Norwegian Citizen Panel

2020, Eighteenth Wave

Methodology report

Øivind Skjervheim

Asle Høgestøl

Olav Bjørnebekk

Amund Eikrem

August, 2020

NORSK MEDBORGERPANEL

IDEAS2EVIDENCE
Bygger kunnskap

TABLE OF CONTENTS

BACKGROUND

This report describes the procedures of data collection in the eighteenth wave of The Norwegian Citizen Panel. Furthermore, the report discusses technical aspects of the data collection before turning to the representativity of the panel and how the weights are calculated.

The Norwegian Citizen Panel (NCP) is one of the main components of Digital Social Science Core Facility (DIGSSCORE) at the University of Bergen. NCP was established as a collaboration between several departments at the Faculty of Social Sciences at the University of Bergen and NORCE.

ideas2evidence is responsible for the panel recruitment, the administration of the panel, and the technical solutions regarding data collection and computing.

TECHNICAL ASPECTS OF THE SURVEY

SOFTWARE

The surveys are administrated through the web-based survey software Confirmit. Confirmit is a "Software-as-a-Service" solution, where all software runs on Confirmit's continuously monitored server park, and where survey respondents and developers interact with the system through various web-based interfaces. This software provides very high data security and operational stability. The security measures are the most stringent in the industry, and Confirmit guarantees 99.7 percent uptime. ideas2evidence does the programming of the survey in Confirmit on behalf of The Norwegian Citizen Panel.

PILOT AND SOFT LAUNCH

The survey went through small-N and large-N pilot testing before data collection. In addition, the survey was tested extensively during the development phase by ideas2evidence and the researchers involved in the project.

The pilot testing was regarded as successful, and no major technical revisions were deemed necessary.

The field period started by inviting a random sample of the respondents (soft launch). This was done in order to minimize the consequences if the questionnaire contained technical errors. No such errors were located/reported, and remaining panel members was therefore invited the following day.

RANDOMIZATION PROCEDURES

Each wave of NCP has an extensive use of randomization procedures. The context of each randomization procedure may vary, [1] but they all share some common ground that will be described in the following.

All randomization procedures are executed live in the questionnaire. This means that the randomization takes place while the respondent is in the questionnaire, as opposed to pre-defined randomizations that are uploaded to the questionnaire. All randomizations are independent from another, unless the documentation states otherwise.

[1] Some examples: sorting respondents in different thematic subsets, randomly allocate treatment value in experiments, randomize order of an answer list/array, order a sequence of questions by random, ask a given question to a subset of the respondents.

The randomization procedures are written in JavaScript. Math.random()[2] is a key function, in combination with Math.floor()[3]. These functions are used to achieve the following:

- Randomly select one value from a vector
- Randomly shuffle the contents of an array

The first procedure is typically used to determine a random sample of respondents to i.e. a control group. Say for example we wish to create two groups of respondents: group 1 and group 2. All respondents are randomly assigned the value 1 or 2, where each randomization is independent from one another. When N is large enough these two groups will be of equal size (50/50).

Here is an example of the JavaScript code executed in Confirmit:

```
var form = f("x1");
if(!form.toBoolean()) // If no previous randomization on x1
{
  var precodes = x1.domainValues();// Copies the length of x1
  var randomNumber : float = Math.random()*precodes.length;
  var randomIndex : int = Math.floor(randomNumber);
  var code = precodes[randomIndex];
  form.set(code);
}
```

The second procedure is typically used when defining the order of an answer list as random. This can be useful for example when asking for the respondent's party preference or in a list experiment. However, since i.e. a party cannot be listed twice, the procedure must take into account that the array of parties is reduced by 1 for each randomization.

Here is an example of the JavaScript code executed in Confirmit [4]:

```
Function shuffle(array) {
  var currentIndex = array.length, temporaryValue, randomIndex;
  // While there remain elements to shuffle...
  while (0 !== currentIndex) {
    // Pick a remaining element...
    randomIndex = Math.floor(Math.random() * currentIndex);
    currentIndex -= 1;

    // And swap it with the current element.
    temporaryValue = array[currentIndex];
    array[currentIndex] = array[randomIndex];
    array[randomIndex] = temporaryValue;
  }
  return array;
}
```

PREVIOUS WAVES OF RECRUITMENT

Existing panel members were recruited in wave 1, wave 3, wave 8, wave 11, wave 14 and wave 16. All samples were drawn from the *National Population Registry* of Norway. This registry holds information on everyone born

[2] Please see following resource (or other internet resources):https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Math/random
[3] Please see following resource (or other internet resources):https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Math/floor
[4] Code collected from Mike Bostocks visualization: https://bost.ocks.org/mike/shuffle/

in Norway, as well as former and current inhabitants. The Norwegian Tax Administration holds the formal responsibility for this registry, but the administration is partly outsourced to the private IT-company Evry. Evry drew the sample on behalf of the Norwegian Citizen Panel after relevant permissions were acquired from the Norwegian Tax Administration.

The samples consisted of people over the age of 18 that were randomly drawn from the registry. The extracted information was a) last name, b) first name, c) address, d) gender, e) year of birth, and f) phone number (the latter was not included in wave 1). The sample excluded persons without a current home address in Norway.

For a detailed description of the recruitment process in wave 1, wave 3, wave 8, wave 11, wave 14 and wave 16, we refer to the respective methodology reports for the respective waves. A short summary of previous recruitment efforts is presented in table 1. Note that there are some differences between the recruitment processes. A detailed description of the recruitment in wave 18 follows in the next section.

The response rate of recruitments 4-7 is substantially lower than later waves of recruitment. The most important explanation is new restrictions enforced by the Norwegian Tax Administration with regards to how many times the Citizen Panel can contact persons in the net sample. Respondents in recruitments 4-7 were contacted twice at most. Recruitment 1 also had a maximum of two contact points, but achieved a response rate of 20 percent. One explanation for why we cannot replicate a response rate of 20 percent in recruitments 4-7 might be that NCP did a lot of promotion of the panel through different media outlets leading up to and during recruitment 1. The promotion of the panel was also done in relation to the Norwegian Parliamentary election that same fall.

Table 1: Summary of recruitment processes

	Sample size	Mode	Contacts	Returned letters	Response Rate (%)
Recruitment 1 (wave 1)	25 000	Postal	2	546	20.1 %
Recruitment 2 (wave 3)	25 000	Postal, phone/SMS	4	543	23.0 %
Recruitment 3 (wave 8)	22 000	Postal/SMS	3	479	19.4 %
Recruitment 4 (wave 11)	14 000	Postal/SMS	2	334	15.1 %
Recruitment 5 (wave 14)	14 000	Postal/SMS	2	389	15.0 %
Recruitment 6 (wave 16)	34 000	Postal/SMS	2	994	14.9 %
Recruitment 7 (wave 18)	15 000	Postal/SMS	2	381	14 %

DATA COLLECTION WAVE 18

RECRUITING A NEW SET OF PANEL MEMBERS

In wave eighteen, the Norwegian Citizen Panel recruited new panel members. This section gives a detailed description of the sample frame, recruitment process and the results.

THE RECRUITMENT PROCESS

As in the preceding waves of recruitment a gross sample was drawn from the population registry. Evry drew the sample on behalf of the Citizen Panel after the necessary permissions were acquired from the Norwegian Tax Administration.

15,000 people over the age of 18 were randomly drawn from the register. The extracted information was as before a) last name, b) first name, c) address, d) gender, e) telephone number(s) (if available) and, f) year of birth. The sample excluded individuals without a current home address in Norway.

RESULTS OF THE RECRUITMENT PROCESS - SURVEY RESPONDENTS AND PANEL MEMBERS

First, letters were sent to everyone in the sample. The letters contained the following information: a) a description of the project, b) the Citizen Panel's policy on privacy and measures taken to protect the anonymity of the participants, c) the time-frame of the project, d) the participants' rights to opt of the panel at any time in the future, e) contact information for the people responsible for the project, f) a unique log-in id and the web address to the panel's web site and g) the estimated time required to complete the survey (15 minutes).

In order to maximize the response rate, an incentive in the form of three gift cards is included in the project. The values of the gift cards are 8,000 NOK. To enter the lottery respondents were required to join the panel and provide their email addresses. Respondents were asked to register on the panel's web site and log into the survey using the unique ID-code provided in their personal letter. Information on the lottery was included in all correspondence with respondents.

The invitational letter was posted 3rd of June 2020.

The second reminder was distributed by SMS or post card. Respondents below 60 years of age registered with a cell phone number received an SMS. Respondents who did not fit this description received a post card reminder. This is different from the first three waves of recruitment. In wave eleven, an experiment was conducted regarding the use of SMS and postcard. That experiment gave the panel more information regarding the effectiveness of different recruitment strategies, and thus gave the opportunity for a more cost-efficient use of reminders in subsequent waves.

Both reminders were sent to respondents who a) had not logged into the survey, or b) had neither completed the survey. Respondents were encouraged to join the panel, with reference to the invitation letter. The unique log-in ID provided in the original letter was included in both the post card and the SMS. The SMS reminder also included a direct link to the survey.

The post card was posted the 17th of June, and the SMS was distributed June 15th.

RESULTS OF THE RECRUITMENT PROCESS - SURVEY RESPONDENTS AND PANEL MEMBERS

It is necessary to make a distinction between panel members and survey respondents. We define panel members as respondents who register their e-mail address, regardless of whether they have completed the questionnaire or not. Survey respondents are respondents who have completed a certain share of the questionnaire, regardless of whether they have entered their e-mail address or not.

Of the 14,941 letters that were sent out, 381 were returned, and 7 respondents opted out. 20.3 percent (2,951) of the remaining 14,553 logged on and accessed the survey. 2,018 individuals completed the questionnaire, while 933 exited the questionnaire before completion, though 3.2 percent of these responses are kept as a part of the survey data. The remaining 903 incomplete responses were excluded from the data set, due to lack of data.

In sum, after subtracting a few cases where the credentials of the respondent did not match the credentials of the invited, this recruitment wave resulted in 2,041 new survey respondents, a recruitment rate of 14 percent. This is lower than what has been achieved in previous recruitment waves.

99.2 percent of the respondents who completed the survey entered their e-mail address. Of the excluded respondents, 0.8 percent entered their e-mail address. In sum, after subtracting respondents with mismatching credentials, 2,032 new **panel members** were recruited to the Norwegian Citizen Panel, resulting in a panel recruitment rate of 14 percent.

Further discussions in this report about new recruits in wave eighteen are based on data on survey respondents. Since there is an almost perfect overlap between survey respondents and panel members, the descriptions are also valid for the panel members.

RESPONSES BY METHOD OF DATA COLLECTION

Table 2 summarizes the effects of the various stages of data collection. The invitational letter accumulated 1,165 responses, the SMS generated 452 responses, and the postcard 424 responses: Resulting in a cumulative response rate of 14 percent.

Table 2: Number of responses and response rates for the new panel members by the various stages of data collection

	Response	Response rate (%)	Cumulative Responses	Cumulative Response Rate (%)
Invitation (3rd of June)	1165	8.0 %	1165	8.0 %
SMS, reminder (15th of June)	452	3.1 %	1617	11.1 %
Postcard, reminder (17th of June)	424	2.9 %	2041	14.0 %

RESPONSES OF EXISTING PANEL MEMBERS

Wave eighteen of the NCP also included data collection from existing members of the panel, recruited in waves 1, 3, 8, 11, 14 and 16. Data collection among existing panel members was conducted in parallel with the recruitment of, and data collection among, new members.

RESPONSES BY METHOD OF DATA COLLECTION

The survey was distributed via email to 23,103 existing panel members on June 2nd 2020. In these e-mails, the basic information about the Norwegian Citizen Panel was repeated, and the individual panel members received unique URLs that led to the questionnaire.

The invitation, the first reminder and the second reminder were all distributed via e-mail. The third, and last reminder was, depending on whether the individual panel member had a registered mobile phone number or not, distributed via SMS or e-mail. Prior to wave eighteen, 32.8 percent of the panel members were registered with a mobile phone number.

Table 3: Responses and response rate for panel members by the different stages of data collection

	Response	Cumulative Responses	Response Rate (%)	Cumulative Response Rate
Invitation (2nd of June)	5,644	5,644	37.9 %	37.9 %
1st reminder (9th of June)	2,037	7,681	13.7 %	51.6 %
2nd reminder (12th of June)	1,291	8,972	8.7 %	60.3 %
3rd reminder – email (18th of June)	1,027	9,999	6.9 %	67.2 %
3rd reminder – SMS (18th of June)	687	10,686	4.6 %	71.8 %

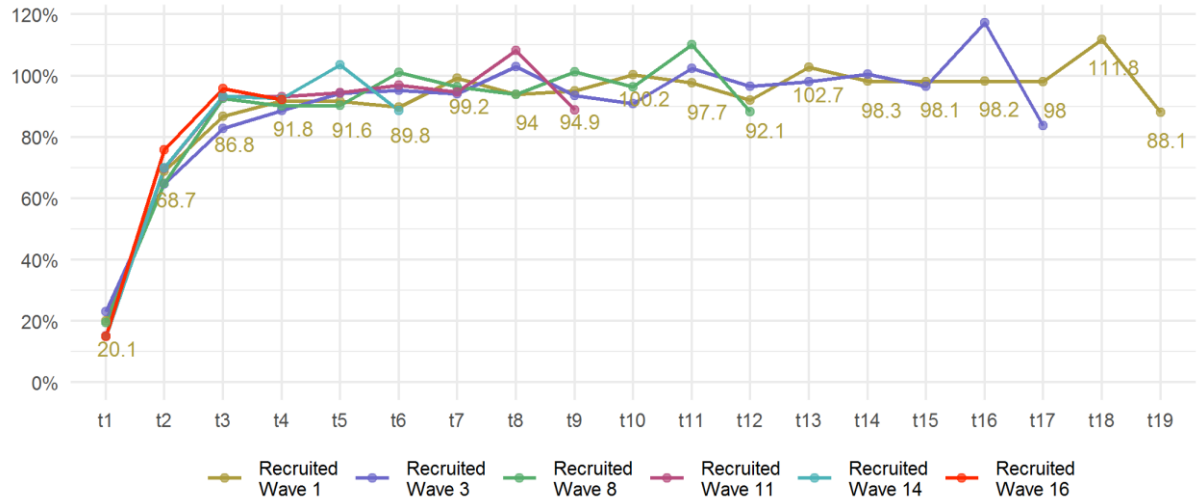
In total, 10,686 existing panel members filled out the questionnaire. A response rate of 37.9 percent was reached between the invitation and the first reminder (June 2nd - June 7th). Following a pattern observed in earlier waves, the email invitation produced a higher number of respondents than the subsequent reminders. For details on the number of respondents after each reminder, see table 3.

When calculating the response rate, we follow the methodology from earlier waves, and exclude respondents who have not participated in any of the last three waves. This leaves us with 14,879 eligible respondents. The overall response rate, as reported in table 3, is **71.8 percent**.

RESPONSE OF EXISTING PANEL MEMBERS OVER TIME

Comparing the number of wave eighteen respondents (10,686), to the number of respondents in the previous wave, Fast track one (12,050), gives an overall wave-to-wave retention rate of 88.7 percent. Fast track one was an extraordinary wave fielded in late march 2020, focusing on matters related to the Coronavirus pandemic. Figure 1 shows that the wave-to-wave retention rate normally increases substantially the first three waves, before stabilizing around a mean of 95 percent. In wave eighteen, however, the retention rate is rather low, following the high Fast track one response rate.

Figure 1: Wave-to-wave retention rate



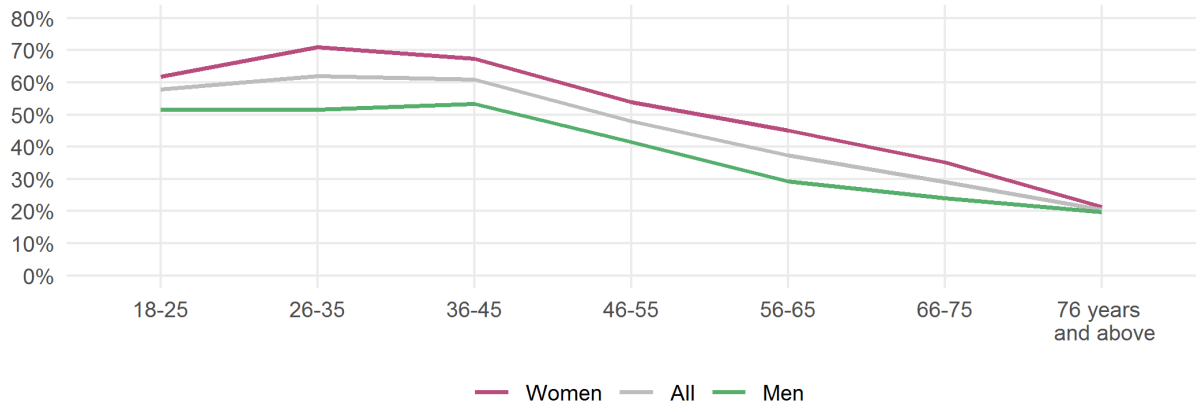
PLATFORMS

The questionnaire was prepared for data input via smart phones. In order to enhance the respondents' experience with the questionnaire, mobile users got a slightly different visual representation of most questions. For instance is a question grid presented as a set of individual questions on the same page, which is different from the desktop presentation where grid questions are presented in a table. 41.4 percent of all survey respondents that opened the questionnaire used a mobile phone.

A small number of survey questions must be answered for a person to be included as a survey respondent. 10.3 percent of the mobile users did not reach this minimum requirement, compared to 14.7 percent for non-mobile users. Opposite to what we have usually observed in the NCP, mobile users were thus more likely to complete the questionnaire than non-mobile users.

The share of mobile users is high among respondents between 18 and 45 of age. As shown in figure 2, the share of mobile users declines substantially with age, starting at age 46-55. Overall, women are more inclined to use a mobile phone to fill out the questionnaire than men are.

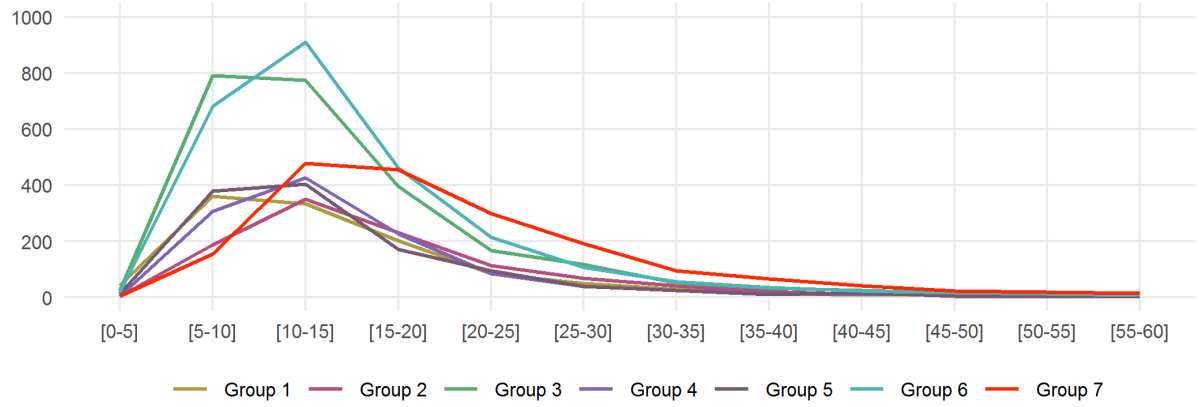
Figure 2: Share of mobile users by gender and age in wave 18



TIME USAGE

The average respondent used 15.9 minutes to complete the questionnaire. This is nearly a minute more than what the respondents were told upon invitation. Measuring average time usage poses a challenge, in that respondents may leave the questionnaire open in order to complete the survey later. This idle time causes an artificially high average for completing the survey. The average of 15.9 minutes therefore includes only the respondents which spent 60 minutes or less completing the survey.

Figure 3: Time usage of survey respondents in wave 18



The wave eighteen questionnaire consisted of seven subsets of questions, given to groups 1-7. The respondents recruited in wave eighteen were assigned to group 7. The remaining panel members were assigned to group 1, 3 or 6, based on participation in previous rounds; or by random to group 2, 4 or 5.

Compared to previously recruited respondents, newly recruited respondents (group 7), on average spent more time completing the questionnaire. This is partly explained by the fact that they, unlike already recruited panel members, were asked to register their email. Unfamiliarity with the NCP can also be an explanation.

Table 4: Average time usage (minutes) in each subset in wave 1

	All	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7
All users	15.9	14.5	17.2	14.6	15.0	14.3	15.1	20.2
Non-mobile users	17.0	15.4	18.2	15.7	16.1	15.2	16.2	21.5
Mobile users	14.3	13.1	15.8	13.2	13.7	13.2	13.7	18.0

As previously observed, mobile users on average use substantially less time filling out the questionnaire, than non-mobile users. In the wave seven methodological report, we found that mobile users spent less time answering 85 percent of the survey questions. The difference was particularly large in questions with long

and/or complex vignettes. Mobile users also gave substantially shorter responses to open text questions, and were, in the case of one particularly complex question more likely to state that they did not know what to respond.

REPRESENTATIVITY

In this section, we describe the representativity of the panel as a whole. First, we will discuss factors explaining representativity. Thereafter we apply demographic variables to present data on representativity by different strata. The data on representativity is the foundation for the section on weighting.

FACTORS EXPLAINING LACK OF REPRESENTATIVITY

There are two main points that can serve as explanations to non-response and lack of representativity when recruiting panel members and maintaining panel members:

- access to and familiarity with the internet (given that a web-based questionnaire was the only response mode made available)
- the motivation and interest of the respondents

The first challenge is strongly related to the age composition of the survey respondents. Although Norway has a very high computer and internet density, the probability of having an e-mail address, and the skills required to access and fill in an online questionnaire, normally decreases with increasing age. The second challenge, motivation and interest, is often explained by the respondents' level of education. In addition to age and education, we added the variables of geography and gender in order to test the representativity of the survey respondents. The variables have the following categories:

- Age: 19-29 years, 30-59 years, 60 and above.
- Highest completed education: no education/elementary school, upper secondary, university/university college.
- Geography: Oslo/Akershus, Eastern Norway, Southern Norway, Western Norway, Trøndelag, Northern Norway.

THE REPRESENTATIVITY OF THE NORWEGIAN CITIZEN PANEL

The sampling frame of the survey equals to the Norwegian population above the age of 18, comprising a population of approximately 4.2 million individuals. Earlier reports have documented a systematic underrepresentation of respondents belonging to the two lowest educational groups, independent of gender and age. The underrepresentation is particularly strong for young men. As expected, individuals with education from universities or university colleges are overrepresented. All of these observations are still true for wave eighteen.

Table 5: Age distribution in the population and the net sample of wave 18

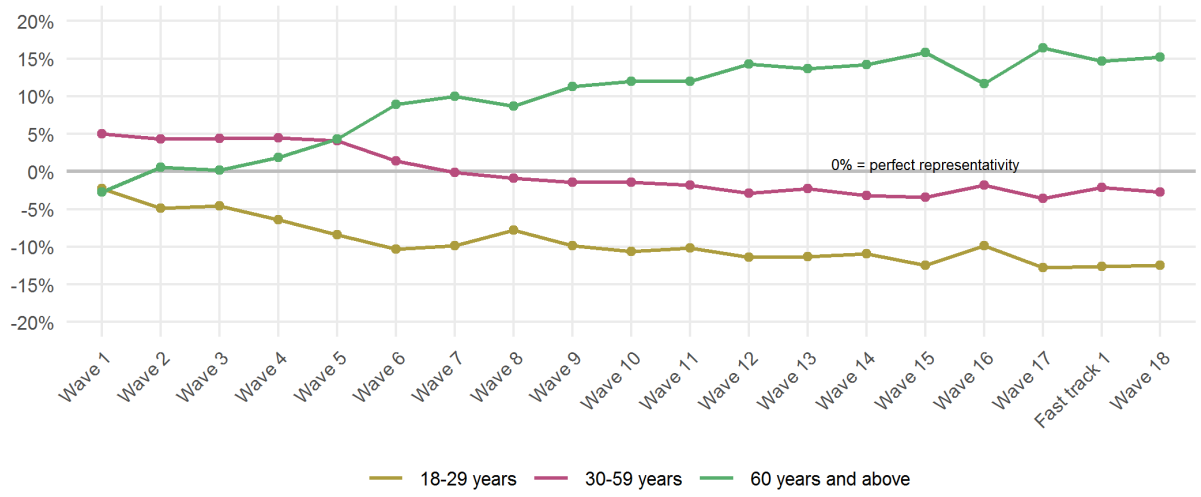
	18-29 years	30-59 years	60 years and above
Population	20.3 %	50.6 %	29.1 %
Net sample	7.9 %	47.9 %	44.3 %

From the age distribution presented in table 5, we see that 18-29 year olds are clearly underrepresented in the net sample of wave eighteen. The age group 30-59 years in the net sample is somewhat underrepresented compared to the distribution in the population, while respondents aged 60 years and above are clearly overrepresented.

Over time, we have observed a drift away from perfect representativity of age groups. While the oldest respondents started out as underrepresented in wave one, they have become increasingly overrepresented over time. The youngest respondents, on the other side, have become increasingly underrepresented. This can be explained by a difference in panel membership loyalty; younger panel members are more likely to stop responding to new NCP waves after having been an active member of the panel.

Representativity is usually improved in recruitment waves (wave 1, 3, 8, 11, 14, 16, and 18), as the newly recruited respondents more closely represent the population. An exception to this, is Fast track one, where representativity was improved, even without the recruitment of new panel members. This exception may explain why overall age representativity is not improved in wave eighteen (figure 4), with the exception of the youngest age bracket which is marginally improved.

Figure 4: Representativity of age groups from wave 1-18



In table 6, the population and net sample are broken down by age and gender. This reveals a gender-age interaction in the panel representativity. Younger men are more underrepresented than younger women, while older men are more overrepresented than women in the same age bracket. Lastly, middle-aged men are underrepresented, while women in this age bracket are slightly overrepresented.

Table 6: Combined distribution of age and gender in the population and the net sample of wave 18

	18-29 years		30-59 years		60 years and above	
	Men	Women	Men	Women	Men	Women
Population	10.4 %	9.9 %	26.0 %	24.7 %	13.8 %	15.4 %
Net sample	3.3 %	4.5 %	22.3 %	25.6 %	23.9 %	20.4 %

The inclusion of educational level in table 7 reveals a systematic underrepresentation of respondents with little or no education, independent of age and gender. The underrepresentation is present in all age brackets, but is especially strong for young respondents.

Table 7: Combined distribution of age, gender and education in the population and the net sample of wave 18

		Population		Net sample	
		Men	Women	Men	Women
No education/elementary school	18-29 years	3.8 %	2.9 %	0.3 %	0.3 %
Upper secondary education		4.2 %	3.3 %	1.6 %	2.0 %
University/university college		2.3 %	3.6 %	1.5 %	2.2 %
No education/elementary school	30-59 years	5.5 %	4.6 %	0.8 %	0.6 %
Upper secondary education		11.2 %	7.9 %	7.6 %	6.1 %
University/university college		9.3 %	12.2 %	14.0 %	19.0 %
No education/elementary school	60 and above	3.1 %	4.4 %	1.8 %	1.7 %
Upper secondary education		6.9 %	7.4 %	8.3 %	6.2 %
University/university college		3.8 %	3.6 %	13.9 %	12.2 %

Respondents that have upper secondary education as their highest completed education are underrepresented in all groups, except men with upper secondary education aged 60 years or above. Those who have university or university college education are clearly overrepresented in the two oldest age brackets, independent of gender.

Figure 5: Representativity of education groups from wave 1-18

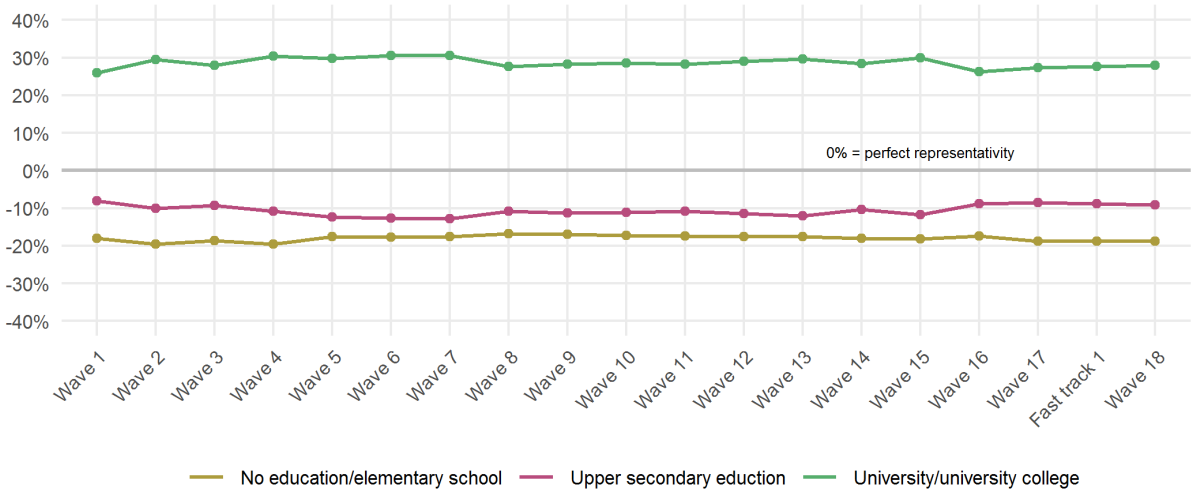


Figure 5 illustrates the representation of education groups since wave one. The general trend is that the highly educated are overrepresented compared to those with less or no education. Except for slight adjustments, improving the representativity of the education groups when new respondents are recruited (wave 1, 3, 8, 11, 14 and 16), the overall pattern has remained stable throughout all waves.

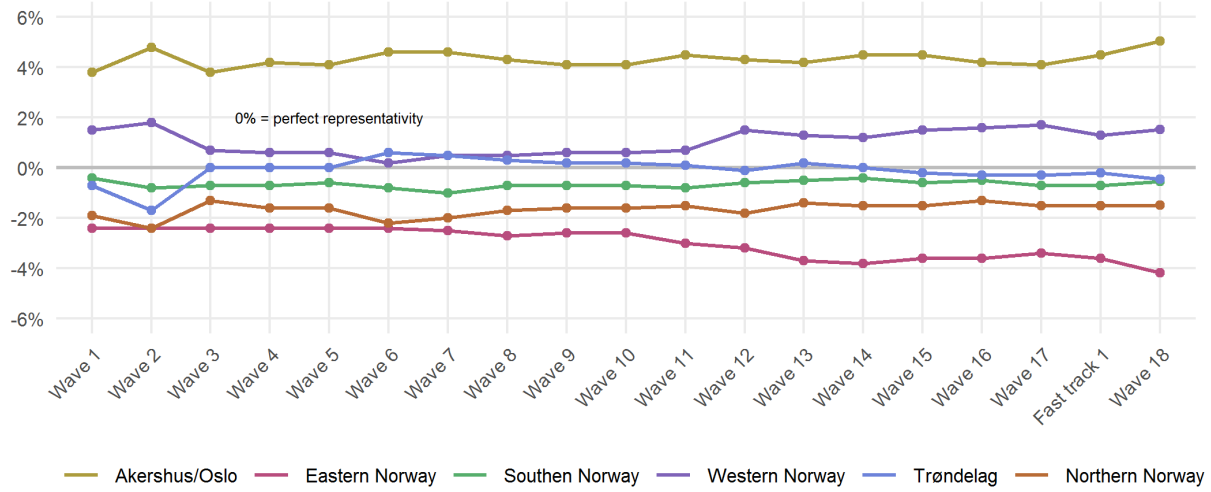
In regard to geography, (table 8) we observe that the representation of panel members living in Trøndelag and Southern Norway are on level with the population, while the capital region (the counties of Oslo and Akershus) is clearly overrepresented. Western Norway is also overrepresented, but not as prominent as the capital region. Northern Norway and Eastern Norway meanwhile are underrepresented among the respondents in the eighteenth wave.

Table 8: Combined distribution of age, gender and geography in the population and the net sample of wave 18

		Population			Net sample		
		Men	Women	Total	Men	Women	Total
Akershus/Oslo	18-29 years	2.6 %	2.6 %	5.2 %	0.9 %	1.4 %	2.4 %
	30-59 years	6.8 %	6.5 %	13.3 %	6.8 %	8.2 %	15.0 %
	60 and above	2.8 %	3.2 %	5.9 %	6.2 %	5.8 %	12.0 %
	In total	12.1 %	12.3 %	**24.4 %**	13.9 %	15.5 %	**29.4 %**
Eastern Norway	18-29 years	2.5 %	2.3 %	4.8 %	0.6 %	0.9 %	1.5 %
	30-59 years	6.6 %	6.4 %	13.0 %	4.4 %	5.1 %	9.5 %
	60 and above	4.1 %	4.6 %	8.7 %	6.0 %	5.3 %	11.3 %
	In total	13.2 %	13.3 %	**26.5 %**	11 %	11.3 %	**22.3 %**
Southern Norway	18-29 years	0.6 %	0.6 %	1.2 %	0.1 %	0.2 %	0.4 %
	30-59 years	1.4 %	1.4 %	2.8 %	1.2 %	1.3 %	2.5 %
	60 and above	0.8 %	0.9 %	1.7 %	1.3 %	1.0 %	2.3 %
	In total	2.8 %	2.8 %	**5.7 %**	2.6 %	2.6 %	**5.2 %**
Western Norway	18-29 years	2.7 %	2.6 %	5.3 %	1.0 %	1.0 %	2.0 %
	30-59 years	6.7 %	6.2 %	12.9 %	6.4 %	6.9 %	13.2 %
	60 and above	3.5 %	3.8 %	7.3 %	6.4 %	5.3 %	11.8 %
	In total	12.9 %	12.6 %	**25.5 %**	13.8 %	13.2 %	**27 %**
Trøndelag	18-29 years	1.0 %	0.9 %	1.9 %	0.4 %	0.6 %	1.0 %
	30-59 years	2.2 %	2.0 %	4.2 %	2.0 %	1.9 %	3.9 %
	60 and above	1.2 %	1.3 %	2.6 %	1.9 %	1.4 %	3.3 %
	In total	4.4 %	4.3 %	**8.7 %**	4.3 %	3.9 %	**8.3 %**
Northern Norway	18-29 years	1.0 %	0.9 %	1.9 %	0.2 %	0.3 %	0.6 %
	30-59 years	2.3 %	2.1 %	4.4 %	1.6 %	2.2 %	3.7 %
	60 and above	1.4 %	1.5 %	2.9 %	2.0 %	1.5 %	3.5 %
	In total	4.7 %	4.6 %	**9.3 %**	3.8 %	4.0 %	**7.8 %**

People aged 60 years and above, living in Akershus or Oslo, are quite overrepresented. This group accounts for 5.9 percent of the population, while making up 12 percent of wave eighteen respondents. Young people in all regions are underrepresented, as is middle-aged people living in Eastern Norway.

Figure 6: Representativity of regions from wave 1-18



The representativity of the regions has more or less been unchanged from wave 1 through wave seventeen (figure 6). Note that Akershus/Oslo and Eastern Norway diverge in wave eighteen. Compared to age and education, geography does, however, not seem to be a strong determinant of survey participation.

Please note that there is a known, but very small, bias, inflating the number of respondents from Oslo/Akershus and Trøndelag by 16 respondents, while deflating the number of respondents from Eastern and Western Norway accordingly. See the chapter on weighting for a more thorough discussion on this bias.

WEIGHTING

To compensate for the observed biases, we have calculated a set of weights. The weights are equal to the relation between a given strata in the population and the total population, divided by the relation between a given strata in the net sample and the total net sample.[5] This procedure returns values around 1, but above 0. Respondents belonging to a stratum that is underrepresented will receive a weight above 1 and respondents belonging to an overrepresented stratum will receive a weight below 1. We have listed the weights of the different strata in table 10 in the appendix.

When calculating the weights, information regarding the respondents' geographical location, gender and age is based on registry data. Information on these variables was included in the sample file we received from the Norwegian National Registry. Information regarding the level of education is collected from the NCP surveys. 2.6 percent of the eighteenth wave net sample have not answered the question about level of education. Because of this, two different weights have been calculated:

- **Weight 1** is based on demographic variables only (age, gender and geography)
- **Weight 2** combines the demographic variables with education. Respondents with missing data on the education variable are only weighted on demography (the education component of the weight is in these cases set to 1).

The variables have the following categories:

- **Age**: 19-29 years, 30-59 years, 60 and above.
- **Highest completed education**: no education/elementary school, upper secondary, university/university college.
- **Geography**: Oslo/Akershus, Eastern Norway, Southern Norway, Western Norway, Trøndelag, Northern Norway.

The method for calculating weights is the same as in previous waves.

When applied, both weights will provide a weighted N equal to the number of respondents in the dataset.

As discussed above, level of education is the greatest source of observed bias. We therefore recommend using weight 2 in most statistical analyses, as this weight provides the most accurate compensation for the various sources of bias in the net sample. Please note that there is some small biases to the wave eighteen weights. In the appendix, we provide an in depth explanation and analysis concluding that the biases are negligible, and that using wave eighteen weights is still recommended.

Table 9 shows the effects of weight 2 on the distribution of self-reported level of education in the net sample. As we can observe, the weight gives the sample a perfect distribution compared to the population. It is, however, important to stress that the unweighted distribution is far from ideal, with a clear underrepresentation of people with low levels of education.

[5] The applied formula for weight wi for element i, in strata h is: $w_i = \frac{N_h/N}{n_h/n}$

Table 9: Effect of weight 2 on self-reported level of education

	Sample - not weighted	Sample - weighted	Population	Difference between sample and population	Difference between weighted sample and population
No education/elementary school	5.5 %	24.3 %	24.3 %	-18.8	0
Upper secondary eduction	31.8 %	40.9 %	40.9 %	-9.1	0
University/university college	62.8 %	34.8 %	34.8 %	28	0

APPENDIX

Table 10: Weights applied to different strata (weight 2)

			Men	Women				Men	Women
Oslo/Akershus	18-29 years	No education/elementary school	14.5	7.9	Western Norway	18-29 years	No education/elementary school	8.6	9.3
		Upper secondary education	2.6	1.5			Upper secondary education	2.4	2.1
		University/university college	1.6	1.4			University/university college	1.3	1.7
	30-59 years	No education/elementary school	8.2	10.5		30-59 years	No education/elementary school	6.8	7.5
		Upper secondary education	1.4	1.2			Upper secondary education	1.3	1.1
		University/university college	0.7	0.6			University/university college	0.6	0.6
	60 and above	No education/elementary school	1.7	2.1		60 and above	No education/elementary school	1.3	2.7
		Upper secondary education	0.8	0.9			Upper secondary education	0.7	1.0
		University/university college	0.3	0.3			University/university college	0.3	0.3
Eastern Norway	18-29 years	No education/elementary school	16.2	13.8	Trøndelag	18-29 years	No education/elementary school	8.5	7.8
		Upper secondary education	3.5	1.6			Upper secondary education	2.3	1.3
		University/university college	1.9	2.2			University/university college	1.4	1.1
	30-59 years	No education/elementary school	7.7	7.1		30-59 years	No education/elementary school	7.1	5.0
		Upper secondary education	1.6	1.7			Upper secondary education	1.5	1.6
		University/university college	0.8	0.7			University/university college	0.6	0.8
	60 and above	No education/elementary school	2.2	2.6		60 and above	No education/elementary school	1.5	3.0
		Upper secondary education	0.9	1.4			Upper secondary education	0.9	1.9
		University/university college	0.3	0.3			University/university college	0.3	0.3
Southern Norway	18-29 years	No education/elementary school	28.5	4.3	Northern Norway	18-29 years	No education/elementary school	51.3	18.4
		Upper secondary education	3	3.8			Upper secondary education	2.7	1.5
		University/university college	1.9	1.6			University/university college	2.3	2.9
	30-59 years	No education/elementary school	5.1	8.5		30-59 years	No education/elementary school	6.4	5.7
		Upper secondary education	1.5	1.1			Upper secondary education	1.8	1.2
		University/university college	0.7	0.7			University/university college	0.7	0.7
	60 and above	No education/elementary school	1.8	2.9		60 and above	No education/elementary school	1.7	3.3
		Upper secondary education	1.1	1.3			Upper secondary education	0.9	1.6
		University/university college	0.3	0.4			University/university college	0.3	0.3

REGIONAL REFORM AND BIAS TO SURVEY WEIGHTS

In January 2020, several counties and municipalities were consolidated in a regional reform. This poses a problem, as we have to calculate the NCP survey weights using a mix of post reform and pre reform data.

When respondents are first recruited, we get a list of contact information from the Norwegian Population Registry. From this list, we know the municipality of residence for all panel members. We use this residential information to place respondents in geographical regions, as a step in the process of calculating the survey weights.

Survey weights are calculated by comparing the survey net sample to the general population. We achieve this by stratifying both the net sample and the general population by the exact same set of characteristics, and comparing the resulting strata.

Awaiting Statistics Norway population data on the post reform regions, we have to calculate weights based on pre reform population data. In wave eighteen, however, recruits (16 percent of the net sample) were drawn from the post reform population, meaning that we have information only on their post reform municipality of residence. As such, there is a mismatch between the geographic units in parts of the survey net sample and in the general population data. Because of this mismatch, we have to infer the pre reform region from post reform municipality.

Figure 7 illustrates the problem. If we know that someone live in what used to be municipality A, we can tell that they are now living in municipality C. If we know only that they are now living in municipality C, on the other hand, we have no way to tell if they previously lived in municipality A or in municipality B.

Figure 7: Illustration of municipality consolidation

This is, however, not a problem for all respondents, even if we cannot infer their exact pre reform municipality of residence. To see when this is a problem, let us examine the possible outcomes of the regional reform for a given municipality:

1. The municipality was unchanged and stayed in the same county as before.
2. The municipality was unchanged, but the county consolidated with other counties.
3. The municipality was consolidated with another municipality belonging to the same county.
4. The municipality was consolidated with another municipality belonging to another county.
    a. The municipalities belonged to the same region
    b. The municipalities belonged to different regions

In case 1, there can be no problem, as nothing changes. In case 2, only higher-level geographic units change, and we can still place the municipality in the correct pre reform region. In case 3, we cannot find the pre reform municipality, but we can still accurately place the respondents in the correct region, because we know the county to which they both belonged.

The problem arises in case 4, but is even more granular yet. If the consolidated municipalities belonged to the same region, as in case 4a, we still have no problem placing respondents in the correct pre reform region. This leaves us with case 4b, where municipalities were consolidated across regional borders, and where we have no way of accurately inferring the pre reform region.

As it turns out, subset 4b consists of only the six municipalities listed in table 11. To arrive at an inferred pre reform region, we had to make common sense judgements based on the population sizes in the pre reform municipalities. As shown in the table, we know that the incorrect region is inferred for three pre reform municipalities.

Table 11: Municipalities where pre reform region cannot be correctly inferred.

Post reform		Pre reform		
Municipality	**Region**	**Municipality**	**Region**	**Inferred region**
Asker	Eastern Norway	Asker	Akershus/Oslo	Akershus/Oslo
		Røyken	Eastern Norway	Akershus/Oslo
		Hurum	Eastern Norway	Akershus/Oslo
Heim	Trøndelag	Hemne	Trøndelag	Trøndelag
		Snillfjord	Trøndelag	Trøndelag
		Halsa	Western Norway	Trøndelag

Already, we have a clear indication that the bias is small, as the number and size of affected municipalities is quite small.

To estimate the resulting biases, we replicate the issue of incorrectly inferring region of residence, using wave sixteen data consisting of pre reform data. Note that the recruitment strategies in wave 16 and wave 18 were identical, with the exception of the number of invitees (34 000 vs 15 000). Therefore, we randomly sample down the wave sixteen data set to the same size as the wave eighteen dataset (12,727 respondents). This serves as our reference data, with which we can calculate weights correctly.

We then create a test data set based on the reference data. In the test data, we simulate the scenario in wave 18 by inferring pre reform region of the respondents living in the municipalities mentioned in table 11. First, we use the pre reform municipality to find the post reform municipality of the respondents. We then use the post reform municipality to infer back to the pre reform municipality. Here we apply the same rules as when creating the wave eighteen weights (table 10). Finally, we arrive at the inferred pre reform region, which we can use to calculate biased weights.

Figure 8: Manipulations for reproducing inaccurately inferred pre reform regions.

In our test, the incorrect region was inferred for only 16 persons, or 0.13% of the respondents.

To further test the bias, we calculate weight 2, on the benchmark data and test data, resulting in weights A and B respectively. Weight B is biased in a similar fashion to the wave eighteen weights, while weight A is correctly calculated, and serves as a benchmark.

Table 12 shows the calculated weights for all strata where the biased weight differs from the weight calculated on pre-reform data. From this reproduced bias, we conclude that the biases to the wave eighteen weights are negligible.

Table 12: Weights A and B, sorted by the difference between the weights.

Sex	Age	Region	Level of education	Weight A	Weight B	Difference
Women	18-29	Eastern Norway	University/university college	2.325	2.389	0.065
Men	60+	Oslo/Akershus	No education/elementary school	1.686	1.654	-0.032
Men	60+	Eastern Norway	No education/elementary school	2.214	2.242	0.028
Men	30-59	Oslo/Akershus	Upper secondary eduction	1.402	1.422	0.020
Men	30-59	Eastern Norway	Upper secondary eduction	1.130	1.115	-0.015
Women	18-29	Oslo/Akershus	University/university college	1.393	1.408	0.014
Women	60+	Oslo/Akershus	Upper secondary eduction	1.376	1.362	-0.014
Women	60+	Eastern Norway	Upper secondary eduction	1.440	1.426	-0.014
Men	30-59	Eastern Norway	University/university college	1.549	1.562	0.013
Men	60+	Oslo/Akershus	Upper secondary eduction	0.988	0.978	-0.010
Men	60+	Eastern Norway	Upper secondary eduction	1.482	1.490	0.008
Women	30-59	Eastern Norway	University/university college	0.768	0.775	0.007
Men	60+	Eastern Norway	University/university college	1.162	1.155	-0.007
Men	30-59	Oslo/Akershus	University/university college	0.779	0.785	0.005
Women	60+	Eastern Norway	University/university college	0.557	0.555	-0.003
Men	60+	Oslo/Akershus	University/university college	0.702	0.699	-0.002
Women	60+	Oslo/Akershus	University/university college	0.309	0.311	0.002
Women	30-59	Oslo/Akershus	University/university college	0.265	0.264	-0.001

Lastly, we test the weights on a survey question. We selected a question that was given to all wave sixteen respondents, and that contained responses that varied across the weighting parameters. While the question itself is of minor importance, we chose the following question: "Which political party/party list did you vote for in the local elections this autumn?" (r16k43_1). As shown in table 13, there are only minor differences in the biased (B) and unbiased (A) weighted distributions. We thereby conclude that we have no reason to believe that using the wave eighteen weights will have any large impact on analyses.

Table 13: Weighted and unweighted distribution of party voted for in municipal election.

	Unweighted	Weight A	Weight B
The Christian Democrats	3.36 %	2.81 %	2.81 %
The Conservative Party	19.14 %	16.42 %	16.41 %
The Progress Party	5.23 %	5.59 %	5.59 %
The Liberal Party	3.95 %	3.52 %	3.51 %
The Socialist Left Party	7.95 %	6.83 %	6.83 %
The Centre Party	12.08 %	12.95 %	12.97 %
The Green Party	8.73 %	8.41 %	8.4 %
The Labour Party	20.73 %	20.26 %	20.27 %
The Red Party	4.78 %	4.67 %	4.66 %
The People's Movement No to More Road Tolls (or similar toll initiative)	2.07 %	2.37 %	2.37 %
A joint party list	0.19 %	0.2 %	0.2 %
A local party list	1.67 %	1.77 %	1.77 %
Voted blank	1.01 %	1.57 %	1.58 %
Did not vote	6.42 %	9.5 %	9.5 %
Not entitled to vote	0.51 %	0.7 %	0.7 %
Other:	2.19 %	2.42 %	2.42 %