

## **Life Event History Calendar, CAPI and the C-TOR Study**

*This summary is based on chapter 15 “An Evaluation Study of the Event History Calendar” of the book “Using Calendar and Diary Methodologies in Life Events Research”, the paper “Training and monitoring interviews in administering CAPI Event History Calendar instruments” delivered to the 63th AAPOR conference in New Orleans, Louisiana in May 15-18, 2008 by Dijkstra W, Smit JH, Ongena YP and the paper “Monitoring interviewer performance by means of behaviour coding in a large scale cross national survey” delivered to the 62<sup>nd</sup> AAPOR conference in Anaheim, California in May 17-20, 2007 by Dijkstra W, Ongena YP, Smit JH.*

Event History Calendar (EHC) interviews have been developed to improve respondent's recall of retrospective information and are becoming a more and more popular replacement of conventional (standardised) interviews. The EHC is usually a matrix with column headings indicating calendar years and/or months and rows representing different life domains (i.e., residence, marriage, employment etc.). Life events entered in the calendar serve as cues, enhancing the retrieval of other events. By inspecting well remembered events from previous domains (e.g. house moves or marriages), the interviewer can help the respondent to retrieve other events, with cross-referencing probes like “Did you quit smoking before or after you moved to Haarlem?”

The main challenge in studies such as C-TOR is the ability of patients to recall past events in their lives. Imperfect memory is an important source of error that compromises data quality. For this reason a state of the art questionnaire was created in order to obtain information not only with regard to smoking history, but also for a large number of other potential risk factors. The study made use of a Life Event History Calendar (LEHC) to obtain these primary data.

This questionnaire is based on interviewing methodologies developed from cognitive psychology on the structure of memory. Throughout the questionnaire, memories are thematically and temporally structured and memory retrieval can be achieved through multiple memory processing pathways which relate events to one another. This technique has shown to be more effective and reliable than simply asking isolated questions which interrupt natural flow of thought. Its methodology has had an impact on population studies, psychology and sociology studies and its use has demonstrated high quality retrospective reporting. In specific, the LEHC has been used in various fields to facilitate recall of events:

- Occupational epidemiology research: Improvement of exposure characterisation of agricultural workers
- Reproductive epidemiology research: Collection of data on synthetic hormone use
- Nursing research: Improvement of adolescent sexual risk assessment
- Interpersonal research: Intimate partner violence reporting

The LEHC is an extremely useful tool for collecting retrospective data on life events with respect to different life domains like residence change, marriage, occupation, medical history and the like. The occurrence of simultaneous transitions in different domains is an important feature of the LEHC. Transitions in one domain serve as cues that help the respondent (1) to remember transitions in other domains and (2) to locate those transitions exactly in time. The EHC also helps the interview to identify potential inconsistencies.

The basic idea of the EHC is that it uses the way events are apparently organised in memory. Remembering a particular event stimulates the respondent to remember related events or situations, that is, events serve as *cues* for other events. Three different types of cueing can be distinguished: top-down, parallel, and sequential cueing (Belli, 1998).

*Top-down* cueing refers to relationships between the top and the bottom of a hierarchy. In the EHC this cueing is reflected by first asking about the more general events in a domain, for example a timeline with names of employers, and next a timeline with more specific information, for example one's position. *Sequential* cueing refers to the chronological order of events within the same domain; events are organised in memory on the basis of what happened earlier versus later in time. *Parallel* cueing refers to associations that exist across domains; this memory process reflects the fact that many aspects of life impinge upon individuals simultaneously or nearly so. For example, a change in employment or getting married may affect the residential situation, finishing an education may affect employment status, and so on.

### ***A Computer-Assisted EHC***

According to Dijkstra, Smit and Ongena, a Computer Assisted Personal Interview (CAPI) version of the EHC delivers a number of benefits. For example, such a computer programme might prevent the interviewer from entering an X if a number should be entered. It can also check for any un-allowed gaps in a timeline; for example, the timeline designed to collect residential history should have no gaps. Introductory scripts can be adjusted depending on whether the respondent is male or female, or has a spouse or not, and so on, thus preventing complex sentences that include multiple possibilities like "he or she" or "If the respondent has a spouse then ask...; if not, then ask..." Questions or timelines can be skipped if not applicable; for example, if the respondent does not have any children, the birth timeline can be skipped to prevent information from being incorrectly entered.

A CAPI version can also contain the introductory scripts, the response alternatives, suggested probes and cross-checks, and a help function where the interviewer can easily find information about how to handle difficult situations.

### ***Interviewer Skills in an EHC Interview***

1. Unlike usual standardised survey interviews, EHC interviews are more flexible and the interviewer has more freedom in administering the interview. The reason for this flexibility is that the interviewer has to apply EHC-specific techniques, to stimulate the respondent's memory in retrieving information from the past. The most important EHC-techniques are:

- Cross-checking

If a respondent reports an event, or a change in his situation in a particular month (e.g., a change in family income), the interviewer may verify this by reminding the respondent of another event that occurred at the same moment in time, reported earlier by the respondent (e.g., a marriage, another job).

- Cueing

If a respondent is unsure about when an event took place, the interviewer may use personal landmarks (other events already reported by the respondent) to identify the exact location in time (e.g. "Was that before or after you started working for company 'X'?").

- Probing for changes

Interviewers should always be aware that the situation of the respondent can change over time. Hence the interviewer has to probe for changes (like "Did your income change after that?").

In applying such techniques, the interviewer has a lot of freedom with respect to the formulation of probes and highly depends on the kind of information provided by the respondent. Nevertheless, some rules have to be applied. For example, “cueing” should always be balanced (like “was that before or after you...”) instead of unbalanced (like “was that after you...”) to prevent conformity bias. Especially if the respondent is unsure about a particular date, he may tend to agree with suggestive cueing like: “Was that before your birthday?” Asking “Was that before or after your birthday?” forces the respondent to make a choice between both alternatives, which in turn forces him to actively think about his birthday in relation to the date of the event he is asked about.

Similarly, in probing for changes the interviewer should preferably ask “Did that situation change or did it remain the same?” Many interviewers appear to ask however something like “And that remained the same?” Here too, if the respondent is not quite sure, simply agreeing with the suggestion is the easiest way. Providing the respondent with a choice will more likely force the respondent to actively try to remember if a change occurred.

Interviewers should be trained to apply such techniques frequently, at the right moments, and in a correct way (e.g., balanced).

2. In addition, they should be taught that despite they have more freedom than in a common survey, part of their behaviour, like question reading, should nevertheless be standardised. For example, in filling out a particular timeline (e.g., income), the interviewer has to read an introductory question (e.g., explaining what exactly is meant with “income”). Because of the more conversational character of administering an EHC, interviewers tend to deviate from scripted texts more easily than in a usual standardised survey with question lists. If the respondent has to select one of a number of response alternatives, also in an EHC the interviewer has to carefully read the alternatives, and prevent unwarranted suggestions and inferences.

Hence, in addition to EHC specific techniques, in a training attention has also to be paid to standardised interviewing techniques, especially the wording of questions and response alternatives.

3. Because of the more conversational character of an EHC, respondents tend to be more talkative than in a usual survey. Interviewers should know how to cope with irrelevant talk. They should be able to distinguish irrelevant information from information that, although not directly pertaining to the requested information, is part of the process of trying to remember relevant events.
4. In an EHC interview it is much more likely that respondents provide information pertaining to next timelines. Quite often, after a number of timelines have been filled out, the interviewer already knows (or suspects) how a next timeline will look like. Hence, the risk of suggestive behaviour, unwarranted inferences and departures from scripted question texts (to account for information already known) is much larger. Interviewers should be trained how to deal with such problems.
5. An EHC is not as straightforward in filling out as a standardised questionnaire, especially an electronic version may be complicated and it may take some time before interviewers are acquainted with filling out the EHC.

## **Background of C-TOR Study and LEHC**

The C-TOR study was designed as a large-scale epidemiological case-control study that ultimately recruited approximately 8,500 patients from hospitals and clinics in five different European countries (Germany, France, Italy, Greece and Slovenia). The main purpose of the study was to examine the risk of developing lung cancer, in particular whether there is a difference in risk related to smoking ultra-low tar (ULT) cigarettes when compared to smoking full-flavour cigarettes. Approximately half of the respondents were cancer patients, the other half comparable hospital patients with non-smoking related diseases. Respondents were recruited by medical doctors.

The structure of the CAPI-LEHC aims at establishing specific experiences in the context of the patient's life. The important events and transitions in the patient's life are used as cues to recall specific events or experiences. Because EHC technique heavily relies on techniques like cross-checking and cueing, quite a number of other aspects about the respondent's life were taken into account. For example, the respondent may very well remember the addresses where he has lived, and when, or the date of birth of children. Such information may help him or her to recall less salient events that are however important to the study. Therefore, the LEHC covered diverse topics like residences, marriages, children, education, jobs, smoking history, drinking history, diets through one's lifetime and medical history.

### ***Description/Structure of the CAPI-LEHC***

The questionnaire used in the study is composed of several sections:

- Introductory questions
- Standardised Mini-Mental State Examination (SMMSE) – cognitive assessment
- LEHC questions
- Socio-economic questions
- Short version of the questionnaire (to be used with patients who score less than 18 on the SMMSE)

Each section of the questionnaire is explained in detail below.

- *Introductory Questions*

The introductory questions were intended to capture administrative details about the patient and interviewer, interview location, date of interview, etc., such that each interview had a unique identifier. This unique identifier was used to merge the CAPI-LEHC data output with the patient's medical record data which was recorded using the case report form.

- *Standardised Mini-Mental State Examination (SMMSE) – Cognitive Assessment*

The accuracy of the information gathering process of the CAPI-LEHC relied heavily on the memory of the patient. Accordingly, it was beneficial, prior to beginning the interview, to assess the patient's cognitive capacity. Because of the nature of this study, it was recognised that many of the patients would be of advanced age and/or in poor health, and that the patient's answers to such cognitive questions would help the interviewer assess the cognitive capacity of the respondent. The C-TOR study employed the Standardised Mini-Mental State Examination (SMMSE), one of the most widely used screening tests of cognition in older adults. Failure by the patient to achieve a minimum threshold score suggested that the patient cognitive capacity was not adequate to impart a reasonable degree of confidence in the accuracy and reliability of the information the respondent was about to be asked to give. If the patient failed to achieve a minimum threshold score of 18, he or she was not asked

questions from the LEHC section of the questionnaire but was asked to answer more general questions about their educational and socioeconomic status and their smoking history.

- *LEHC Questions*

The main body of the questionnaire was focused on a series of LEHC questions, whereby answers to questions were stored in combination with a year or range of years. The questions were logically grouped into domains of questions relating to similar subjects concerning aspects of the respondent's life, such as: Residences, Life Events, Education, Occupation, Smoking History, Health and Medical History. These domains served as conceptual anchors and encouraged top - down processing.<sup>1</sup> The specific timelines in each of the domains were used to provide cues which would set the context in which the patient may have used cigarettes as well as to record data on possible effect modifiers and confounders. While CAPI-LEHC asked a number of personal questions from patients regarding partners or children, including their names, and past and current residential, school, college and work addresses this information was only used as cues to aid the patient in remembering information during the interview and was not saved in the study database. The questionnaire outlined which information was used for data passage and which information was solely used by the interviewer to cue the patient.

- Residence Domain

The Residence Domain consisted of three timelines: (1) the address timeline which recorded all of the places in which the respondent has lived during his or her entire lifetime; (2) the setting timeline where the patient characterised the setting of the residence (e.g., farm or a big city); and (3) a population timeline which asked for a self-characterisation of the population density for each of the places in which the patient lived.

- Life Events Domain

The Life Events Domain recorded major life events which have occurred in the life of the patient. This domain included timelines on: (1) the number of marriages the patient may have had; (2) marriage-related events, such as permanent separations, divorces, or widowhoods; (3) the number of partners with which the patient lived as if married; (4) the number of children which may have been born, adopted or taken care of by the patient; (5) important deaths which may have affected the patient; (6) the number of persons living in the same household as the patient; and (7) the number of smokers, not including the patient, who had been living with the patient.

- Education Domain

The purpose of the Education Domain was to collect information about patient's formal educational experiences. This domain had two components: a schooling timeline and a degrees timeline to record the patient's highest level of education. Appropriate schooling types (educational institutions) and degrees (educational levels) were automatically displayed based on the patient's country of residence. The customised response alternatives were standardised based on the International Standard Classification of Education (ISCED).

---

<sup>1</sup> Wiebe, E.F. and Landis, K.R. 2000. Evaluation of an electronic event history calendar. In Proceedings of the Survey Research Methods Section, American Statistical Association, Alexandria, VA, U.S.A. American Statistical Association

- Occupation Domain

The Occupation Domain aimed to capture information about the working life of the patient. It comprised several timelines: (1) an employer timeline which sought to gain information on the periods during which a patient worked for particular employer, including military service and self-employment; (2) an occupational exposure timeline to determine the type of industry or profession in which the patient may have worked against occupations known to be confounders or effect modifiers in this study; and (3) an “others smoking” in the workplace timeline to capture the frequency with which a patient may have been exposed to colleagues smoking in the workplace.

- Smoking History Domain

The Smoking History Domain sought to capture information regarding the patient’s smoking behaviour throughout their life. This domain, through several timelines and additional pop-up questions, determined:

(1) If the patient had ever smoked; (2) if the patient was a current smoker; (3) the periods during which the patient smoked or did not smoke; and (4) the brand names and quantities of cigarettes smoked.

This section of the CAPI-LEHC included a database of pictures of cigarette packs to help the patient identify the specific cigarette brand they smoked. Allowances were also made to collect information for cigarette brands which may not have been included in the database.

The tobacco domain also included timelines to collect information on other forms of tobacco products that the patients may have used, such as cigars, pipes, cigarillos, snuff, or chewing tobacco. These timelines also included additional pop-up windows to determine if the patient was a current pipe or cigar smoker.

- Health Domain

The Health Domain captured information on the general self-perceived healthiness of the patient’s: (1) diet; (2) fruit and vegetable consumption; (3) consumption of fatty foods; (4) the level of alcohol consumed; and (5) weight at a given period.

- Medical History Domain

The Medical History Domain featured timelines for certain illnesses and conditions that the patient may have been diagnosed as having at various times and which are of interest in the C- TOR study. The timelines included in this domain sought to capture information on: (1) chronic lung disease; (2) chronic vascular disease; (3) coronary or heart disease; (4) cancer; and (5) diabetes.

- *Socio-Economic Status Questions*

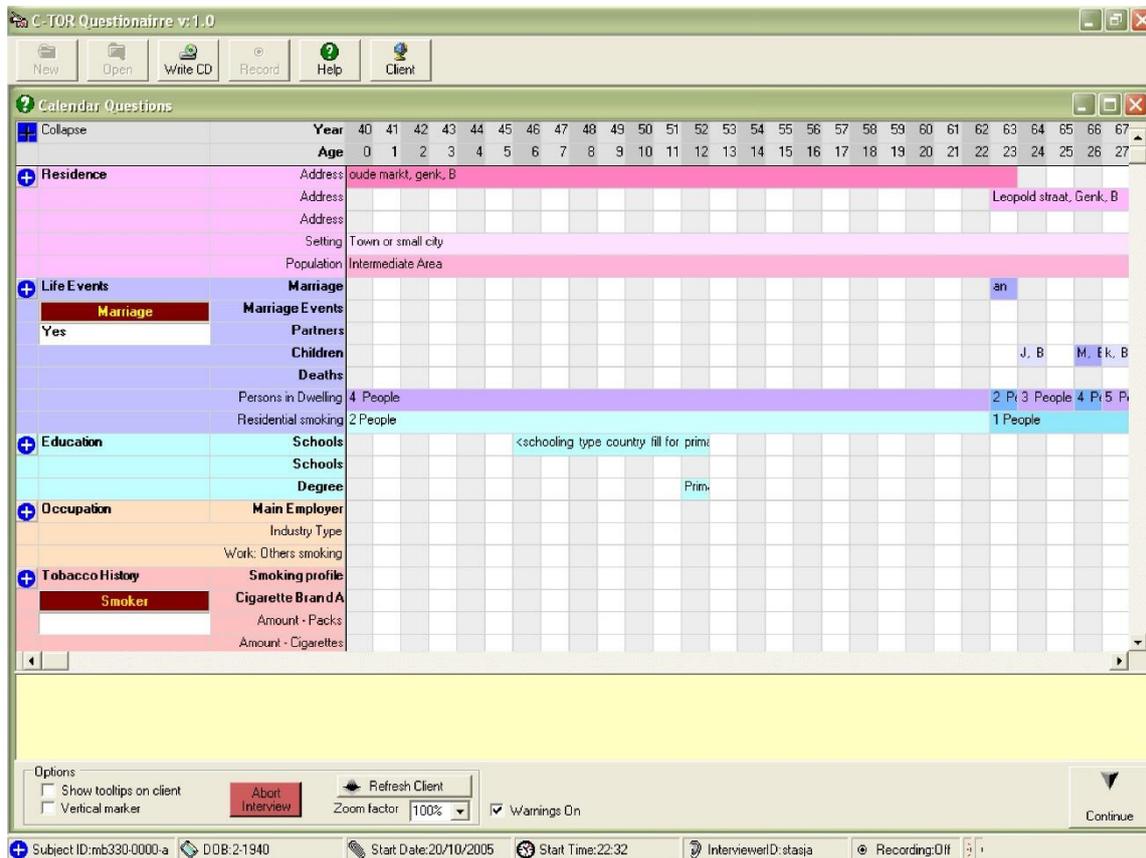
The CAPI-LEHC contained one final set of questions which collected general background information of the patient and members of the household. Questions in this section of the interview attempted to collect information on the patient’s income, self-perceived quality of life, and educational attainment of the patient’s and of the patient’s parents. The socio-economic section also included employment-related questions regarding other members of the household and their educational attainment. These questions were designed such that a socio-economic index for the patient could be calculated. This index was a general index applicable across all the study countries.

- *Short Version of the Questionnaire*

Patients who score less than 18 points in the SMMSE were not asked questions from the LEHC section of the questionnaire but were asked to answer more general questions about their educational, socio-economic, and marital status and their smoking history. The short questionnaire was designed to gather general background data on patients who fail the SMMSE and to minimise the potential for distress among patients who may not be suitable for the full interview and may become upset that the interview was discontinued because they “failed the test”. Cases scoring less than 18 on the SMMSE cognitive test were not matched with a control and if a matched control scored less than 18 on the SMMSE cognitive test another control for the documented case was sought.

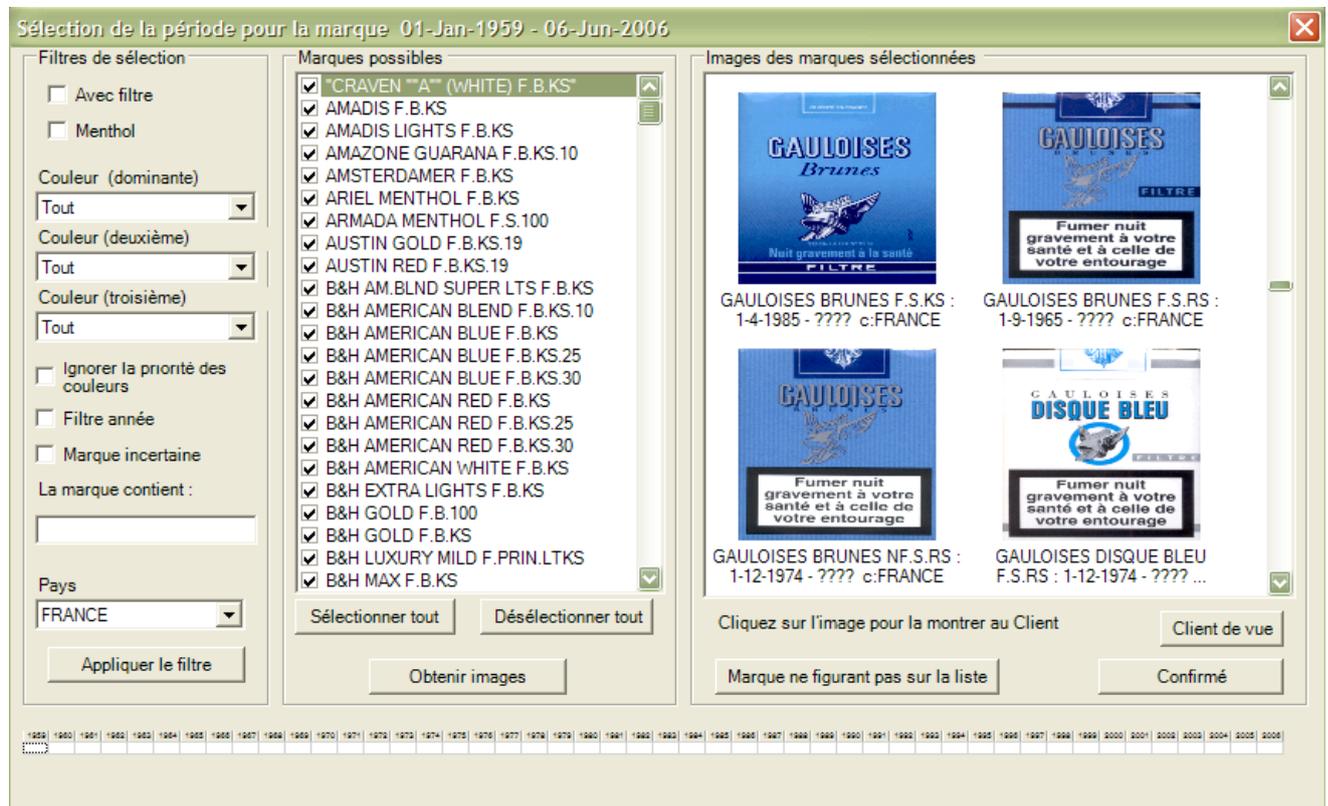
### *Life Event History Calendar*

An electronic version of the LEHC was developed to conduct the interviews and minimise recall bias. The software development and data management of the Computer-Assisted Personal Interview (CAPI)-LEHC was handled by 2 clinical research organisations (CROs) (Kendle and PRA International). Figure 1 shows a screenshot of the CAPI-LEHC for the C-TOR project as seen by the interviewer.



**Figure 1: Life Event History Calendar.** Screenshot of part of the electronic EHC. On the left side, there is a standardised set of major life cues which act as conceptual anchors: residences, marriages, marriage events (matrimonies, divorce, widowhood, etc.), births of children, important deaths, schools attended, degrees obtained, and occupation.

Two laptops were used to administer the interview, one for the interviewer, and one for the respondent. The respondent viewed one laptop on which could be seen the development of the EHC. Interviews lasted approximately two hours. There were in excess of 100 of these two-laptop sets used in the course of the study. Interviewees were also provided with visual display of cigarette packages from the cigarette pack image database on their screen in order to be facilitated in retrieving information regarding their past smoking habits.



**Figure 2: Visual Display of Cigarettes.** Cigarette package display on the interviewer's laptop.

Interviews were conducted by interviewers from CROs (Clinical Research Organisations) in the different countries at location (i.e. hospital or clinic). After an interview was completed, data were burnt on CD. All interviewers were trained and certified in a training programme developed by the Free University of Amsterdam (Vrije Universiteit Amsterdam, VUA) before performing the interviews. All interviews, subject to the permission of the interviewees were audio-recorded, using the interviewer laptop as a recording device.

#### *Validation Studies of the Questionnaire*

Two formal validation studies and dozens of pre-testing sessions were conducted by university-affiliated experts to evaluate and validate the LEHC and the CAPI. An initial validation study was conducted during August and September 2004 on a paper and pencil version of the C-TOR LEHC.<sup>2</sup> The validation study was conducted by six experienced interviewers who administered the questionnaire to 30 subjects aged 60-79. This validation study explored several issues including; interviewers' ability to grasp the LEHC concept, interviewer instruction, training and performance, attitude of patients towards the LEHC; as

<sup>2</sup> Dijkstra, W. and Smit, J.H. 2004. Validation of the EHC Questionnaire for the ULTORC Project. Report for THE WEINBERG GROUP, LLC.

well as design improvements to the LEHC. Based on the findings of the validation study extensive modifications were made to the LEHC questionnaire in terms of wording, layout and design, including the decision to computerise the LEHC. Modifications were also made to the instruction and training of interviewers in order to ensure optimal questionnaire administration and data gathering.

The CAPI version of the C-TOR LEHC was developed between September 2004 and March 2005, with many pre-testing meetings held to improve CAPI usability and LEHC readability. A validation study of the CAPI-LEHC was conducted in April 2005. For this study six interviewers administered the CAPI-LEHC to 30 subjects. Subjects were 65 years or older and had a previous or current history of smoking during part of their lifetime. This study focused on testing the reformulated version of the CAPI-LEHC; identifying and rectifying particular problems interviewers may have with the CAPI-LEHC; improving the documents concerning manuals and instructions for the interviewer; improving trainer skills for the main C-TOR Master Class training; improving the training procedures; and testing and improving the interviewer monitoring instrument. The CAPI-LEHC was reviewed and improved upon based on the results of the validation work, ensuring that the final version of the C-TOR CAPI-LEHC was ready to be implemented in the field.

One of the observations from the pilot studies conducted by the team from the Free University of Amsterdam on data quality was that interviewers tend to relax the requirements of standardisation, apparently because of the character of the LEHC, which allows quite a lot of freedom for the interviewer, for example by providing the respondent with cues already available from the calendar, like “Was that before or after the birth of your first child in 1969?”. Nevertheless, questions should be posed as worded in the LEHC, that is, as they appeared on the interviewer’s screen. However, in many cases, the interviewer completely rephrased the questions. In addition, it was of importance that interviewers correctly applied EHC techniques (cueing and cross-checking) to a sufficient degree. For these reasons a special three-day training was developed, including exercises to become acquainted with the CAPI (Computer Assisted Personal Interview) programme, role-plays, demonstration videos and interviews with a real respondent. In addition it was decided to monitor the interviewer’s performance during the interview.

#### *Data Quality*

Interviewer performance was monitored throughout the study in order to ensure that the final interviews were conducted to the high quality standards set for the study. Interviews were selected for behaviour coding according to the requirements of the C-TOR Protocol. For each interviewer the first, second, fifth and tenth audio recorded interviews were selected for behaviour coding, followed by a random selection of 10% of interviews for each interviewer. The procedure for random selection ensured that interviews that were not audio recorded were not selected. If the first, second, fifth or tenth interview was not audio recorded, this interview was replaced by another interview.

#### *Coders*

Behaviour coders (the individuals assigned to perform the behaviour coding) were selected by the CROs based on the criteria that they have to be a native speaker for the interviews to be coded and that they have an adequate level of proficiency in English. Coders were assigned only after training and certification by the Free University of Amsterdam. Additionally, coders were only assigned to interviews from interviewers affiliated with a different CRO from the coder.

Coders first had to read the interviewer training manual, because this manual explained in detail the EHC, EHC techniques and requirements with respect to the behaviour of the interviewer. In addition, they studied the coder manual, explaining both programmes for behaviour coding and coding rules. Next they had to code a test interview, designed for this purpose. After obtaining feedback they had to do a number of exercises, essentially consisting of small interview transcripts with behaviour that was more difficult to code. After this exercise they had to code a second test interview and do a second exercise. If their performance was acceptable, they were authorised to start coding.

### *Tools*

Two computer programmes were developed to assist the coders in behavioural coding. One programme was called “CAI player”. After insertion of the audio CD, this programme decrypted the audio and log files, and allowed the coder to play the audio file. The second programme, the Behaviour Coding Programme Proper, made a selection of those parts of the LEHC that were subjected to Behavioural Coding. To check the wording of questions, the programme included the different question texts, in each of the different languages. After an interview had been behaviour coded, the programme produced a file, containing all codes assigned by the coder. This file was electronically sent to the Free University and added to a specially designed database programme. This programme in turn generated a report of the interviewer’s performance, which was next used for feedback.

### *Translation of the Questionnaires*

The C-TOR translation approach used a multi-stage team-based process consisting of Translation, Review, Adjudication, Pre-testing and Documentation (TRAPD). The TRAPD procedures were incorporated in a team approach, reflecting cutting-edge developments in survey questionnaire translation and assessment practice.

The C-TOR translators were skilled practitioners who had received training on translating questionnaires. The C-TOR Study called for two translators and a coordinator per questionnaire. Translators translated out of English into their strongest language. The translators shared the translations with each other. At a reconciliation meeting, translators and the translation coordinator went through the questionnaire question by question discussing versions and agreeing on a version. At the end of the one-day meeting the coordinator-adjudicator in cooperation with the two translators decided on which translation options to adopt. Suggestions were also made for countries sharing languages to allow for harmonisation where appropriate and a documentation component was included to facilitate assessment and inform later versions. The final version of each language in each case was the product of the decisions taken on the draft translations by a group that brought together survey knowledge for the relevant population, translation and source language skills and an intimate understanding of the C-TOR survey instrument.

### *Data Privacy*

Regarding data privacy, it is significant to mention that personal data stemming from a number of personal questions asked to the interviewee were only used as cues to place events in context. They were not saved in the study database and the appropriate electronic data handling practices and requirements for maintenance of patient privacy have been applied as governed by international codes and as required by national and local regulatory requirements.