Decentralised CATI Versus Paper and Pencil Interviewing: Effects on the Results in the Swedish Labour Force Surveys

Lars R. Bergman, Karl-Erik Kristiansson, Anita Olofsson, and Margaretha Säfström[1]

Abstract: The results of decentralised computer assisted telephone interviewing (DCATI) and decentralised paper and pencil interviewing (DPAPI) were compared in a large methods study within the Swedish Labour Force Surveys. In the main study ($n = 103,161$ and $23,420$ for DPAPI and DCATI, respectively) key estimates were compared for the two methods. Very similar results were obtained except for (a) the category weak attachment to the labour market for which a 2% difference was found, believed to have been caused by a different layout of the response alternatives in DCATI and DPAPI, and (b) the percentage working 40 hours for which also a 2% method difference was found. Results from a reinterview study ($n = 1,800$ for both DPAPI and DCATI) indicated that the two methods had produced data of very similar quality. In a coding study ($n = 4,127$), the quality of the information provided by the interviewers on occupation and industry in DCATI was compared with previously obtained information using DPAPI. Both methods seemed to have produced almost equivalent information.

Key words: CATI; CAPI; computer assisted interviewing; telephone interviewing; data quality; measurement error; labour force survey.

1. Introduction

This paper examines the quality of labour force survey data from two different modes of data collection: (1) decentralised computer assisted telephone interviewing (DCATI) and (2) decentralised paper and pencil telephone interviewing (DPAPI). There is a paucity of such studies in the literature as interest has focused primarily on centralised computer assisted telephone

interviewing (CCATI) or computer assisted personal interviewing (CAPI). It is therefore natural to begin with a brief outline of CCATI and CAPI.

Emerging in the 1970s, CCATI is now considered the standard method of conducting telephone interviews. Compared to centralised paper and pencil telephone interviewing (CPAPI), it has been claimed that CCATI offers numerous advantages, e.g., a shorter post-interview processing and lower costs. The second claim has not received unequivocal support and may not be generally valid (Nicholls and Groves 1986). What is clear is that, normally, the CCATI/CPAPI cost ratio decreases with

an increased sample size (for further discussion, see Weeks 1992).

Another claim is that CCATI improves data quality because the computer (a) guides interviewers through correct skip patterns, (b) makes monitoring easier, (c) conducts consistency checks, (d) allows for customised wording, and (e) in general, leads to greater standardisation of interviewer behaviour (Fink 1983; Groves 1983). Nevertheless, hard evidence has been slow to emerge and Groves and Nicholls (1986, p. 117) conclude that "... with few exceptions, there is little reliable empirical evidence that CATI affects data quality." More recent findings profess data quality improvements with CCATI (Catlin and Ingram 1988; Hubble and Wilder 1988) and in a recent review article Weeks (1992, p. 456) concludes that "... the CATI experiments reported in the literature suggest that CATI can improve the level of data quality over a comparable telephone PAPI survey."

For a variety of reasons, CAPI has been slower to emerge but many of the early concerns about CAPI have now been put to rest (for instance, worries about technical feasibility and respondent and interviewer acceptance (Baker and Bradburn 1991)). The evolution of powerful and light weight portable PCs and improved software also makes this method more attractive; a good example is the introduction of the Blaise system in CAPI at Statistics Netherlands (Hofman and Keller 1993).

In a study of the quality of labour force survey data, Catlin and Ingram (1988) compared CCATI and CPAPI in a study of 10,000 Canadian households. They found a higher proportion of nonresponse in CCATI than in CPAPI (about 2% higher rate of "no contact/absent"). Their explanation was that the CCATI interviews took about 20% longer, which gave the interviewers less time for contact attempts than the CPAPI interviewers had. On the other hand, in their study of the labour force survey processing system, which provided a discrepancy or edit failure rate for each item, they found less than half the number of errors in CCATI compared to CPAPI.

At the U.S. Bureau of the Census, Shoemaker, Bushery, and Cahoon (1989) compared a test group (56% CCATI, the rest DPAPI or personal interviews) and a control group (DPAPI or personal interviews) in a large sample in the Current Population Survey (CPS) development program. The unit nonresponse rates were about the same for the two groups but some small differences in item nonresponse rates were found. When key labour force estimates were compared, the results, with some minor exceptions, were found to be similar. An exception was the CLF rate which was 1.1% higher for the test group than for the control group. The authors concluded that, generally speaking, the CCATI method produced results which were comparable to those from the regular CPS. Later CCATI results from this program based on independent reinterviews indicate that the response variance estimates were significantly lower in CCATI for the categories employed, unemployed and not in the labour force than in the paper and pencil data collection (Waite and Rogers 1991).

In a preliminary report, Baker and Bradburn (1991) compared the quality of data and the survey costs of CAPI and of conventional personal interviews in an embedded methods study in the National Longitudinal Survey of Labour Market Experience/Youth Cohort. An experimental design was used with interviewers randomly selected to either the CAPI group ($n = 2,814$) or the control group

($n = 2,715$). The authors interpreted the results as very positive for CAPI. Missing data were reduced due to elimination of erroneous skips, respondents were presumably somewhat more candid in their answers to some questions about birth control practice, and on the average the interview was about 10 minutes shorter in CAPI than in the control condition. However, with regard to a series of questions aiming at establishing wage rates for employed persons, the results indicated a difference in the way the interviewers presented the questions to the respondents. It appeared that in CAPI the questions concerning wage rates were viewed and asked separately from each other while in the control condition the interviewer could see all the questions on the same page of the questionnaire.

2. The Swedish Labour Force Survey (SLFS)

2.1. The SLFS sample and survey design

The SLFS population consists of all persons in Sweden who are in the Register of the Total Population (RTP), ages 16 through 64. In the version of the RTP that is used as a sampling frame, individuals are listed according to personal identity numbers (PIN), which is equivalent to a listing by age.

The SLFS is based on three separate rotating samples, one for each month in the quarter. In a given month, the total sample of about 18,000 persons consists of eight rotation groups, where the persons in rotation group 1 are interviewed for the first time, the persons in rotation group 2 are interviewed for the second time, etc., up to rotation group 8 which consists of persons interviewed for the eighth and last time. The rotation groups are successively

replaced, one at each measurement period. For each of the rotating samples this takes place every third month. Thus, a selected person participates every third month and is, after two years, rotated out of the survey. To simplify sub-sampling etc., a random group number system has been constructed by which one of 40 random numbers is assigned to each person in each rotation group.

Telephone interviews are conducted by Statistics Sweden's decentralised interviewer organisation. For about 0.6% of the sample the data have to be collected in personal interviews. The first time a person is interviewed (i.e., normally when he/she is part of rotation group 1), there is a more extensive **first interview** than in rotation groups 2–8 when no change in job status has occurred. This shorter interview is called a **continuation interview**.

The population is stratified by county, sex, nationality (Swedish, non-Swedish), and employment situation according to an employment register. In this way, a total of 192 strata are obtained. A systematic sample is drawn within each stratum, and the sample size is proportional to the size of the stratum. The estimation of population totals from the sample values is based on the sampled persons' sampling probabilities and on monthly figures from the RTP.

For further information about SLFS, the reader is referred to Statistics Sweden (1990).

2.2. DCATI in the SLFS

Despite the fact that DCATI has not been given much attention in the literature, it is a natural data collection method for a decentralised survey organisation. Many advantages that CCATI has over CPAPI also transfer to a decentralised environ-

ment. In addition, the cumbersome conventional mail-based distribution and collection procedures transfer well to electronic media and the same computer hardware and software can be used for both DCATI and CAPI.

Statistics Sweden has developed methods, computer programs, and procedures for decentralised computer assisted interviewing for both telephone (DCATI) and personal interviews (CAPI), see Blom (1985). The first implementation of DCATI occurred in the SLFS.

Since the numbers produced by the SLFS (e.g., level of unemployment) are subjected to a high degree of scrutiny by users, the switch from DPAPI to DCATI should not in any way affect the main estimates. Or if it did, we must be able to show that the new method produced better estimates. This requirement led to extensive methodological work to establish whether the change from DPAPI to DCATI could be achieved without jeopardising the quality of the results. Different data quality indicators were compared between methods (completion rates, coding levels of occupation and industry, estimates of labour force variables, estimates of bias, and errors of measurement in individual observations). The results will be reported in Section 4.

The present version of DCATI uses a central mini-computer and each interviewer enters survey results on a portable PC. At the time this study was conducted, a Toshiba 1200 was used with a 20 Mb hard disk, 1 Mb RAM and a background illuminated LCD screen. Software which administers the interviewer's scheduling and assignment for the interviewing period are stored on the PC's hard disk. Responses from earlier rounds are also on the interviewer's PC. The interviewer program, locally installed in each PC, includes

built-in checks and consistency controls. There is also a communication program for receiving and sending information via modem to the central computer automatically during the night. The central computer receives the sample from the mainframe, communicates with the PCs, monitors the work flow and returns the preliminary result file to the mainframe. A description of the system is found in Statistics Sweden (1991). The system can, of course, also be used in CAPI applications, which are currently being tested.

3. Design of the Methods Study

3.1. Sample and time period

Beginning in August 1989, part of the SLFS sample was interviewed according to the DCATI procedure. The methods study had three main parts:

a. The **main study** included all SLFS data that were collected from August 1989 through February 1990. The DCATI sample numbered 23,420 and the DPAPI sample 103,161; for information about completion rates etc., see Table 1.

b. The **reinterview study** covered the same time period (except November and December when no reinterviews were performed) with sample sizes of 1,800 in both DCATI and DPAPI.

c. The **coding study** used the DCATI data for September and October for all persons who were employed and who had participated in the June or July SLFS. From this earlier time period the corresponding DPAPI data were used. The sample included data from two points in time for 4,127 persons.

The DCATI sample in the main study

was a probability sample from the Swedish population based on the above mentioned random group number system and drawn in the same way as the regular SLFS sample but with one difference: In the September to February rounds the DCATI sample in rotation group 8 was overrepresented by a factor of 2.58 and in rotation groups 1–7 underrepresented by a factor of 0.77. In the August sample, rotation group 8 was overrepresented by a factor of 4.71 and in rotation groups 1–7 underrepresented by a factor of 0.47. Since the total SLFS sample was divided between DCATI and DPAPI this implies that the DPAPI samples were overrepresented where the DCATI samples were underrepresented and vice versa (though to a less extent since the DPAPI sample was larger). Rotation group 8 was overrepresented in DCATI to obtain sufficient numbers of unemployed in the DCATI reinterview sample where the extensive reinterviewing procedures would not affect subsequent interviewing rounds.

In the reinterview study for both DCATI and DPAPI, all unemployed in rotation group 8 (according to the original interview) were selected for reinterviewing, while 25% of the not unemployed were randomly selected. In rotation groups 4–7, 6% of the DCATI sample and 0.9% of the DPAPI sample were randomly selected for reinterviewing. Due to logistic problems (interviewer illness, etc.) the reinterviewing had to be cancelled in November and December which disrupted the schedule of successive sampling from all the rotation groups. Rotation groups 1–3 were therefore not included in the reinterview sample.

The varying sampling probabilities were taken into account when constructing the estimators and when computing their variances.

3.2. Some details about interviewers and interview procedures

Originally 35 interviewers were selected and trained in DCATI. The selection was done in a two-step procedure: First a number of regions were selected that would be representative of all of Sweden and then either all interviewers or a random selection of interviewers in each region were selected (depending on the size of the region). However, a year later when the methods study was to begin, 10 interviewers had either quit or were on leave. This attrition rate of 29% should be compared to the corresponding figure for the whole interviewer staff which was about 22%. An additional 18 interviewers were then selected (using administrative criteria) and trained making a total of 43 DCATI interviewers. Of these, 39 participated during the entire seven months of the methods study. When the age and the number of employment years of the DCATI interviewers were compared with those of the entire interviewer force, some differences were found. The main one was that 45% of the entire work force were over 54 years old as compared to 33% of the DCATI interviewers. Clearly the DCATI interviewers were not fully representative of the total interviewer staff.

A first interview rather than a continuation interview was performed the first time a person was interviewed by the DCATI method, irrespective of the rotation group the person belonged to. This was done to avoid that the DCATI interview results should be influenced by the results of the first, more extensive, interview which was in DPAPI.

A specific procedure termed **delayed, indirect reconciliation** was employed in the reinterview study to try to measure individual errors and bias as carefully as

possible. The procedure is characterised in the following way:

1. The reinterviews were carried out centrally by telephone by 15 specially selected, trained and supervised interviewers using CPAPI, instructed to do the very best interviews possible. The reinterview was independent of the original interview in the sense that the reinterviewer did not know the results of the original interview. However, it was not a replication, since the reinterview involved the use of elite interviewers, working under unusually favourable conditions.

2. While conducting the reinterview, the interviewer noted the respondent's labour force category and attachment to the labour market on a special sealed form. Immediately after the SLFS questions, the seal was broken and the interviewer could see, adjacent to the answers from the reinterview, the corresponding information from the original interview. If there was no discrepancy, he/she would continue with some additional questions, mainly to make sure that the reference week was right. If there was a discrepancy, both close-ended and unstructured questions were used to determine in detail the respondent's working conditions during the reference week. The interviewer was to probe conditions where a discrepancy had been found. A summary statement of the work situation was also checked with the respondent.

3. All information was evaluated by an SLFS expert who decided whether the original answer or the reinterview answer would be counted as "true." In a few cases the respondent had to be recontacted. In the presentation of

the results of the reinterviews, these corrected and evaluated answers are considered as true. Due to the design of the reinterview study, only the information about labour force category and attachment to the labour force can be assumed to indicate true values.

The reinterviews were carried out, on the average, two weeks after the original interviews. Due to the faster communication of interview results in DCATI as compared to DPAPI, the DCATI reinterviewing had to be delayed to avoid a difference in time intervals between the original interview and the reinterview.

4. Results

4.1. Completion rates

In Table 1 completion rates and specified noncompletion rates are given.

In the main study DCATI has a slightly lower completion rate than DPAPI (87.9% as compared to 88.5%). The percentage not contacted is lower in DCATI than in DPAPI as well, and the percentage refusals is higher. However, in the DCATI sample, rotation group 8 is overrepresented, and it has been shown that there is a rotation group bias in SLFS in the direction of the reported DCATI-DPAPI differences (Kristiansson 1984). If the mode comparison is restricted to rotation group 8, the differences almost disappear. (They are 0.1% and 0.3% for not contacted and refusals, respectively.) In the above mentioned study by Kristiansson (1984) rotation group bias was also studied for labour force status but no appreciable effects were found.

4.2. Errors in the coding of occupation and industry

When introducing DCATI it was noted that

Table 1. Sample sizes and participation (%) in the SLFS method study of DCATI

	Main study		Reinterview study	
	DCATI	DPAPI	DCATI	DPAPI
Completion rate	87.9*	88.5	93.2	91.5
Not contacted	5.5*	6.2	4.1*	5.7
Refusals	5.7*	4.7	1.3	1.2
Other noninterviews	0.8*	0.6	1.3	1.7
	100.0	100.0	100.0	100.0
Sample size	23,420	103,161	1,800	1,800

Note. * stands for $p < 0.05$ for a two-tailed test of the difference between the DCATI and DPAPI rate. Due to rounding errors the percentages do not add up to exactly 100.0%.

many interviewers did not type well. The question then arose whether this might affect the information on which the coding of occupation and industry was to be based. This has also been a concern of others (e.g., Groves and Nicholls 1986). To throw light on this issue, a study was made of all respondents for whom data were collected in September or October using DCATI and who also had participated in the SLFS in June or July (at these two earlier dates all information was collected using DPAPI). The code obtained using DCATI was compared with the

Table 2. Error classification of occupation codes and industry codes given to persons for which data were collected using both DCATI and DPAPI. Percentages of the total sample

Error classification	Occupation	Industry
Reason for different codes		
A. Incomplete information in DCATI which resulted in		
erroneous classification	1.2	0.8
omitted classification	0.6*	0.4*
B. Incomplete information in DPAPI which resulted in		
erroneous classification	1.2	0.7
omitted classification	0.1	0.1
C. A true change in occupation/industry occurred between the DPAPI and DCATI data collection	7.5	5.4
D. Identical information was given in DCATI and DPAPI but different codes were assigned (coding error)	8.9	5.5
The DCATI and DPAPI codes were identical	80.5	87.1
All	100.0	100.0
Sample size	4,127	4,127

Note. * stands for $p < 0.05$ in a two-tailed test of the difference between the DCATI and DPAPI rate.

Table 3.   Overview of the variable sets for which differences between DCATI and DPAPI estimates were studied

Variable set	Number of variables studied	Differences found
Labour force category	5	No sign. diff.
Attachment to the labour market	3	3 sign. diff.
Number of hours of work	10	4 sign. diff.
Occupation	9	No sign. diff.
Industry	8	1 sign. diff.

Note. Within a variable set, each response category is counted as a variable. A difference is reported as significant if $p < 0.05$ using a two-tailed z-test.

code obtained for the same person three months earlier using DPAPI. In case of a discrepancy an expert coder decided the correct code. The results are given in Table 2.

Table 2 shows very similar results for DCATI and DPAPI. There is a significant tendency for a higher percentage of omitted codes in DCATI than in DPAPI, but the difference is very small (0.5% and 0.3% for occupation and industry, respectively). Presumably, the difference was caused by the interviewers' providing less comprehensive information in DCATI than in DPAPI and presumably this problem will decrease as the interviewers become more experienced in typing.

4.3.   Estimates

Mode comparisons of estimates were

originally made for the variables in the sets reported in Table 3.

For lack of space, detailed results are presented only for labour force category (Table 4) and for the variable sets where a pronounced method difference was found (attachment to the labour market, Table 5, and number of hours of work, Table 6).

The estimated percentages of persons belonging to different labour force categories are given in Table 4.

The differences between the methods are small and not significant.

Table 5 gives the estimated percentages of persons characterised by different degrees of attachment to the labour market. Here we can see a small but significant difference between methods, most apparent in the category weak attachment. The percentage is 13.9% for DCATI as compared to

Table 4.   Comparison of estimates of percentages belonging to different labour force categories

	DCATI	DPAPI	Diff.	sd of diff.
Labour force category				
Employed, at work	70.9	71.0	−0.1	0.4
Employed, not at work	12.5	12.3	0.2	0.3
Unemployed	1.2	1.2	0.0	0.1
Not in the labour force	12.9	13.1	−0.2	0.3
Unable to work	2.5	2.3	0.2	0.1
	100.0	100.0		

Note. Due to rounding errors the percentages do not add up to exactly 100.0%.

Table 5. Comparison of estimated percentages on different degrees of attachment to the labour market

	DCATI	DPAPI	Diff.	sd of diff.
Attachment of the labour market				
Strong attachment	76.8	75.8	1.0*	0.4
Weak attachment	13.9	15.8	−1.9*	0.3
No attachment	9.3	8.4	0.9*	0.3
	100.0	100.0		

Note. * stands for $p < 0.05$ in a two-tailed z-test of the difference between the DCATI and DPAPI rate. Due to rounding errors the percentages do not add up to exactly 100.0%.

15.8% for DPAPI. In trying to understand this difference, the exact questions and questionnaire layouts of the two modes were compared. The questions were identical but it was found that there was a difference in the layout of the response alternatives. In the DPAPI questionnaire the response categories were clearly grouped with those indicating strong attachment in the first column and those indicating weak attachment in the second and third columns. In the DCATI mode, the response categories were exactly the same but they were not grouped; they were placed in one column beginning with response categories indicating strong attachment. This lack of grouping may have decreased the interviewers' overview of the different response categories. The reinterview results (Table 8) give some support to the conclusion that the DPAPI estimate is the more correct one. For this reason, in March 1990 the DCATI questionnaire was revised to correspond to the DPAPI layout. In March through September 1990, the significant difference

Table 6. Number of hours of work during the reference week and during a normal week for employees at work in DCATI and DPAPI

	DCATI	DPAPI	diff.	sd of diff.
Number of hours during the reference week				
1–19	7.7	7.3	+0.4	0.2
20–34	21.6	22.1	−0.5	0.5
35–39	9.1	8.8	+0.3	0.3
40	36.6	38.5	−1.9*	0.6
41 or more	24.9	23.4	+1.5*	0.4
	100.0	100.0		
Number of hours during a normal week				
1–19	3.9	3.7	+0.2	0.2
20–34	20.5	20.8	−0.3	0.5
35–39	12.4	12.0	+0.4	0.3
40	50.8	52.5	−1.7*	0.6
41 or more	12.3	11.1	+1.2*	0.4

Note. * stands for $p < 0.05$ in a two tailed test of the DCATI and DPAPI rate. Due to rounding errors the percentages do not add up to exactly 100.0%.

for the weak attachment category disappeared. Percentage estimates were 14.1% and 14.4% for DCATI and DPAPI, respectively, the difference having a standard deviation of 0.8.

In Table 6 the estimated number of hours of work for the gainfully employed during the reference week and during a normal week are compared for the two modes.

Table 6 shows that the estimated percentage working "40 hours" (which in many jobs equals full time) is lower in DCATI than in DPAPI and that the reverse is true for the estimated percentage working "more than 40 hours." One tentative explanation, which is supported by field observations, is that in the DPAPI procedure the interviewers have to calculate the difference between the number of hours worked during the reference week and during a normal week. In addition, some questions are not asked if the number of hours worked is the same for the reference week and a normal week. This may cause some interviewers to round off the answer, especially to 40 hours a week. In DCATI, the calculations

are made by the computer. Thus it seems possible that the results concerning the number of hours of work are of a slightly higher quality in DCATI than in DPAPI. Unfortunately, reinterview results rendering true values were not collected for these variables.

4.4. Reinterview results

In this section, the resuls of the reinterview study are compared for the two modes. In the tables the figures have been weighted to take into account that the reinterview sample was stratified (see Section 3.1). As described previously, a reinterview method using delayed, indirect reconciliation was employed. This method is believed to produce reliable and valid data coming fairly close to true values, and it also permits a study of bias and individual errors.

Ten percent of the reinterviews identified a discrepancy either in labour force status or in attachment to the labour market among the combined sample interviewed by DCATI or DPAPI. After reconciliation it

Table 7. Labour force category according to the original interview and according to the reconciled reinterview. Percentages of cases in DCATI and DPAPI

Original interview	Reconciled reinterview			
	Employed	Unemployed	Not in labour force	All
DCATI				
Employed	82.8	0.2	0.4	83.4
Unemployed	0.1	1.0	0.1	1.2
Not in labour force	1.3	0.1	14.0	15.4
All	84.2	1.4	14.4	100.0
DPAPI				
Employed	82.7	0.0	0.7	83.4
Unemployed	0.1	1.0	0.2	1.2
Not in labour force	0.8	0.3	14.4	15.4
All	83.6	1.2	15.2	100.0

Note. Due to rounding errors the percentages do not always add up to the corresponding totals. The figures have been weighted to take the stratification into account.

Table 8. Degree of attachment to the labour market according to the original interview and according to the reconciled reinterview. Percentages of cases in DCATI and DPAPI

Original interview	Reconciled reinterview			
	Strong attachment	Weak attachment	No attachment	All
DCATI				
Strong attachment	74.9	1.8	0.2	76.8
Weak attachment	1.0	12.2	0.2	13.4
No attachment	0.6	0.9	8.3	9.8
All	76.5	14.8	8.7	100.0
DPAPI				
Strong attachment	75.7	1.0	0.0	76.6
Weak attachment	0.7	14.0	1.0	15.6
No attachment	0.3	0.4	6.9	7.7
All	76.7	15.4	7.9	100.0

Note. Due to rounding errors the percentages do not always add up to the corresponding totals. The figures have been weighted to take the stratification into account.

was judged that the reinterview had provided the correct information in 79% of the discrepant cases.

For the two modes separately, Table 7 gives estimated percentages of persons characterised by a certain combination of labour force category codes according to the original interview and according to the reconciled reinterview.

The results indicate that the two methods give similar results. The discrepancies between the original results and the reconciled reinterview results (considered approximately true) are small and similar for both modes. The total percentage of respondents with an incorrect labour force category code is estimated as the sum of the off-diagonal percentages. This figure is 2.2% in DCATI and 2.1% in DPAPI. For the three labour force categories employed, unemployed, and not in the labour force the following gross difference rates (GDR) were obtained: 2.0, 0.5, and 1.9 (DCATI) and 1.6, 0.6, and 2.0 (DPAPI). These GDRs are similar for both methods (GDR

is expressed in per cent and is computed as the ratio of the number of cases classified differently in the original interview and the reconciled reinterview to the total number of cases).

Let us now turn to bias. Assuming that the reconciled reinterviews yield approximately true values, the row and column totals in Table 7 can be compared for each of the two modes. For both modes, the discrepancies between the estimates obtained and the corresponding true values are small. However, the estimated bias is somewhat larger for DCATI than for DPAPI for the categories not in the labour force and employed. (1.0% and −0.8% for DCATI and 0.2% and −0.2% for DPAPI.) This indication of a methodological difference is not statistically significant, however.

For the two modes separately, Table 8 gives estimated percentages of persons characterised by attachment to the labour market according to both the original and the reconciled reinterviews.

It can be seen that the modes give fairly similar results. Assuming that the reconciled reinterviews yielded approximately true values, the number of erroneously classified persons is estimated to be 4.7% and 3.4% for DCATI and DPAPI, respectively. There is a significant tendency for the DPAPI method to have smaller classification errors ($p < 0.10$ using a two tailed z-test). The following GRDs were obtained for the three attachment categories: 3.6, 3.9, and 1.9 (DCATI) and 2.0, 3.1, and 1.7 (DPAPI).

The largest difference in bias found in Table 8 concerns the category weak attachment for which there is a significant estimated negative bias of 1.4% for DCATI as compared to a positive bias of 0.2% for DPAPI. This difference is consistent with the findings of the main study in Table 5 and it gives some support to the conclusion drawn in Section 4.3 that DPAPI presumably yielded more valid data on this variable. However, the mode difference found in Table 5 for the category strong attachment is not supported by the reinterview results.

For DCATI a small but significant positive bias was found for the category no attachment in the reinterview study (Table 8) and a corresponding difference in the estimates was found in the main study (Table 5). Going through the reinterview results case by case several small sources of error were identified that could have contributed to this difference but no single major cause was found.

5. Discussion

The main finding of the SLFS methods study is that the differences between the DCATI and DPAPI modes were small or negligible with respect to completion rates, codings of occupation and industry, and

usually small with respect to important estimates, classification errors, and bias. On the basis of these results, Statistics Sweden has implemented DCATI as the main data collection method in SLFS.

The largest methodological differences we found concerned the percentage working 40 hours a week (where we believe that the DCATI estimate is the more accurate one) and the percentage characterised by weak attachment to the labour market. The estimate of weak attachment is 1.9% lower by DCATI than by DPAPI, and in the reinterview results, a corresponding negative bias is found for DCATI. As discussed in Section 4.3, it presumably was caused by a difference in how response alternatives were presented in the questionnaires. The layout used in the DCATI mode was inferior to the DPAPI layout. It thus appears that exactly the same question and the same response alternatives yielded different answers depending on layout details.

When analysing reinterview data in our study, almost no differences were found between the GDRs for DCATI and DPAPI with regard to labour force categories, but in the CPS study (Waite and Rogers 1991) the GRDs tended to be lower for a CCATI mode than for a paper and pencil mode. Two possible reasons for this are:

a. In the CPS reinterview study independent replication was attempted and there were also systematic differences between the reinterviews performed on those originally interviewed by CCATI and those originally interviewed by paper and pencil methods.

b. In the SLFS reinterview study both those originally interviewed by DCATI and those originally interviewed by DPAPI were reinterviewed

by CPAPI. This might possibly bias our results in favour of DPAPI.

After reviewing all the evidence from our methods study (including field observations and reconciliation protocols aimed at explaining observed discrepancies), we conclude that the two methods appear to have produced data of very similar quality, but that there are possibilities for quality improvement within the DCATI procedure. As the procedure is presently designed, it does not fully utilise the potential the computer offers; in some respects it is shackled by conventional paper and pencil thinking. Two promising areas for improvement are more extensive and sophisticated feedback on unreasonable or insufficient answers during the interview and a closer central monitoring of the data collection.

The lack of evidence of improved data quality when switching from DPAPI to DCATI should not be generalised to other settings. For one thing, the DPAPI procedure used in SLFS must be considered an expensive, high quality procedure. It included, for instance, follow-up contacts with all persons whose data were incomplete.

Reinterview results from the American CPS were slightly more positive for CCATI than for DPAPI/personal interviews; and Hubble and Wilder (1988) have interpreted results for crime variables from a comparison between a mainly CCATI mode and a mainly DPAPI mode in favour of CCATI. One can, of course, speculate whether the differences found in these two studies could be caused by the effects of a centralised versus a decentralised environment rather than by the effect of a computerised versus a paper and pencil data collection. Catlin and Ingram (1988), however, reached similar conclusions in a study where all interviews were performed

in a centralised environment. Nevertheless, the results are not unequivocal and our findings are in line with the conclusions of some other studies (Groves and Mathiowetz 1984; Groves and Nicholls 1986; Tortora 1985). The conservative conclusion seems to be that the effects of computer assisted telephone interviewing on data quality often are small and sensitive to the exact procedures used (both in the CATI and PAPI mode), the kind of variables under study, and the quality aspects that are in focus.

In our study, a reinterview procedure using delayed, indirect reconciliation was used. If the purpose of the reinterviews is to estimate response variance, an alternative approach is to use independent replication (though it can be discussed how "independent" the replication actually can be). If the purpose is to estimate bias, some kind of reconciliation procedure is needed which can be assumed to yield approximately true values. As we see it, from the viewpoint of estimating response variance, justified criticism can be launched against a standard reconciliation procedure since it tends to underestimate that variance. This is pointed out by Biemer and Forsman (1990) in the discussion of some results obtained in the American CPS. They also reported results from an Independent Replication Experiment in the CPS which indicated that the reinterviewers in the regular replication mode did not report all the discrepancies observed. However, the special type of indirect, delayed reconciliation used here should to a great extent avoid many of the problems of a standard reconciliation procedure. The reasons are:

a. It guarantees that the reinterviewer is not aware of the results of the original interview until after the reinterview is

finished and coded and the follow-up interview is to start.

b. After a discrepancy is found, the follow-up interview (reconciliation interview) concentrates on obtaining as complete information as possible about the respondent's work conditions. Nonetheless it is not a regular reconciliation interview since the interviewer as much as possible avoids asking the respondent why the recorded answers differ. Normally the respondent is not aware of the discrepancy.

c. If there is a discrepancy between the results from the original interview and the reinterview, all the material is gone through by an SLFS expert who decides the "true" answer. This is done even if the discrepancy disappears after the reconciliation interview.

We believe that the delayed, indirect reconciliation procedure used here can be recommended in certain reinterview settings, but it should be noted that it is a complicated procedure that demands special training as well as careful testing. It is best suited for a centralised reinterviewing environment and should preferably use a CCATI system (which was not used in the SLFS methods study).

No reinterviews were conducted in rotation groups 1–3 (see Section 3.1). This means that in the DPAPI mode almost all reinterviews concerned respondents originally interviewed according to the continuation questionnaire. In the DCATI mode, the complete first-time interview was used the first time a person was interviewed even when the person belonged to rotation groups 2–8. This might have biased the results in favour of DCATI, if it is assumed that a first-time interview is of higher

quality than a continuation interview. However, a study comparing first-time and continuation interviews for paper and pencil interviewing found no appreciable systematic differences in estimates.

It should be kept in mind that we used a DCATI system, which in many ways is different from a CCATI system. The procedure we used has many similarities to the CAPI procedure that is being introduced in other surveys at Statistics Sweden. The successful testing of the DCATI system gives some assurance that the CAPI system will work as well. However, whether the introduction of the computer in the respondent's home and in personal interviews will affect data quality cannot, of course, be answered by the present study.

6. References

Baker, R.P. and Bradburn, N.M. (1991). CAPI: Impacts on Data Quality and Survey Costs. Paper presented at the 1991 Public Health Conference on Records and Statistics, Washington, DC, July 15–17, 1991.

Biemer, P.P. and Forsman, G. (1990). On the Quality of Reinterview Data with Application to the Current Population Survey. Paper presented at the Sixth Annual Research Conference, U.S. Bureau of the Census, Washington, D.C.

Blom, E. (1985). Computer Assisted Data Collection at Statistics Sweden. Paper prepared for Meeting on Statistical Methodology, UN/CES, 17–20 February 1986, Statistics Sweden.

Catlin, G. and Ingram, S. (1988). The Effects of CATI on Costs and Data Quality: A Comparison of CATI and Paper Methods in Centralised Interviewing. In R.M. Groves, P. Biemer, L. Lyberg, J.T. Massay, W.L. Nicholls II, and J. Waksberg (Eds.), Telephone

Survey Methodology, New York: John Wiley.

Fink, J.C. (1983). CATI's First Decade: The Chilton Experience. Sociological Methods and Research, 12, 153–168.

Groves, R.M. (1983). Implications of CATI: Costs, Errors, and Organization of Telephone Survey Research. Sociological Methods and Research, 12, 199–215.

Groves, R.M. and Mathiowetz, N.A. (1984). Computer Assisted Telephone Interviewing: Effects on Interviewers and Respondents. Public Opinion Quarterly, 48, 356–369.

Groves, R.M. and Nicholls II, W.L. (1986). The Status of Computer-Assisted Telephone Interviewing: Part II – Data Quality Issues. Journal of Official Statistics, 2, 117–134.

Hofman, L. and Keller, W.J. (1993). Management of Computer Assisted Interviews in the Netherlands. Journal of Official Statistics, 9, 765–782.

Hubble, D.L. and Wilder, B.E. (1988). Preliminary Results from the National Crime Survey (NCS) CATI Experiment. Paper presented at the 1988 Joint Statistical Meetings of the American Statistical Association, New Orleans, August 22–25, 1988.

Kristiansson, K.-E. (1984). Gross-Flow Estimates in the Swedish Labour Force Surveys. Statistical Journal of the United Nations ECE 2, 357–367.

Nicholls II, W.L. and Groves, R.M. (1986). The Status of Computer-Assisted Tele-phone Interviewing: Part I – Introduction and Impact on Cost and Timeliness of Survey Data. Journal of Official Statistics, 2, 93–115.

Shoemaker, H.H., Bushery, J.M., and Cahoon, L.S. (1989). Evaluation of the Use of CATI in the Current Population Survey. Paper presented at the Joint Statistical Meetings of the American Statistical Association, Washington, D.C., August 1989.

Statistics Sweden (1990). The Swedish Labour Force Surveys. Bakgrundsfakta till arbetsmarknadsstatistiken, No. 3.

Statistics Sweden (1991). Report from the Field Test of the DATI Project in the Labour Force Survey August 1989 to January 1990. English translation of R & D Report 1990: 11, Statistics Sweden.

Tortora, R. (1985). CATI in an Agricultural Statistical Agency. Journal of Official Statistics, 1, 301–314.

Waite, P.J. and Rogers, P. (1991). 1989 and 1990 Current Population Survey CPS/CATI Reinterview Response Variance Report. Memorandum, U.S. Bureau of the Census.

Weeks, M.F. (1992). Computer Assisted Survey Information Collection: A Review of CASIC Methods and Their Implications for Survey Operations. Journal of Official Statistics, 8, 445–465.