Development of a Complex Longitudinal Computer-Assisted Questionnaire for Studying Infant Feeding[3]

James M. O'Reilly, Research Triangle Institute USA[4]

1. Introduction

We describe here the computer-assisted interviewing (CAI) systems developed for a longitudinal research project. The research project studies the infant feeding practices of a representative sample of women enrolled in the Special Supplemental Food Program for Women, Infants, and Children (WIC) of the U.S. federal government.

Of particular research interest are events and influences affecting the mother's decision to attempt to breastfeed, continue breastfeeding, terminate breastfeeding, and feed the child prepared foods. Researchers are interested in the timing of major changes and possibly associated factors such as work and schooling activities, interpersonal influences, and feelings and opinions. Previous research usually addressed these matters with a survey conducted about a year following the child's birth. A key problem is that many mothers cannot recall accurately detailed information about these events months later.

To address these problems, the Food and Consumer Service (FCS) of the U.S. Department of Agriculture specified that the 1994-1995 WIC Infant Feeding Practices Study would be conducted by monthly interviews with mothers or caretakers of the sample infant. The much briefer recall period would enhance accuracy in reporting events, activities, and subjective influences. A central virtue of this approach is that in the interview immediately following a significant change in feeding practices, detailed questioning can be conducted on circumstances surrounding the change at a time when recall would be less difficult.

Longitudinal data collection designs are superior to cross-sections for looking at changes over time. They are less often used, however, because of the greater cost and complexity. Cost is higher because of the multiple contacts with respondents. Complexity is greater because each interview after the baseline must include mechanisms to recall previous responses and conditions in order that change can be detected in the current interview and acted upon.

FCS also recognized that the WIC design would benefit significantly from computerized administration. A computer-assisted interview (CAI) approach can better handle the complexities of a longitudinal study--asking the proper question for each interview, instantly identifying that a key change has been reported, and asking appropriate questions about the change.

Working with FCS, Battelle developed the questionnaire and the CAI systems to conduct the study. The field work began in August 1994 and will be completed in late 1995. While other multi-wave surveys have been conducted with computer-assisted systems, we believe the WIC system breaks new ground in terms of the number of interviews (10), the rapid, monthly cycle, and the degree of integration of the longitudinal design into the CAI instrument. For a general description of the study design, development, and field experience see Williams et al. (1995).

2. Study design

An important issue in designing the study was that many of the low-income WIC mothers would be difficult to contact by telephone. Therefore a standard computer-assisted telephone interviewing (CATI) approach from a centralized calling facility would be problematic. One alternative was a centralized CATI interview followed by in-person interviews of a

supplemental sample of WIC mothers without telephones. A major difficulty of this approach is the substantial costs of interviewing the non-telephone households, since traveling interviewers would be required.

As with all longitudinal research, a critical design issue was retaining subjects in the sample through each cycle. Even a relatively small attrition rate per cycle may result in a final, cumulative response rate that jeopardizes some of the research goals of the study. With a sample population of younger, lower income women, many of whom may change living arrangements over the course of a year, retention seemed likely to be especially challenging.

Our experience in other demanding data collection situations suggested strongly that the key to retention would be establishing a close rapport with the respondent through one-on-one interactions with the same inteviewer over the entire duration of the study. The relationship between the interviewer and respondent could build through the course of the survey. As well, the interviewer could make use of her/his knowledge of the respondent and her household to solve problems of recontacts and tracing in later rounds.

Battelle proposed an innovative design in which there are be two roughly equal subsamples--one subsample of cases for CATI-only interviewing and one subsample of cases to be interviewed by CATI or by CAPI, if no telephone contact is possible. This design is able to produce the required estimates of the national WIC recipient population in both telephone and non-telephone households. An essential feature of this design was conducting the CATI work from interviewers homes, most of whom lived in or near the sampled areas.

Exhibit 1: Study sample design:

	Followup Mode		
	CATI-only	CATI-CAPI	Total
Sites	23	19	42
Participants per site	15	19	
Allowance for non-response	7	8	
Total participants selected	22	27	
Total participants (all sites)	506	513	1009

The design allows the same interviewers to conduct the CATI and CAPI interviews, providing them with sufficent work to retain their participation over the year-long interview period. The alternative of a centralized CATI system, supplemented by a local CAPI interview force to handle non-telephone households, provided many fewer work hours per site and inhibited establishing a close rapport with respondents.

The distributed CATI-CAPI design increases significantly the importance of the systems design for the field work. It must be capable of managing and monitoring interviews and interviewers spread across the country, and handle scheduling, case transfers, and many other functions that are taken as a given in a centralized CATI environment. Exhibit 1 describes the study sample design.

3. Questionnaire structure

The 329 substantive questions were organized into 31 sets which share similar content and timing of questioning. These are shown in Exhibit 2.

Exhibit 2: Description of WIC instrument question groups

1 Demographics of mother, child, household
2 Child and birth
3 Government support
4 Schooling
5 Work history
6 Work plans
7 Maternity leave
8 Child care and feeding at child care
9 Health problems during pregnancy
10 Prenatal care
11 Health problems following birth
12 WIC voucher use
13 Information/advice received from WIC office on
14 Knowledge of WIC feeding recommendations
15 Details on baby's birth and initial feeding
16 Baby's health and eating over past 7 days
17 Current breastfeeders--detailed feeding
18 Current breastfeeders--feeding problems
19 Current breastfeeders--feeding at work/school
20 Stopped breastfeeding since last time
21 Stopped breastfeeding since last time & currently formula feeding
22 Currently formula feeding
23 Breast and formula feeding
24 Formula feeding--detailed feeding
25 Currently feeding neither breast or formula and did formula last time
26 Feeding during past 7 days
27 Discussions how you intend to feed you baby with
28 Ever been told you should breastfeed by:
29 Prior breastfeeding influences
30 Informationi on breastfeeding from other than WIC
31 Agree/disagree about breastfeeding attributes, pluses, minuses

4. Distribution of questions across interviews

Three hundred and twenty nine substantive questions is not an especially large number for a survey. Yet, if asked entirely in one sitting, the interview would be an hour or longer. A virtue of the fully implemented longitudinal CAI design is that only the appropriate questions are asked for each interview. Exhibit 3 shows the counts of questions normally asked in the 11 wave.

Exhibit 3: Number of substantive questions "normally" asked per wave

Prenatal	94
Birth screener	7
MONTH 1	190
MONTH 2	161
MONTH 3	135
MONTH 4	101
MONTH 5	101
MONTH 6	116
MONTH 7	88
MONTH 9	101
MONTH 12	109

The number of questions which individual respondents are asked in a wave is substantally less than that shown in Exhibit 3. For example, all post-natal waves ask questions about why the mother stopped breastfeeding only in the month following that event. Similarly, the series of questions on formula feeding or feeding the baby other foods only apply to a subset of women in any month. As a result, the average length of the monthly interview is about 15 minutes. We believe enlisting and retaining cooperation in the study has been enhanced by the ability to structure the interview this way.

Orchestrating these questions is challenging, however. Each must be asked in the proper waves for the appropriate person. Special situations must also be handled correctly, such as making up key questions when a prior round interview was missed. Exhibit 4 shows part of a question matrix report developed so that project staff can determine which question are to be asked in which wave. For each question is shown

- whether it is normally asked in each of the wave PN, BS, M1 to M12 (Prenatal, birth screener, month 1 to month 12),
- whether the question is to be asked at the next wave if the earlier wave was missed (Ask Next) and
- whether the question is only to be asked of the natural mother not a caretaker (Mom Only).

Waves marked with an 'x' means to ask the question. Those with an 'xn' mean not to ask in subsequent waves if missed.

Exhibit 4: Section of Questions-Asked Matrix report

QUESTION	PN	BS	M1	M2	M3	M4	M5	M6	M7	M9	M12	Ask Next	Mom Only
BABYDOB		x										y	
TWINS		x										y	
BABHOSP		x										y	

MOMHOSP		x						y	y
DOBMOM	x							y	
RACEMOM	x							y	
BORN_US	x							y	
COUNTRY	x							y	
YRSINUS	x							y	
MRTLSTAT	x			xn			xn	y	
HHSIZE	x			xn			xn	y	
HSEHOLD	x			xn			xn	y	
HHREL	x			xn			xn	y	
BABYSDAD	x			xn			xn	y	y
HHONWIC	x							y	
MOINCOME	x							y	
FSPGRM	x		x		x		x		
AFDCPGRM			x						
MAIDPGRM			x						
MOMSMK	x							y	y
EDUCCOMP	x							y	
CURRSCHL	x		x	x		x	x		
HISTSCHL			x	x		x	x		
HRSSCHL	x		x	x		x	x		
PLANSCHL	x								y
CURRWORK	x		x	x		x	x		
PRENWORK	x								
HISTWORK			x	x		x	x		

The 329 questions are asked normally a total of 1095 times (3.32 times per question). Normal means asked in a specified wave. A number of more important questions are asked out of the ‘normal’ sequence when a prior wave is missed. These are asked in the next interview. Combining normal and make-up question occasions, the questions may be asked in a maximum of 1695 occasions (5.15 times per question).

Another important factor affecting which questions are asked is whether the baby’s caretaker is the natural mother or someone else. Some 208 questions are addressed only of the natural mother and the remaining 121 are asked of either the natural mother or other caretaker of the baby.

5. Tracking respondent information across the study

A key part of the design is dynamically tailoring the interview questionning based on information collected in previous months. For example, mothers are asked each month if the baby was immunized since the last interview. Once she responds ‘Yes’, the question is not asked in the subsequent wave. Many other questions which are dependent on previous responses relate to infant feeding.

Perhaps the most important for this study are those to women who are breastfeeding. Each month mothers are asked, how was the baby fed for the last 7 days--breastfeeding only, formula only, both, or neither.When a previously breastfeeding-only mother reports any other response, then a substantial series of questions are asked about the types of other feeding and when and why the change took place. Fourteen substantive questions are tracked from month to month to control these queries. These items are:

Baby vaccinated
Date of vaccinations
Current infant feeding method
Baby ever drank from a cup
Baby held a cup while feeding
Baby spoon fed
Baby fed from an infant feeder

Baby fed self
Mother's relatives breastfed
Mother's friends breastfed
Baby fed cereal,
Baby fed fruit,
Baby fed vegetables,
Baby fed meat

6. Developing the WIC System

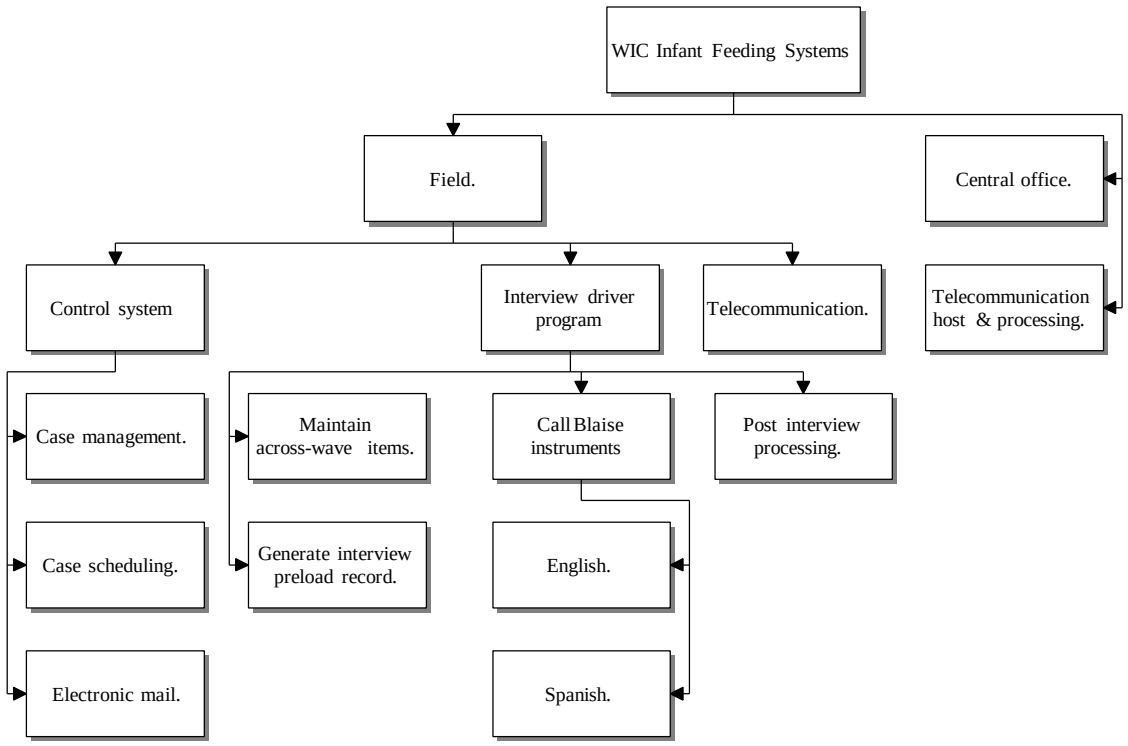
6.1 Systems design

In developing of the WIC CAI system a number of key characterics of the study had to be kept carefully in view:

1. The study is an original. No prior model would be available to follow.
2. The questionnaire was being created from scratch. One could expect changes even late in the process.
3. The system must function as either CATI or CAPI at locations anywhere in the U.S without any local support.
4. The sample population of younger, less educated, low-income women would be difficult to locate, track, and contact by phone. The ability to transfer cases from location to location, to specialized interviewers, and to conduct interviews from the central project office must be provided.
5. The interview cycle would be tight and demanding. Interviews for a given wave must take place in a narrow window of 15 days on either side of the baby's monthly birthday.
6. Cycling interview information from wave to wave so that changes are identified properly is crucial.
7. A distributed processing model is needed, rather than one depending on centralized processes to organize and manipulate monthly data and prepare it for the following month.
8. Daily automated telecommunication of data and mail from interviewers to the project staff is essential for data security, to transfer cases, to provide coordination and support.

Exhibit 5 displays the structure of the WIC project information systems. The central office and central telecommunications host system will not be discussed in detail here. The field system runs on the Toshiba 486/33 Model 1950 laptop PC in the interviewers home. The control system was written in Foxpro by staff of the Battelle/SRA St. Louis office. It's three major functions are general case management, case scheduling, and electronic mail. All are important, but for this project case scheduling is particularly critical because of the narrow window when cases can be interviewed for a wave.

Exhibit 5: Diagram of WIC project systems



6.2 Field telecommunications system

The field telecommunications system was a custom product developed in the Battelle/SRA Durham office. This is a general purpose program designed to handle the full range of automated data communications functions needed for field survey research. Nightly each interviewer selects the automated telecommunication function in the control system. At a preset time between 11pm and 7am each interviewer system calls into the telecommunications host. It uploads all new and changed data and mail files and downloads any files from the host system which the project staff wants the field system to receive. These may be mail messages, cases being transfer to the interviewer, commands for the field system to transfer a case out, or even new software systems. Because this system is completely automated and allows the central office to monitor and control all field system comprehensively and conveniently, managing a demanding and dispersed field study such as this is made much easier.

6.3 Design and development of CAI system

Key considerations in designing the Computer-Assisted Interviewing (CAI) system were:

1. Simplicity of design--because changing and revising the CAI instrument would be required late in the process with minimal time to do the work.
2. Capability of implementing a complete translation into Spanish after the English version become final.

The major programming issue was whether to develop a single comprehensive CAI instrument which included all the intricate logic to handle the myriad contigencies across all

11 waves, or develop 11 separate and more simple CAI instruments, one for each wave. Our first tack was to follow the orthodox systems wisdom--

- keep it separate, simple, and modular;
- avoid complex, comprehensive approaches.

However, after some initial design efforts and early prototyping, we concluded that a single complete instrument could be constructed. And by doing so we would avoid the burden of trying to maintain and change 11 separate versions of slightly different instruments.

The comprehensive approach meant there would be a single instrument and data structure for all waves. The essential unit of observation is a person-wave. Having the data record for all waves the same significantly simplified construction of the analysis file and the analysis itself.

6.4 Blaise Version 2.5 Development System

The CAI development software used is the Blaise system from Statistics Netherlands. We believe the use of the Blaise system was critical to the success of this work. It's conceivable that other systems might have been made to do what we accomplished in Blaise. But, it is very hard to believe they could have done it as well, as quickly (3-4 months), or for the 350 hours of programmer time expended.

One crucial characteristic of the overall questionnaire allowed us to develop essentially 11 separate, but related questionaires within one general Blaise instrument. The questions were always asked in the same order. This means that, for example, for Q43 and Q44, For any of the 11 interviews, any combination of these questions might be asked. But if they both are asked in a round, Q44 is always asked after Q43.

This is important because Blaise 2.5 has strict, structured IF-THEN-ELSE logic for control questionning. No 'GO-TO' statements to jump back against the question flow is permitted. If for one wave, Q44 could be asked before Q43, while in others it is the reverse, then using Blaise 2.5 would have required significant 'work-arounds', tricks, and other efforts.

6.5 Spanish version

Interviews are conducted in English or Spanish. To develop a Spanish version of the instrument our strategy was to wait until the English version of the questionnaire had reach a final state. Once the English CAI instrument was fully programmed, tested, and accecpted , and then we used Blaise tools generate an ASCII version of the questionnaire from the Blaise code, This version was translated into Spanish and the translated text transferred into the Spanish version of the Blaise instrument.

The structure of the Blaise language separates the question definitions--the text, fills, and answer definition--from the questionnaire process--the question order, skips, and related dynamic elements. This was a significant help with the Spanish version. We had to replace the question definitions. But the logic of the instrument was identical for both versions. As well, for interviewing there was one Blaise data file for an interview, whether it was conducted in English or Spanish.

6.6 Questionnaire logic

To understand how the questioning is handled in the Blaise instrument so that different sets of questions are asked in different waves and asked in the next appropriate wave is a wave is missed, Exhibit 6 shows the logic used to control when a set of 30 agree/disagree questions are asked. These questions, BOTLEASY (#350) to DIARRHEA (#379) are asked only of the

natural mother. They are to be asked normally in the prenatal and Month 2 waves. But if either of those wave are missed, then they are asked at the next interview.

Exhibit 6: Example Blaise code

```
if (((WAVE = PN) and (SKIP_PN <> 'Y')) or (WAVE = M2) or (MkUpM2 =
yes)) and
   (RisMom = YES) then
   BOTLEASY  ;  { 350 }
.
.
.
   DIARRHEA  ;  { 379 }
endif;
```

The Boolean expression in the if statement says: if the current wave is Prenatal and the Prenatal wave is not flagged to be skipped, or the current wave is Month 2 or the Month 2 is flagged for makeup (MkUpM2 = YES), and the respondent is the mother (RisMon = YES), then ask the questions. Otherwise don't ask. So a key element was the careful contruction of the Boolean expression which determine whether question are asked. Compared to the 'GO-TO' flow controls which would have be required in many other CAI languages, this was not difficult to do in the first place or, just as importantly, to understand weeks later as the instrument was changed.

6.7 Managing interviews and passing data between monthly interviews

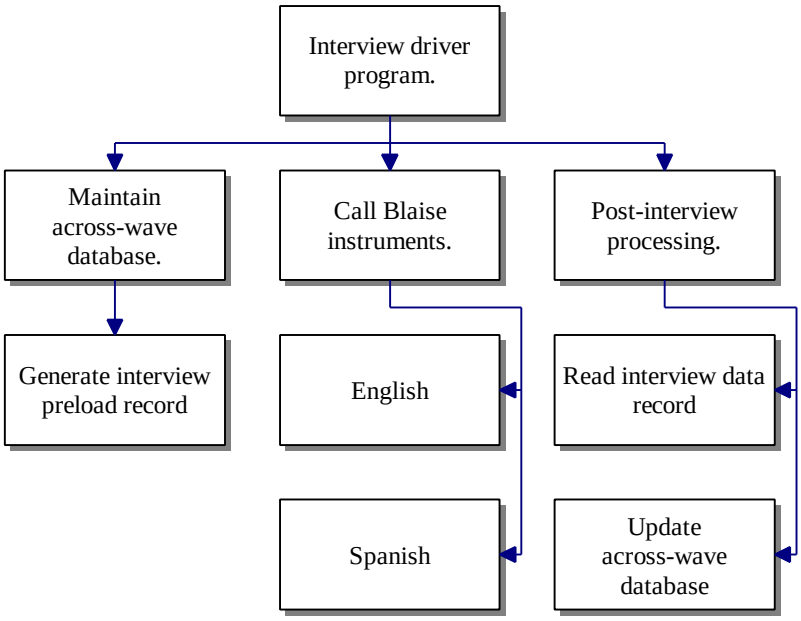
Blaise 2.5 is a powerful, robust, structured computer-assisted interviewing system. It is able to handle relatively large surveys. But it was never designed to manage very long interviews of more than one hour. Also it was not designed to manage complex multi-wave longtitudinal surveys.

However, it is reasonably straight-forward to extend the Blaise system to accomplish these more complex and elaborate surveys. This is possible because of the systems' strong development architecture for interview instrument programming, its data management system, and its DOS system tools for very quickly reading and writing Blaise interview data files and importing external data into the interview process. Together these provide a rich tool set for creating long and complex interview applications.

We did this previously in a very complex and lengthy CAI study, the 1993 National Survey of Family Growth pretest (O'Reilly, 1993). In that study, the instrument exceeded a number of fixed Blaise limits. As well, the instrument included a specialized event history module programmed in Foxpro and an Audio Computer-Assisted Self-Interviewing (ACASI) section. The solution was to develop seven major sections as separate Blaise 2.5 instruments. The overall instrument was managed by a custom Foxpro for DOS driver program which called each section. The driver also generated external Blaise data files of relevant responses from earlier sections, which were accessed by the current Blaise interview. This architecture worked successfully during the 1993 NSFG pretest (Lessler, et al., 1994).

A similar approach was used for managing the WIC longitudinal interviewing process. Each Blaise interview is unique for a study person and interview month. During each interview the instrument must access information about the study person's previous interview history to ask correct questions. This was done using a combination of a Foxpro "driver" program and Blaise tools, as shown in Exhibit 6.

Exhibit 7: Diagram of interview processes.



The major steps are:

- The Foxpro driver program maintains a database record for each case with information about interview status for each wave and responses for all interview data fields which must be passed to subsequent wave.

- When an interview is to begin, the driver program generates an ASCII data record for the case with the key previous interview data fields. Then it calls the Blaise data conversion program which reads the ASCII data record and generates a Blaise 'external' data file which the interview program will read.

- The drivers starts the Blaise interview program (in either English or Spanish), passing as parameters the case ID, type of interview (new or a restarted) and the names of the LIPS data file in which the interview data will be stored.

- Following the interview, when control returns to the Foxpro program, it calls a Blaise program to generate an ASCII copy of the Blaise interview data file, reads the ASCII file, and updates the case's Foxpro data record for information from the latest interview.

This approach worked well. On the notebook PC's, even though often the system was doing many steps to move from process to process--writing data files, calling DOS programs, running Blaise, etc.--the speed was quite adequate.

7. Conclusions

The 1994-95 WIC Infant Feeding Survey has now been in the field for one year and will be completed in a few months. Definitive results on how well the design and systems performed are not yet available. Yet, substantial preliminary statistics on the interviewing, data, and related areas all point to a very successful outcome. Wave-to-wave retention rates have been very high. Overall cooperation of this normally difficult to interview population have been quite strong. See Williams, et al. (1995) for detailed results.

In terms of the information systems, the conclusions look similarly positive. We believe much of the overall success of the survey is a function of the innovative distributed CATI/CAPI design. Gratifyingly, the systems described here were flexible, powerful and robust enough to implement the challenging design. The Blaise system was able to render effectively virtually every features of the questionnaire the substantive experts required. The complex web of questionning based on the child's caretaker, previous interviews missed, previous interview responses and other variables was accomplished to near perfection.

The system also succeeded in managing the overall data systems of some 20 distributed interviewers and a central office so that cases are interviewed at the appropriate times, transferred to other interviewers when necessary, and are available for a central office telephone interview in the case of an incoming call from the respondent. The custom Battelle field telecommunication systems worked successfully managing the nightly automated calling.

We are pleased that this systems design, built around the strength, integrity, and extensibility of the Blaise 2.5 system, has succeeded in supporting this important research.

References

Lessler, J.T. Weeks, M.F., O'Reilly, J.M. 1994 "*Results from the National Survey of Family Growth Cycle V Pretest*". A paper presented to the Annual Meetings of the American Statistical Association. Toronto, Canada. August 1994.

O'Reilly, J.M. 1993. "*Lessons Learned Programming a Large, Complex CAPI Instrument*." A paper presented at the International Blaise Users Conference. London, England. October 1993.

Williams, R.L., O'Reilly, J. M., Vesper, E. 1995. "*Design and Implementation of a Longitudinal Infant Feeding Survey*". Paper presented to the Annual Meeting of the American Statistical Association. Orlando FL, August 1995.