



This module is part of the

# Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

**Theme: CATI Allocation**

**Contents**

General section ..... 3

- 1. Summary ..... 3
- 2. General description..... 3
  - 2.1 Introduction ..... 3
  - 2.2 On the scheduling and allocation problems in CATI allocation..... 4
- 3. Design issues ..... 5
- 4. Available software tools..... 6
- 5. Decision tree of methods ..... 6
- 6. Glossary..... 6
- 7. References ..... 6

Interconnections with other modules..... 7

Administrative section..... 8

## General section

### 1. Summary

To be able to perform CATI interviews sample elements have to be linked to CATI interviewers at some point, so that the interviewers can call and interview them by telephone. In order to be able to do this, several steps have to be taken. An important one is that interviewers have to be scheduled in a timetable, taking various conditions into account, such as days and/or parts of the day where interviewers are not able to work, the surveys for which the interviewers are trained and for which they can be deployed, who has to be called (in our case: businesses) and the time of day when this preferably should be done in order to increase the response rate. If the interviews are conducted from a call room, say at the premises of a national statistical office, its maximum capacity has to be taken into account (if the number of interviewers that can be employed simultaneously exceeds this capacity). The interviewers present in the call room (at a certain day, part of day (DPoD) combination) have to be fed with telephone numbers of sample elements to be contacted, as they were drawn into respective samples, so the interviewers can call them for interviews. The most difficult problem is the allocation of interviewers to DPoDs and surveys. This can be done by hand, but (preferably) by using optimisation models. It is possible to take deadlines for surveys into account in these models, or (salary) costs of the interviewer corps that should be adequate for its task, but not bigger than necessary. When these allocations have been made (and the interviewers know when they should come to work and for what surveys), they have to be 'fed' telephone numbers. We assume that the interviewers receive such telephone numbers 'on their demand', whenever they have indicated to be ready for a new interview.

A special issue is about the updating of an allocation that has been determined for a planning period. This is necessary as things change. Time flows and plan and reality tend to diverge more and more from each other as time passes. Interviewers get ill, go on vacation, have to take leave, quit their job, new ones get hired and trained. Also as time passes, new surveys come into sight that also have to be carried out. All these changes require a schedule to be updated, very regularly.

It should be noted that instead of a single method, in reality there is actually a complex of models, that are rather strongly related and that differ by using certain alternative constraints (or discarding such constraints), or certain objective functions that are different (but with the same set of constraints).

### 2. General description

#### 2.1 Introduction

The subject of this theme is about allocation problems in CATI surveys. Ultimately it is about allocating sample elements to interviewers. This allocation happens in a few steps. As we assume that the interviewers work from a call room (say at the premises of an NSI) it has to be made clear which interviewer works when (on what DPoD combinations), and on what surveys (among those that are 'active' during the planning period). We assume that the allocation of sample elements to interviewers (and surveys) for CATI surveys is in the following order of steps:

1. **Scheduling:** The interviewers are allocated to admissible Day-Part of Day (DPoD) combinations, as well as to surveys they are supposed to work on.

2. **Allocation of workplace:** The interviewers are allocated to a specific workplace in the call room before they start their work on a particular DPoD combination.
3. **Allocation of telephone numbers:** Telephone numbers are allocated to interviewers when they indicate to be ready for a next interview. These telephone numbers are the direct links between interviewers and sample elements.

The first step is difficult and leads to all kinds of optimisation problems, depending on the goals that one pursues.

The second step is generally taken by some supervisor who knows the interviewers well and who knows which interviewers should sit together and which not. This step is not suitable for formalisation and automation, as it is likely to have little added value. For that reason it is not described here.

The third step is often carried out with specialised software, namely CATI call management software. Such a system assigns a suitable telephone number to an interviewer who asks for a number. The program then picks a suitable telephone number from a list. It 'knows' to which sample element it belongs, and in particular it 'knows' for which particular survey the sample element is intended. It also 'knows' for which surveys the interviewer asking for a new telephone number is qualified for. There are several variants possible for this step.

After the third step the interviewer can call the telephone number and try to interview the sample element to which this number belongs. This interviewing is not part of the allocation step considered in the present module. See, however, other modules in the topic "Data Collection".

The steps above describe a static situation with a fixed planning period. In practice the planning is not static but dynamic. For all sorts of reasons reality may diverge from a planning: interviewers get sick, or move to another job, etc. As time proceeds, new days become available where new interviews can be planned. So in order to keep the planning up to date, new information has to be 'injected' into an existing (or current) planning. So this corresponds to a fourth step:

4. **Update the schedule:** Recent update information is used to update the current schedule.

We consider this step as a separate one, but it could be well considered an integral part of the first step. It is likely to be carried out with the same software.

## 2.2 *On the scheduling and allocation problems in CATI allocation*

The scheduling problem is a matching problem, where interviewers are linked to DPoD combinations and surveys. The matching uses various constraints, which reflect wishes and demands from the NSI that conducts the surveys, on one hand, and from the interviewers, on the other hand. Relevant input data for the scheduling problem are the following:

- It should be known which surveys are 'active' in the planning period considered.
- For each interviewer it should be known for which 'active' surveys he/she is qualified. More than one is possible. This means that such a person is allowed to conduct interviews for such surveys. This qualification is a result of successfully completing a relevant training.
- It should be known for each interviewer on which DPoD combination he/she is available.

- The interviewers are supposed to work from a call room, sitting at a desk (a workspace). This call room may physically consist of several rooms. Important in this case is that the capacity (the number of interviewers it can take simultaneously) is limited. This should be taken into account when planning the work. Another option would be that the interviewers work from home. In this case there is no limitation to the number of interviewers that can work simultaneously.
- It is possible to take into account specific wishes and demands. For instance the deadline of a survey and the maximum work load for interviewers. In case of a deadline, one should be able to estimate the number of hours needed to finish the work on time, that is, before the deadline.

There are also various objectives that can be considered, each resulting in specific objective functions. The scheduling problem is translated into an optimisation problem, which tries to maximise an objective function under a set of constraints. The problems that may arise in this way might be difficult to solve (currently), in which case one should look for simplified models that are tractable, and that yield approximate solutions that are good enough to be useful in practice.

The method allows for telephone numbers that have been called but without getting a response and can be called again. How often a number is called is part of the contact strategy. This is a separate topic, outside the present theme, but obviously related to it. It is obvious that in case call-backs are allowed, it is necessary to estimate how many there will be and how much time they take to handle. Otherwise it is impossible to say anything about deadlines.

Once the interviewers are at work in the call room they need to be provided with telephone numbers of sample subjects that they should call and interview. It is assumed that each interviewer gets a telephone number each time it is requested. A request indicates that an interviewer is ready for a next interview. There are various possibilities how telephone numbers can be allocated to interviewers.

As time proceeds, reality may deviate from the schedule calculate. As explained above this requires an update step, which is a separate step in the CATI allocation process.

Solving the CATI allocation problem can be done by formulating an appropriate optimisation model. See, e.g., Willenborg (2012) for details.

### **3. Design issues**

We assume that it has been decided that a survey is needed to collect certain data from a target population. The details about this are not of interest to our allocation problem. What matters is that the elements in the sample have to be contacted and that this contact should result in information useful for the survey.

There are several options ('modes') available to contact sample elements: PAPI, CAPI, CATI, CAWI and mixed forms of these modes. What option to choose is a matter of weighing several factors: appropriateness, amount of work for the interviewee to provide the data using a particular interview mode, amount of work for the statistical office to prepare and execute a survey using a particular mode of data collection, effort needed to collect and prepare the data, quality of the data collected. (For more information on this see the theme module "Data Collection – Design of Data Collection Part 1: Choosing the Appropriate Data Collection Method".) No mode is perfect. However, there is a trend to move away from PAPI interviews to modes that yield electronic data right away, that is CAPI, CATI

and CAWI. The first two of these employ interviewers, which has several advantages. CAWI is self-administered: it is (by far) the cheapest variant of the three mentioned, but it lacks the help and guidance of interviewers. CATI is much cheaper than CAPI, as no travel is needed from interviewers. But it poses limitations to the questionnaires being used concerning the kind, amount and complexity of information that can be gathered.

So depending on the kind of information that one wants to collect, and the amount of money available, a decision has to be made on what mode to use. It is even possible – and quite modern – to consider surveys that use a combination of these basic modes (mixed mode). More information can be found in the theme module “Data Collection – Mixed Mode Data Collection”.

#### **4. Available software tools**

Blaise® is a flexible and powerful computer assisted interviewing (CAI) system and survey processing tool for the Windows® operating system. Part of Blaise® is a CATI Call Management System. See <http://www.blaise.com/Description> for more information.

#### **5. Decision tree of methods**

#### **6. Glossary**

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

#### **7. References**

<http://www.blaise.com/Description> (information on the Blaise system)

Willenborg, L. (2012), Allocation of sample units to interviewers in CATI surveys. Report, Statistics Netherlands, The Hague.

## **Interconnections with other modules**

### **8. Related themes described in other modules**

1. Data Collection – Design of Data Collection Part 1: Choosing the Appropriate Data Collection Method
2. Data Collection – Mixed Mode Data Collection

### **9. Methods explicitly referred to in this module**

- 1.

### **10. Mathematical techniques explicitly referred to in this module**

1. Scheduling / planning – calculation of a plan / roster for interviewers so they know when to work and on which surveys
2. Matching

### **11. GSBPM phases explicitly referred to in this module**

1. Data collection

### **12. Tools explicitly referred to in this module**

1. Blaise (CATI call management system)

### **13. Process steps explicitly referred to in this module**

1. Scheduling of CATI interviewers
2. Updating of CATI schedules
3. Allocation of telephone numbers to CATI interviewers

## Administrative section

### 14. Module code

Data Collection-T-CATI Allocation

### 15. Version history

Version	Date	Description of changes	Author	Institute
0.1	16-02-2012	first version	Leon Willenborg	CBS (Netherlands)
0.2	03-06-2012	remarks reviewers applied; references to other modules have been removed; technical terms have been removed; readability of text has been improved	Leon Willenborg	CBS (Netherlands)
0.3	12-12-2012	further remarks of reviewers applied; small corrections	Leon Willenborg Manuela Murgia	CBS (Netherlands) ISTAT (Italy)
0.4	19-11-2013	comments of EB review processed	Leon Willenborg	CBS (Netherlands)
0.4.1	19-11-2013	preliminary release		
1.0	26-03-2014	final version within the Memobust project		

### 16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:50