$$

The estimate for the indicator/proportion ( $\hat{p}$ ) was then

$$\hat{p} = \sum_{h=1}^H w_h \cdot \hat{p}_h$$

where  $\hat{p}_h$  is the estimated proportion of that key indicator in stratum  $h$ , and the variance for the estimate of the indicator proportion ( $\hat{p}$ ) is

$$\widehat{Var}(\hat{p}) = \sum_{h=1}^H w_h^2 \cdot \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1}$$

where  $n_h$  is the number of persons in the sample in stratum,  $h$ . The coefficient of variation ( $CV$ ) is then calculated as a percent

$$CV = \frac{\sqrt{\widehat{Var}(\hat{p})}}{\hat{p}} * 100$$

Previous stratification variables used register labour-force status for the time period valid at the time of interview. In reality, the stratification variables will be up-to-date at the time of sampling but suffer from time-lag over the two-year data collection period, particularly labour market status. The effect of stratification on lowering variance through optimal allocation will therefore likely be less than that shown in the previous study.

To estimate the variance and CV, a time-lag was given to the strata. The NLFS is a two-year long panel survey and the effect of the stratification used in selecting the sample will degrade over time. The correlation between the stratification variables and the key indicators weakens as the time

between selecting and interview increases. Strata were created using register information from one-year prior to interview were used to consider this. This should be approximately an average of the time-delay effect.

The key indicators calculated were proportions for employment (in population), unemployment, outside the workforce, self-employed and under-employed. Table 3.2 gives the estimates for CV using the new, optimal allocation compared to the previous design and proportional allocation. It shows an improvement in the unemployment precision which has been the main consideration for the new sampling design. Precision levels are similar for employment and outside the work force rates compared to the previous allocation, while slightly worse for self-employed and under-employment rates.

**Table 3.2 Estimate of Coefficient of variation for five key NLFS indicators for three sample allocation. The new allocation is that using optimisation, allocation prior to 2021 and a proportional allocation**

	New allocation	Allocation prior to 2021	Proportional allocation
Employment rate	0.39	0.39	0.38
Unemployment rate	4.10	4.85	4.63
Outside work-force rate	0.85	0.86	0.83
Self-employment rate	3.97	3.86	3.60
Under-employment rate	6.03	5.80	5.68

Source: Statistics Norway

### Non-response

While the NLFS is a compulsory survey, there is still some non-response. This is mainly due to difficulties in getting in contact with individuals. Prior to 2021, the non-response rate for the NLFS was around 15-18 percent. The non-response rate is not consistent throughout the population groups. There is observed higher non-response in younger age groups and among those registered unemployed for example. When we increase the sample size of these strata compared to that prior to 2021, the non-response will naturally increase.

Accordingly, we increased the gross sample size in these groups slightly to allow for the additional non-response expected. This was done using an iterative process in the optimal allocation algorithm. The expected response rates were calculated based on data from the 10-quarters prior to allocation, i.e. 1<sup>st</sup> quarter 2018 to 2<sup>nd</sup> quarter 2020. The estimated response rate under the new allocation was 80 percent.

Additionally, we were aware that dropping the use of proxy interviews would likely contribute to an increase in non-response rates. Pilot testing using a person sampling unit and no proxy interviews indicated that this was not a significant problem and was therefore not adjusted for.

### 3.5. Supplement of 15-year olds

Due to the nature of the sampling, without supplementation there would be no 15-year olds and very few 16-year olds in the later waves of the NLFS. This is because the survey has a panel design with eight consecutive waves over two-years. The youngest group sampled are 15 years at the time of sampling. As some additional modules of the NLFS target a subset of waves, ignoring this could lead to a problematic bias. This is particularly the case in estimating indicators targeted towards young people such as the NEET rate (Not in Employment, Education or Training).

Additional 15-year olds are sampled and added to later waves to ensure a balance of ages in each wave. Based on 3<sup>rd</sup> quarter 2020 data, there are an estimated 46 15-year olds in each wave under the new sampling design. We therefore supplement with an additional twelve 15-year olds in sample waves 2-8 to represent the 1/4 of approximately 46 that turn 16 years in the quarter. This equates to 84 additional 15-year olds at each sampling point spread over waves 2-7.

### 3.6. Sample of older persons (74-89 years)

The extension of the NLFS population definition to include those between 75-89 years creates some additional challenges in the sampling design. Table 3.3 gives an overview of the number and proportion of the older population (74-89 years) which makes up around 9 percent of the NLFS population. The employment rate for this age-group is around 4.2 percent (in 2019) and using the current NLFS definition, this age-group cannot be defined as unemployed.

**Table 3.3. Overview of population 74-89 years and employment/self employed status. January, 2019**

	Population, 75-89 years
Population size	345 795
Population as proportion of LFS population	0.09
Number register-employed	14 573
Proportion register-employed	0.04
Number register self-employed	8 388
Proportion register self-employed of those employed	0.58

Source: Statistics Norway

If this group was sampled using proportional to the population allocation, the main sample in the core age-groups (15-74 years) would need to be reduced. This would affect the precision of the estimates for the main core age-groups which is not desirable. There are no plans to publish statistics specifically on this group and respondents will only be included in more general figures for employment/unemployment. Therefore, the sample size for this group was set to a minimum in the new NLFS sampling plan.

A sample size of 120 for this age group is used in the new sampling plan. This equates to 15 in each of the 8 waves. The sample is stratified by register employment with an allocation of 60 registered employed and 60 outside the labour force. A consideration taking when deciding the sample size of this group relates to the calculation of the estimation weights. This age-group contains three 5-year age spans (75-79, 80-84, 85-89 years). These three 5-year age spans are required by Eurostat for weight calibration when possible. The new sampling plan therefore contains a minimum number of these groups to produce reasonable estimation weights. If there is too few in the sample, the estimation weights will be extremely high which is both undesirable and problematic when delivering data to Eurostat. A sample size of 5 in each age-group, in each wave, would theoretically allow estimation weights within acceptable limits.

### 3.7. Calculation of selection probabilities

Selection probabilities give the probability that an individual is selected to the NLFS and are calculated at the time of sampling. They are composed of two parts, one for the selection to the main NLFS sample and an addition probability for 15-year olds to be selected as a supplement in one of the waves.

For the population aged 16 to 89 years, the selection probability for individual  $i$ , in stratum  $h$  to be selected to the sample,  $S$ , for a given wave is given by the first part only as

$$P(i \in S) = \pi_i = \frac{n_h}{N_h}, i \in U_h$$

where  $n_h$  is the sample size for the stratum for the wave,  $N_h$  is the population size in the stratum and  $U_h$  is the population sampling frame for stratum,  $h$ . For the population aged 15 years, the selection probability is that described above, plus an additional probability associated with the supplementation. The selection probability is therefore

$$\pi_i^* = \pi_i + \pi_i^+, i \in U_{h,15}$$

where

$$\pi_i^+ = \frac{n_{15}^+}{N_{15} - \hat{n}_{15}} \frac{1}{7} (1 - \pi_i), i \in U_{h,15}$$

The number of additional 15-year olds selected at a given time,  $n_{15}^+$ , is 84. This is multiplied by the selection probability of being selected into a given wave (1/7). The number of 15-year olds selected in the first stage,  $\hat{n}_{15}$ , is estimated as

$$\hat{n}_{15} = \sum_h \frac{n_h N_{h15}}{N_h}$$

where  $N_{h15}$  is the population size for 15-year olds in stratum  $h$ .

### 3.8. Sub-sampling

From 2021, the NLFS includes interviews of all adult household members for a sub-sample of one wave of individuals. Questions are asked in wave two and include those targeted to establish labour market status for all members of the household. This is then used to create household-level statistics, for example the number of jobless households.

It has long been a requirement from Eurostat to report household statistics for labour force indicators. This is something Norway has not been able to fulfil due to the use of a narrower family concept. The family unit previously used in the NLFS did not include unmarried couples, which are common in Norway. To address this, a broader definition was used to encompass a better household concept more aligned with that required by Eurostat (2018). The household for the NLFS is defined by the new household register at the time of sampling and is not adjusted at the time of interviewing. This is mainly due to the costs associated with additional interviews and processing. The calculation of household selection probabilities and weights will be described in a separate document.

## 4. Conclusion

By moving to a person sampling unit in 2021, new opportunities for stratification and allocation were presented. New variables (age and register employment status) are now used in the stratification with an allocation that favours good precision of unemployment statistics. Including age has allowed for under-sampling of the new older age-group in the survey (75-89 years), thereby minimizing the negative impact of reducing the sample size in the remaining age-groups.

The new sampling plan described in this report was first implemented in the first quarter 2021. It is being implemented one-wave at a time and the entire sample won't be under the new design until the end of 2022. This gradual implementation reduces the impact of the change and is practically easier. However, it has resulted in a larger overall sample size during this phase in period, creating some pressure on the interview team. The total sample size should be relatively stable from the start of 2023. The household size is not currently controlled for in the sampling plan which means that the sub-sample of additional household members will vary from quarter to quarter. If this is shown to be problematic, an additional household size stratification can be considered to stabilise this.

Finally, the optimal allocation was run using data from a 10-quarter period prior to 2021. Changes in the labour market affect the variances in the strata and the allocation should be reviewed on a regular basis. Non-response rates will also change over time, particularly given that proxy interviewing is no longer performed. This also affects the sampling design as the gross sample size is predicted based on an estimate of the non-response for the strata. Again, a revision of the allocation should also include revised non-response rates for the strata.

## References

- Barcaroli, B., Bulielli, T., & De Vitiis, C. (2015). *MAUSS-R Multivariate allocation og units in sampling surveys*. Italian National Institute of Statistics.
- Bethel, J. (1989). Sample Allocation in Multivariate Surveys. *Survey Methodology*, pp. 47-57.
- Bø, T. P., & Håland, I. (2015). *Dokumentasjon av Arbeidskraftundersøkelsen (AKU) etter omlegging i 2006*. Oslo: Statistics Norway.
- Eurostat. (2018). *Main changes to be implemented in the future Labour Force Survey under the Integrated European Social Statistics Framework Regulation. Document for item 2.1 of the agenda. F3/LAMAS/02/18*. Luxembourg: European Commission.
- Eurostat. (2022, 10 12). *EU labour force survey – main features and legal basis*. Retrieved from [https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU\\_labour\\_force\\_survey\\_%E2%80%93\\_main\\_features\\_and\\_legal\\_basis#Legal\\_basis](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=EU_labour_force_survey_%E2%80%93_main_features_and_legal_basis#Legal_basis)
- Hagesæther, N., & Zhang, L.-C. (2007). *Om estimeringsusikkerhet og utvalgsplan i AKU*. Oslo-Kongsvinger: Statistics Norway.
- Hagesæther, N., & Zhang, L.-C. (2009, 25 3). A Note on the Effect of Auxiliary Information on the Variance of Cluster Sampling,. *Journal of Official Statistics*.
- Hamre, J. I., & Jentoft, S. (2019). *Sampling unit in the Norwegian LFS* Methodology report for sub-action 9. Unpubl.: Statistics Norway.
- Jentoft, S., & Mevik, A.-K. (2014). *Arbeidskraftundersøkelsen. Imputering av overtid ved indirekte intervju*. Oslo-Kongsvinger: Statistics Norway.
- Li, L. (2018). *Proxy interviews in the Norwegian labour Force Survey*. Oslo-Kongsvinger: Statistics Norway.
- Oguz-Alper, M. (2018). *New estimation methodology for the Norwegian labour force survey*. Oslo-Kongsvinger: Statistics Norway.
- Regulation (EU) 2019/1700 of the European Parliament and of the Council. (2019). *Official Journal of the European Union*.
- Solheim, L., & Zhang, L.-C. (2001). *AKU: Utvalgsplanen og husholdninger (unpub.)*. Oslo: Statistics Norway.
- Statistics Norway. (2003). *Labour Force Survey 2001*. Oslo: Statistics Norway.
- Statistics Norway. (2020). *A-ordningen*. Retrieved from [ssb.no: https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/a-ordningen](https://www.ssb.no/data-til-forskning/utlan-av-data-til-forskere/variabellister/a-ordningen)
- Statistics Norway. (2022). *KLASS*. Retrieved from [ssb.no: https://www.ssb.no/klasse/klassifikasjoner/106/endinger](https://www.ssb.no/klasse/klassifikasjoner/106/endinger)
- Thomsen, I., & Rideng, A. (1974). *Oversikt over arbeidet med ny utvalgsplan*. Oslo: Statistics Norway.

- Thomsen, I., & Zhang, L.-C. (2001, 17 2). The Effects of Using Administrative Registers in Economic Short Term Statistics - The Norwegian LFS as a Case Study. *Journal of Official Statistics*.
- Vedø, A., & Rafat, D. (2003). *Sammenligning av utvalgsplaner i AKU*. Oslo: Statistics Norway.
- Vedø, A., & Zhang, L.-C. (2004). *Omlegging av utvalgsplanen for AKU*. Oslo: Statistics Norway.
- Villund, O. (2010). *Household data in the Norwegian Labour Force Survey*. Oslo: Statistics Norway.
- Zhang, L.-C. (2008, 24 4). On Some Common Practices of Systematic Sampling. *Journal of Official Statistics*.
- Zhang, L.-C., Thomsen, I., & Kleven, Ø. (2013, 81 2). On the use of auxiliary and paradata for dealing with non-sampling errors in household surveys. *International Statistical Review*.

## Appendix A: Sampling ratio for NLFS 2021

**Tabell A1** Table of the sampling ratios for the 56 strata in the main NLFS 2021

Region	Age	Labour market status in register	Sampling proportion
Oslo og Viken	15-24	Employee	1.06
Oslo og Viken	15-24	Other	1.57
Oslo og Viken	25-54	Employee	0.67
Oslo og Viken	25-54	Other	1.86
Oslo og Viken	55-66	Employee	0.44
Oslo og Viken	55-66	Other	0.75
Oslo og Viken	67-74	Employee	0.31
Oslo og Viken	67-74	Other	0.18
Oslo og Viken	15-74	Unemployed	2.98
Innlandet	15-24	Employee	1.95
Innlandet	15-24	Other	2.90
Innlandet	25-54	Employee	1.29
Innlandet	25-54	Other	3.41
Innlandet	55-66	Employee	0.83
Innlandet	55-66	Other	1.34
Innlandet	67-74	Employee	0.74
Innlandet	67-74	Other	0.34
Innlandet	15-74	Unemployed	5.38
Agder og Sør-Østlandet	15-24	Employee	1.01
Agder og Sør-Østlandet	15-24	Other	1.53
Agder og Sør-Østlandet	25-54	Employee	0.67
Agder og Sør-Østlandet	25-54	Other	1.78
Agder og Sør-Østlandet	55-66	Employee	0.44
Agder og Sør-Østlandet	55-66	Other	0.68
Agder og Sør-Østlandet	67-74	Employee	0.51
Agder og Sør-Østlandet	67-74	Other	0.18
Agder og Sør-Østlandet	15-74	Unemployed	2.82
Vestlandet	15-24	Employee	1.01
Vestlandet	15-24	Other	1.51
Vestlandet	25-54	Employee	0.66
Vestlandet	25-54	Other	1.89
Vestlandet	55-66	Employee	0.44
Vestlandet	55-66	Other	0.70
Vestlandet	67-74	Employee	0.31
Vestlandet	67-74	Other	0.18
Vestlandet	15-74	Unemployed	2.88
Trøndelag	15-24	Employee	1.59
Trøndelag	15-24	Other	2.29
Trøndelag	25-54	Employee	1.04
Trøndelag	25-54	Other	2.68
Trøndelag	55-66	Employee	0.67
Trøndelag	55-66	Other	1.11
Trøndelag	67-74	Employee	0.79
Trøndelag	67-74	Other	0.28
Trøndelag	15-74	Unemployed	3.92
Nord-Norge	15-24	Employee	1.54
Nord-Norge	15-24	Other	2.25
Nord-Norge	25-54	Employee	1.02
Nord-Norge	25-54	Other	2.72
Nord-Norge	55-66	Employee	0.68
Nord-Norge	55-66	Other	1.09
Nord-Norge	67-74	Employee	0.65
Nord-Norge	67-74	Other	0.27
Nord-Norge	15-74	Unemployed	4.51
All	75-89	Employee	11.86
All	75-89	Other	0.52

Source: Statistics Norway