

Multi Topic Implementation at Statistics Norway

Jan Haslund and Trond Båshus, Statistics Norway

1. Introduction

Statistics Norway has used the CATI-functionality in Blaise since 2005. At the start in combination with list based offline interviewing (CAPI) and later also with CAWI. Until recently we have used the traditional single topic approach. This approach has significant shortcomings such as high workload for the supervisors who actively have assign specific surveys to individual interviewers. During daytime when there are few interviewers working it is a challenge to handle appointments because they must switch between surveys. In case of small surveys there were also a common problem that the daybatch quickly became exhausted which required supervisor intervention.

We started exploring Multi Topic in 2023 to if multi topic could solve the challenges with traditional approach. This involved much technical testing and consultation with supervisors, interviewers and other users. The first survey in production in the Multi Topic solution was the Media Survey in September 2024. Later together with a second survey Housing Panel Survey. Both surveys had relatively few cases and were only overlapping for one week. This experience was positive, and we decided to run all newly deployed CATI surveys in Multi Topic from January 2025. This includes the Labour Force Survey, the Omnibus survey and EU-SILC. The Labour Force Survey, Housing Panel Survey and Media Survey are also multi-mode (CAWI).

In preparation for Multi Topic we have changed the questionnaire templates, started using the appointment control in combination with operation times. When it is necessary to prioritise specific respondents, we will make use of the sub priority functionality in Blaise, in place of groups, which was very difficult to manage for the supervisors. We are also using milestones to correctly manage the workload. We have made extensive use of data from the dial history of previous rounds to calculate correct workload for each response.

2. Technical Exploration in 2023 and Transition

We had been considering Multi Topic for quite some time but only began our first experiments in 2023. The initial impressions were positive, but various technical issues, combined with a lack of time, brought the exploration to an early halt.

The first survey we conducted using Multi Topic was the Media Survey in September 2024. This was a multi-mode survey, allowing respondents to participate via both web and CATI. Additionally, another survey overlapped for one week: the Housing Panel Survey, which was also a multi-mode survey with a few CATI interviews conducted in the final week of data collection.

The experience gained from running these two smaller surveys using Multi Topic was positive, leading us to decide on full implementation for all surveys from January 2025. Since it remains possible to initiate surveys in the traditional way, we considered the risk to be minimal.

3. Challenges with data collections

We face several challenges in our data collection processes. Almost all of our data collections differ in structure and methodology.

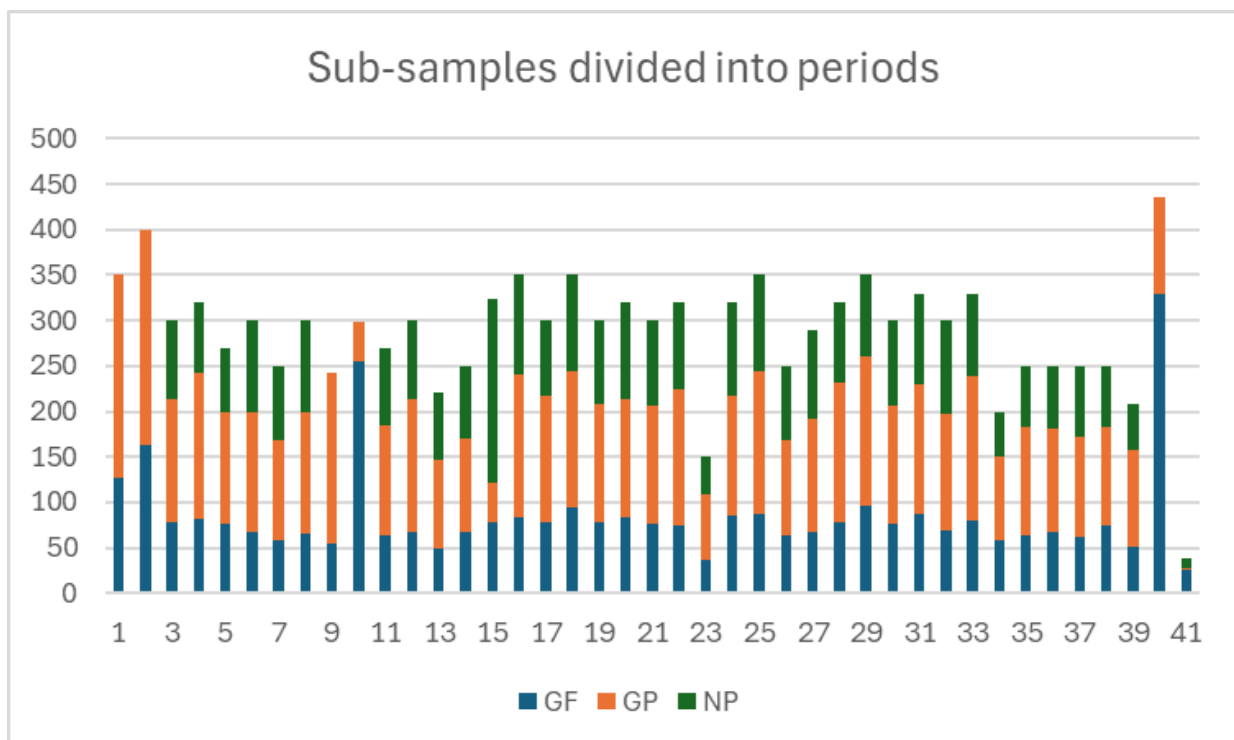
- **Housing Panel Survey:** This is a multi-mode survey that starts with web-based responses for one week, followed by CATI follow-ups in the second week, while the web option remains open. Since this is a panel of addresses, it can be challenging to track new residents when the

previous ones move out. Although we have a registry, we depend on new residents registering their updated addresses.

- **Labour Force Survey:** This survey is divided into periods, ensuring approximately the same number of respondents each week. Until now, it has been conducted exclusively via CATI. However, starting from the second quarter, it will transition to a multi-mode approach, with some respondents answering online and others via CATI.
- **Media Survey:** The goal of this survey is to collect a consistent number of responses for each day of the week, as most questions refer to the previous day. It is also conducted in a multi-mode format.
- **Cultural Use Survey:** This survey consists of two samples—a main sample and a supplementary sample of immigrants. Both samples follow a multi-mode approach: the main sample starts online, while the supplementary sample begins with CATI. After a few weeks, they switch modes.
- **EU-SILC:** This is a panel survey where each respondent participates for four years. As a result, each annual sample consists of three groups:
 1. Respondents who participated the previous year.
 2. Non-respondents from the previous year.
 3. New panel sample.

These groups have different response rates, with last year's non-respondents generally having the lowest response rate, while the new panel sample falls somewhere in between. The sample is divided into nearly 40 periods, but the distribution of the three groups across these periods is uneven.

This creates challenges in predicting the weekly response rate. As illustrated in the figure, the three groups are not evenly distributed across the periods. Additionally, we have divided the sample into many periods because we aim to send email and SMS reminders to respondents shortly before initiating phone calls. New periods are introduced twice a week to optimize response rates.



GF = Non-response last year.

GP = Interview last year

NP = New panel sample

Since many of the surveys are conducted in a multi-mode format, we continuously update the CATI system with results from the web data collection. This process is managed by running a script that updates the launcher database every night. Additionally, web responses received during the day are automatically removed from the daybatch to prevent interviewers from contacting respondents who have already completed the survey online.

4. Learning by Doing: Initial Experiences with Multi Topic

We started production with a "learning by doing" approach, fully aware that challenges would arise along the way. Our goal was to address these challenges as they emerged.

Week 1: Initial Implementation

In the first week, we launched the Labour Force Survey and the Omnibus Survey, both conducted exclusively via CATI. Although efficiency was slightly lower than expected, likely because Multi Topic was new to the interviewers, the system functioned as anticipated.

Week 2: Adding the Housing Panel Survey

In the second week, we introduced the Housing Panel Survey alongside the two existing surveys. While we encountered technical issues with this survey, they were unrelated to Multi Topic. These issues had some impact on data collection for the other surveys, but overall, the system performed as expected.

Week 3: Challenges with EU-SILC Training

In the third week, we launched the EU-SILC survey under the Multi Topic framework. Since all interviewers could now receive respondents from this survey, proper training was required before they could start. The plan was for interviewers to:

1. Read the instructions.
2. Conduct a training interview before engaging with real respondents.

However, training all interviewers on the same day would have significantly reduced the number of interviews for the other surveys. To avoid this, we implemented a phased approach:

- A selected group of interviewers began EU-SILC training on Monday.
- More were added on Tuesday.
- By the end of the week, all interviewers were trained.

To solve this, we introduced a "EU-SILC skill" to the interviewers that has past the training. What we had not anticipated was that the Monday-trained interviewers, once active in Multi Topic, also started receiving respondents from other surveys. As a result, we conducted too few EU-SILC interviews while maintaining expected interview volumes for the other surveys.

To resolve this, we adjusted the system so that EU-SILC interviewers received a direct link to the survey. This correction ensured the survey ran smoothly for the rest of the week.

Week 4: System Bugs and Workload Distribution Issues

By the fourth week, almost all interviewers were trained, and we assigned the "EU-SILC skill" to all, enabling a full Multi Topic rollout. However, challenges persisted:

- **Daytime Challenges:** During the day, only a few interviewers were available, primarily handling scheduled appointments. This made it difficult to assess whether the system was functioning correctly.
- **Bug in skills Handling:** On Monday, a system issue caused interviewers to receive only appointment-based respondents, rather than a mix of appointments and default cases. To fix this, we had to remove the startup skill and reinstall the survey on the server. This resolved the issue, and the system functioned as expected for the next couple of weeks.

Appointment or random vs. Weighted Workload algorithm

We also tested the appointment or random delivery feature for Multi Topic. Our expectation was that the system would randomly select respondents, giving surveys with a higher number of available respondents in the daybatch a greater likelihood of being chosen.

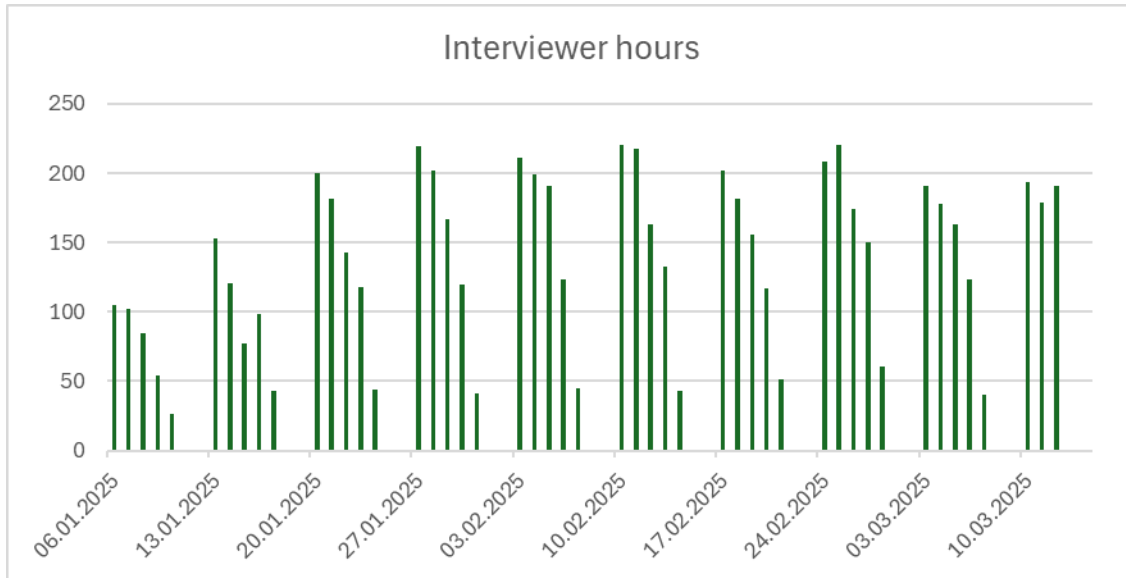
However, the system instead picked surveys randomly, meaning every survey had an equal probability of selection, regardless of the number of available respondents.

To address this, we explored the weight factor but discovered that it only worked with the weighted workload algorithm, not with appointment-based or random selection methods.

Weighted Workload Algorithm and Weighting Issues

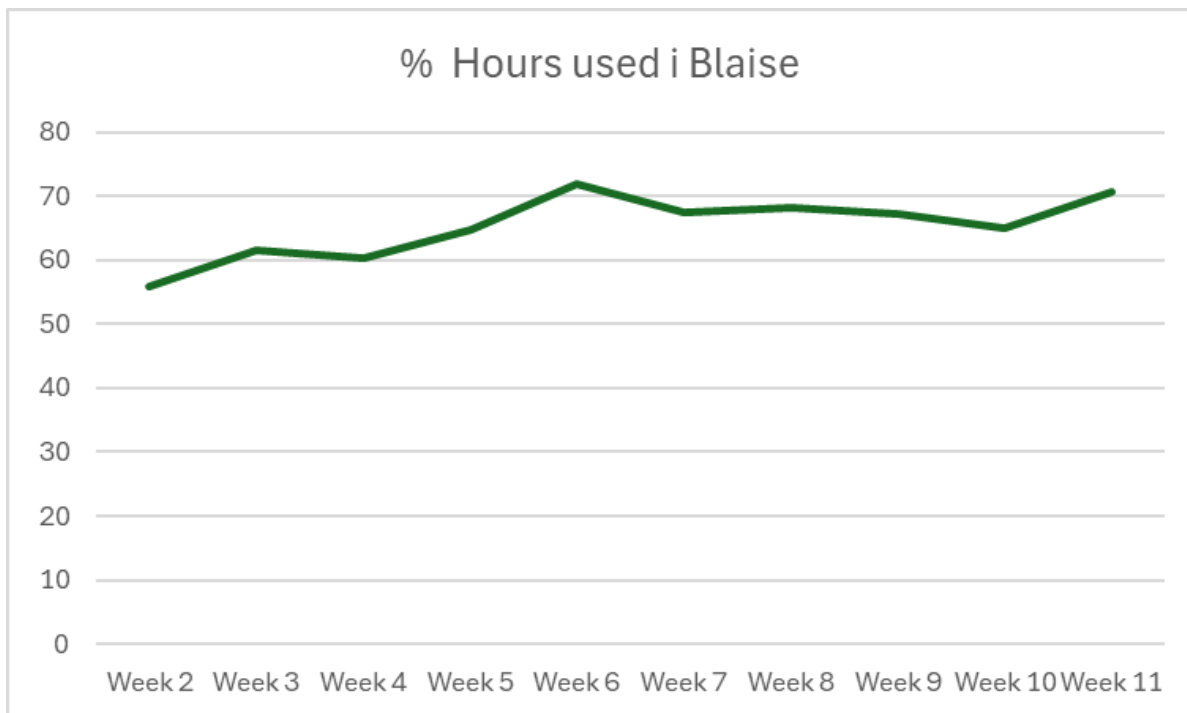
Later, we encountered another major challenge: workload distribution. One day, we noticed that the workload had dropped to zero. The issue stemmed from the linear nature of the weighted workload algorithm, which did not align with our data collection needs.

- **Mismatch Between Algorithm and Data Collection Goals:**
 - The algorithm assumes a steady response rate over time.
 - Our approach requires a higher number of interviews at the start and fewer towards the end of the survey period.
- **Testing Milestones:**
 - We initially set a milestone of 75% completion after two weeks, expecting it to prioritize interviews early in the process.
 - Instead, the system waited until the two weeks had passed before realizing it was behind schedule. At that point, it suddenly increased the workload, leading to an uneven distribution of interviews.



We have a fieldwork plan to allocate interview hours efficiently across different surveys. As shown above, the number of interview hours is highest on Mondays and gradually decreases throughout the week, reaching its lowest point on Fridays.

At the beginning of the week, we allocate approximately 200 interview hours per day on Monday, which then declines to 40–50 hours by Friday. This distribution reflects our strategy for optimizing interviewer availability and balancing the workload over the course of the week.



The recorded interview hours are measured in Blaise, not the hours interviewers are paid for. The data above shows percentage of the total time which is spent in Blaise, and we have observed that the percentage of time spent in Blaise increased during the first week. It now appears that approximately two-thirds of the total time is spent conducting interviews in Blaise. Some variation is expected due to interviewers practicing new surveys, attending briefings, breaks, and participating in other meetings.

We also monitor the workload for each survey on a daily basis. Once we understood how to set milestones correctly, ensuring that they do not result in a workload of zero, we realized that we could calculate a weight factor for each survey. This allows us to achieve a near-optimal distribution of interviews across surveys.

1. Remaining Challenges

1. Impact of Appointments on Workload Distribution

- When there are many scheduled appointments, especially for long interviews like EU-SILC, the survey may take up more time than intended.
- However, our priority remains ensuring that interviewers spend their time conducting interviews rather than making repeated call attempts.

2. Overly Optimistic Targets for Surveys

- One major issue is that our survey response rate targets are often too ambitious.
- Interviewers frequently call in sick, leading to fewer available interviewer hours than planned.
- As a result, we fail to reach the weekly target for completed interviews, leading to complaints from survey managers who feel their surveys are not being prioritized.
- This challenge is not unique to Multi Topic, but before its implementation, complaints were directed at the staff responsible for distributing interviewers to surveys. Now, frustrations are often directed at the "system" itself.

3. Understanding the Weighted Workload Algorithm

- The weighted workload algorithm remains difficult for all involved users to fully understand.
- This lack of clarity makes it challenging for field managers to efficiently distribute interview hours between surveys.

5. Balancing Interview Hours Across Surveys

Since we face challenges in distributing interview hours across surveys on a daily basis, we needed to find a viable solution. After testing multiple approaches, we found a method that works reasonably well for us.

We now use a combination of workload calculations from Blaise and hour allocations determined by field managers for each specific day.

Example Calculation

In this example, the field managers allocate hours as follows:

	Hours from field managers	% of hours	Workload from Blaise	% of workload	factor
Survey A	60	42,9	541215	30,6	140
Survey B	30	21,4	452151	25,5	83
Survey C	50	35,7	777777	43,9	81
Total	140	100	1771143		

To calculate the adjustment factor for each survey, we divide % of hours by % of workload and multiply by 100.

For example, for Survey A:

$$(42,9 \div 30,6) \times 100 = 140$$

These factors are then entered into the weight factor settings in our dashboard to ensure that the final distribution aligns with the field managers' allocations.

Remaining Challenges and Adjustments

1. Interview Duration Bias

- Longer interviews tend to consume more time than shorter ones, leading to imbalances in time distribution.
- Appointments are given higher priority than randomly assigned respondents, which can further skew allocations.

2. Workload Reaching Zero

- If the workload for a survey drops to zero, the weight factor no longer has any effect.
- To prevent this, we introduce milestones when the workload is lower than the allocated hours.

3. Milestone Adjustment

- We considered setting a milestone at 100% on Day 1 to ensure that the workload remains above zero for as long as there are available interview hours.
- This approach helps maintain a steady flow of respondents and prevents the system from deprioritizing surveys too soon.

So far, this hybrid approach appears to be working well. While some challenges remain, the combination of field manager allocations, Blaise workload calculations, and milestone adjustments has improved the distribution of hours across surveys.

Throughout the transition to Multi Topic, we have maintained extensive dialogue with the Blaise team at Statistics Netherlands to discuss Multi Topic functionality and clarify various system-related issues. This collaboration has been invaluable in helping us gain a deeper understanding of the system, especially given that the help documentation can sometimes be sparse and difficult to navigate.

6. Old strategies used for CATI surveys not necessarily suitable for Multi Topic

The mix of different respondent types within batches makes it challenging to manage time effectively in the system. Some EU-SILC batches that were activated consisted primarily of non-respondents from previous years, who are generally more difficult to reach and require significantly more time for interviewers to handle.

7. Other advantages of using Multi Topic

Apart from its benefits for CATI management, Multi Topic offers several other advantages. The most obvious is the separation of personal data, such as names, addresses, and telephone numbers, from survey data. However, this advantage is somewhat diminished when using a household roster or when external data is incorporated into the survey. Examples of this include employment information in the

Labour Force Survey and the household roster in EU-SILC, as they are currently implemented at Statistics Norway.

Another key advantage is that all CATI management functions have been moved from the questionnaire to the Launcher. This significantly simplifies the handling of multi-mode surveys (CATI and CAWI). Previously, certain logic required for CATI, at least as implemented at Statistics Norway, made it more difficult to transition between modes, such as when following up on non-respondents in a different mode.

At the same time as the transition to Multi Topic, a decision was made to reduce the use of groups for prioritizing cases. Groups are an excellent tool for matching respondent and interviewer language or directing non-response follow-ups to interviewers with specific skills. However, at Statistics Norway, groups were also used to prioritize cases that required special attention. The move to Multi Topic has shifted some of this prioritization logic, making case distribution more flexible and adaptable.

8. Suggestion for changes and improvements for the system

Considering the complexity of our surveys, particularly with regard to respondent composition in panel surveys, and the way we activate new batches of respondents weekly in the daybatch, the weighted workload algorithm does not function optimally for our needs in its current form. Additionally, the declining availability of interviewers throughout the week presents an additional challenge.

As an alternative, we propose a simpler method based on our current implementation strategies. Instead of relying on the weighted workload algorithm, we suggest allowing users to directly input the desired workload distribution (in percentages) into the dashboard.

This approach does not require tracking workload from previous survey days to function effectively. However, it should still monitor total time spent across all surveys for reporting and monitoring purposes.

Example of direct workload distribution input:

	% of workload
Survey A	30,6
Survey B	25,5
Survey C	43,9
Total	100

Currently, the random or appointment algorithm selects a survey at random, without considering the number of available cases in each survey.

We propose modifying the algorithm so that it selects randomly from all available cases in the daybatch, regardless of which survey they belong to.

Additionally, this alternative method could incorporate other selection criteria, such as:

- The number of previous call attempts to a respondent.
- Prioritization based on survey-specific quotas or response needs.

This adjustment would improve workload distribution and ensure a more balanced and efficient selection process for interviewers.

The portal does not currently work, but we hope this will be fixed in a future bug fix release soon. The portal will be useful when we need to allocate interviewers to specific surveys, for example, when we have newly recruited interviewers who must start with a specific survey because of training.

9. Conclusion

The implementation of Multi Topic at Statistics Norway removed the need for manual survey allocation and reduced the workload for the CATI supervisors. Appointment handling during daytime is more efficient, however we don't know if the overall data collection on CATI has become more efficient.

We have separated survey data and personal data. This makes the data processing simpler, more secure, and faster. Multi-mode is easier to maintain. All the personal data is in one place. The interviewers are satisfied and are happy with the variation in surveys. They don't need to switch surveys during the day.

Managing workload distribution, particularly for longer interviews and appointment-heavy surveys, requires continuous adjustments. The weighted workload algorithm, while useful in principle, does not fully meet the needs of Statistics Norway's survey structure. To address this, a simpler method of defining workload distribution directly in the dashboard has been proposed.