PSID A national study of socioeconomics and health over lifetimes and across generations

Technical Series Paper #95-02

1995 Panel Study of Income Dynamics Results of an Experiment Designed to Evaluate Central and Decentral Data Collection Methods

Report prepared by:
Stephanie Chardoul

With assistance from:
Tom Gonzales
Lesli Scott

This project was supported by funding from the National Science Foundation (SES 9022891).

1995 PANEL STUDY OF INCOME DYNAMICS

Results of An Experiment Designed to Evaluate Central and Decentral Data Collection Methods

Report Prepared By:
Stephanie Chardoul

With Assistance From:
Tom Gonzales
Lesli Scott

Report Prepared For:
Sandy Hofferth
Frank Stafford
Beth-Ellen Pennell

Panel Study of Income Dynamics Project
Survey Research Center
University of Michigan
Ann Arbor, MI 48106-1248

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Between the months of January and July, the Field Section of the Survey Research Center (SRC) implemented the data collection activities for the 1995 Panel Study of Income Dynamics. As part of the data collection plan, half of the data collection activities were completed by Interviewers working in an SRC based central computerized telephone facility and half were completed by decentrally located Interviewers across the nation who worked with laptop computers from their homes. This report describes the design of this central and decentral mixed data collection effort and presents the results of analytical comparisons made between central and decentral modes.

A. BACKGROUND

The PSID's first data collection wave occured in 1968, with personal interviews collected by decentralized Interviewers. During the '70s and '80s, the data collection continued to be carried out by the decentralized staff, but the interviews switched from face-to-face to telephone mode. Due to cost and data quality considerations, the PSID was given the directive to both centralize and computerize their data collection efforts. This process began in 1991, with a small number of production cases being assigned to the Ann Arbor Telephone Facility and the beginning of development of the CATI application. The majority of data collection was centralized in 1992, and CATI development continued until its full implementation in 1993.

Until 1995, all decentralized PSID interviewing was done using paper questionnaires (which were then manually entered into the CATI system). Throughout the early '90s, several large-scale SRC projects successfully used computer-assisted interviewing in a decentralized setting. Due to the improvements to the decentralized sample management systems, and to the rising cost of maintaining parallel paper and computerized procedures, both the Survey Operations Unit and the PSID staff felt comfortable with converting the decentralized interviewing efforts to a computer-assisted mode in 1995.

In order to make the conversion to computer-assisted decentralized interviewing cost effective, a larger number of cases needed to be completed using the decentral mode. The 50%/50% split between central and decentral locations was decided on (partly due to the fact that a smaller split would not have the power to be indicative of variance), with a goal of evaluating both data quality and cost issues.

B. PRIMARY OBJECTIVES OF DESIGN

The primary objectives of the mixed data collection design are as follows:

- Determine the procedural and administrative requirements associated with mixing central and decentral data collection efforts
- Identify the comparative cost advantages associated with both the central and decentral modes
- Identify the comparative productivity advantages associated with both the central and decentral modes
- Identify the comparative quality advantages associated with both the central and decentral modes
- Identify flexibility gains and constraints associated with mixed mode design.

This report contains various analyses, providing information on the above issues.

II. SUMMARY OF KEY FINDINGS AND RECOMMENDATIONS

- The procedural and administrative systems allowed interviewing to effectively occur in both the centralized and decentralized settings. Transfers from the central facility to decentral Interviews was easily done; transfers from the decentralized setting to the central facility were more cumbersome, and we recommend that this procedure be improved to maximize flexibility and effectiveness. There were very few accounting systems for the face-to-face cases, which did detract from our analysis. It should be noted that there was a high proportion of experienced Interviewers on this year's project and also a high level of training and quality control procedures for the new staff.

- The analysis shows the cost factors to be fairly even. Minimizing the decentralized training costs may give the decentralized setting a cost advantage. It was found that decentral interviews take an average of 1.3 fewer calls to complete and are an average of 6.4 minutes shorter than central interviews. Decentralized interviewing was also more efficient, averaging 2.2 hours per interview compared to the centralized 2.4 hours per interview.

- Productivity was also found to be fairly even, taking into account the faster start of the decentralized interviewing--due largely to the higher experience level of the decentral Interviewers. The decentralized staff was able to maintain a higher level of production, taking on the majority of cases (including many of the most difficult ones).

- The quality analysis showed differences in the application of standard interviewing techniques. Behavior coding showed significantly higher occurances of non-standard interviewing behavior in the decentralized staff. The difference seems to be related to both training and quality control, which we recommend be a stronger focus for future decentralized trainings. A review of general error types showed great similarity between the centralized and decentralized staffs; a review of specific error types showed higher occurances of problems with key concepts in the decentral interviews. No significant differences existed in the central and decentral responses to persuasion letters.

- The flexibility of having both modes available kept costs down by being able to easily move cases to where they would be best worked. The number of production hours was also reduced from previous years; we are estimating that the production hours used will be significantly less than the amount budgeted.

- The findings from our various analyses are presented here, while we realize that further investigation is needed to explain elements of variance between the centralized and decentralized staffs (for example, an explanation of the longer interview length for centralized cases). This further analysis will be done in the near future.

III. METHODOLOGY

A. GENERAL DESCRIPTION

The Panel Study of Income Dynamics (PSID) is a panel, longitudinal survey that has been following and interviewing the same families since 1968. Data collection is conducted annually with original "core" respondents and additional "latino" respondents who were added in 1990. Until 1992, all interviewing was done by a decentralized interviewing staff, using paper and pencil questionnaires; in 1992, data collection efforts were centralized in Ann Arbor, and in 1993, the centralized interviewing was conducted using computer assisted methods. In 1995, the decision was made to split the sample between the centralized and the decentralized interviewing staffs.

In 1995, the data collection period runs from the end of February through the middle of August with a final goal of 8880 completed Core interviews. The predicted overall average interview length is 34.0 minutes.

B. RESEARCH EXPERIMENT DESIGN AND PROCEDURES

Sample Management

Both the Core and Latino samples were split evenly between the centralized and decentralized interviewing staffs. First, those cases that were known to need face-to-face visits (and their related sample lines) were flagged to go to the decentralized Interviewers. The remaining cases were divided by 68ID, with even-numbered cases being assigned to the centralized staff and odd-numbered cases being assigned to the decentralized staff, with the assumption that there were no significant differences between the odd and even 68IDs.

The sample lines were released in a series of five batches. The original plan and the original production goals called for each release to be split approximately evenly between the centralized and decentralized staffs. This occurred until the final release when, due to staff changes in the centralized Telephone Facility, the decision was made to send the majority of the final Core cases to the decentralized staff. Table 1 shows the division of cases by release, and also the date on which each release became available to the Interviewers.

Table 1. Number of Cases Sent to Central and Decentral Staff in Each Sample Release

	Release 1 Core [2/21]	Release 2 Latino [3/2]	Release 3 Latino [3/2]	Release 4 Core [3/27]	Release 5 Core [4/24]	**Total**
Central	2376	664	351	881	364	**4636**
Decentral	2406	794	412	960	1957	**6529**
Total	**4782**	**1458**	**763**	**1841**	**2321**	**11165**

General Operating Procedures

To keep the data collection efforts as comparable as possible, individual case assignment by Interviewer was adopted in the Telephone Facility; this was the first time that this type of sample management has been utilized by the centralized staff. An approximately even number of Interviewers were hired and trained for the 1995 PSID project in both the Central (66) and Decentral (68) settings, and the study-specific portion of the training was designed to be as similar as possible (given some physical limitations) for the two groups. The centralized staff were very familiar with using computer-assisted systems to collect PSID data (the project has been computerized in the central setting since 1993); 1995 was the first time that the decentralized staff used laptop computers to collect PSID data, but all of the Supervisors and Interviewers had used computer-assisted interviewing on other projects.

The decentralized staff were led by five Regional Field Supervisors and five Team Leaders; the centralized staff were led by the Telephone Facility Assistant Manager, who coordinated the work of four Shift Supervisors and five Team Leaders. A Project Manager and Research Assistant oversaw the entire project, ensuring that sample management and data collection procedures remained consistent between the two locations. Please see Appendix 1 for a complete description of quality control and production reporting procedures.

Data for this analysis (reflecting cost, productivity, quality, and flexibility) have been collected and examined throughout the production period. These data are still being compiled, and a final analysis cannot occur until the completion of the project.

(See Appendix 1: Central and Decentral Mode Approaches to '95 PSID Data Collection.)

C. CONSIDERATION OF DESIGN LIMITATIONS

Considering all of the above factors, we believe that a meaningful analysis can be conducted. There are, however, several design limitations that contribute to slightly less than ideal conditions.

One of these concerns is the feasibility of separating out costs for the face-to-face cases. Although face-to-face cases are identified in the Control File, Interviewers do not distinguish them in their cost reporting. For this analysis, we will report on face-to-face cases separately wherever possible, but will not be able to distinguish them in some analyses.

Another concern is the lack of standard interviewing quality comparisons, and the low-level amount of behavior coding that we have been able to complete (due to both time and monetary constraints). We realize that 75 cases from each location is not a large enough number to demonstrate statistically significant differences between the two modes, but, for our purposes, we believe that the data from this mini-experiment will be sufficient to make general comparisons.

The last limitation is the inability to compare study close-out success, an important factor of a large project like the PSID. Because of the disproportionate division of Release 5 cases (a decision made for the benefit of the overall project, not this experiment), the centralized and decentralized operations faced very different final weeks of the project. Also, it should be noted that Release 5 contained cases known to be historically difficult for Interviewers to complete. By the end of June, all cases remaining with the centralized staff were transferred to the decentralized staff (to increase overall efficiency). Due to this, all comparisons include data only through the middle (or end) of June.

IV. RESULTS

A. COST COMPARISONS

The cost comparisons made for this report consider the aspects of the 1995 costs that would be relevant under steady state conditions. These include project management, general operations (control programming, administrative, clerical support), interviewing, supervisory, non-labor (telecommunications, equipment, supplies), and training costs. These exclude computer system design costs which uniquely occur during the 1995 data collection period.

Measures of cost include:

- Overall budgeted and actual costs for steady state cost categories
- Hours per interview throughout the data collection period
- Average number of calls required to attain interviews
- Average length of interviews

In addition to the results found by comparing costs in the central and decentral modes, this report presents information about:

- Comparative costs of experienced and new interviewers
  - -- Comparison of average hours per interview
  - -- Comparison of average calls per interview
- Cost advantages that could only occur by mixing central and decentral modes

---

Overall Budgeted and Actual Costs

Table 2 (Overall Steady State Budget Comparisons for Central and Decentral Modes) compares Core costs between the decentral and central modes. All costs except Interviewer and Supervisor Travel are included in this comparison, including costs associated with the face-to-face cases in the various decentral line items.

Though the original budget assumed an even split of completed interviews between the central and decentral modes, approximately 60% of the interviews have been completed by the decentralized interviewing staff. This additional effort has caused an increase in all decentral cost categories, except Training.

Costs are shown both excluding and including Interviewer Training. Because the decentral staff had not had a formal PSID training in over five years, the 1995 training was used to bring a large number of Interviewers onto the project. In future years, this extensive training effort will not be necessary with the decentral staff.

Decentral General Operations costs are inflated, due to staff members charging disproportionally. This discrepancy will be corrected before the final analysis.

The average cost per interview excluding training shows decentral interviews costing slightly less than central interviews; the average cost per interview including training shows central interviews costing slightly less.

Table 2. Overall Steady State Budget Comparisons for Central and Decentral Modes
(data as of 7/1/95)*

ITEM	DECENTRAL CORE		CENTRAL CORE	
	BUDGET	ACTUAL	BUDGET	ACTUAL
# of Interviews	4440	4931	4440	3224
COST CATEGORIES				
General Operations	$ 60,576	$ 84,448	$ 92,789	$ 65,195
Project Management	$ 21,296	$ 10,482	$ 21,297	$ 11,078
Supervision	$ 57,723	$ 48,482	$ 61,929	$ 35,029
Interviewing	$ 157,985	$ 132,443	$ 116,554	$ 66,160
Non-Salary	$ 58,018	$ 67,011	$ 74,466	$ 48,910
Training	$ 67,274	$ 59,479	$ 25,012	$ 28,917
COST EXCLUDING TRAINING	$ 355,598	$ 342,866	$ 367,035	$ 226,372
COST PER IW EXCLUDING TRAINING	**$ 80.10**	**$ 69.53**	**$ 82.67**	**$ 70.22**
TOTAL COST INCLUDING TRAINING	$ 422,872	$ 402,345	$ 392,047	$ 255,289
COST PER IW INCLUDING TRAINING	**$ 95.24**	**$ 81.60**	**$ 88.30**	**$ 79.18**

* Decentral Core budget and costs include face-to-face cases (145 interviews, as of 7/1/95)

Hours per Interview

Figure 1 (Production Hours per Interview) maps the average hours per Core interview, by week, throughout the production period (data not available prior to 3/26/95). The graph shows that the decentral interviewing remained more efficient than the central interviewing, with a sharp increase in the hours per interview with the release of the last batch of cases (as noted earlier, these cases are the most difficult to complete--due to both tracking and persuasion efforts).

The plots on the graph represent weekly hours per interview figures. As of 20 June, the Core **cumulative** average hours per interview for the centralized staff was 2.4 and for the decentralized staff was 2.2.

Average Number of Calls

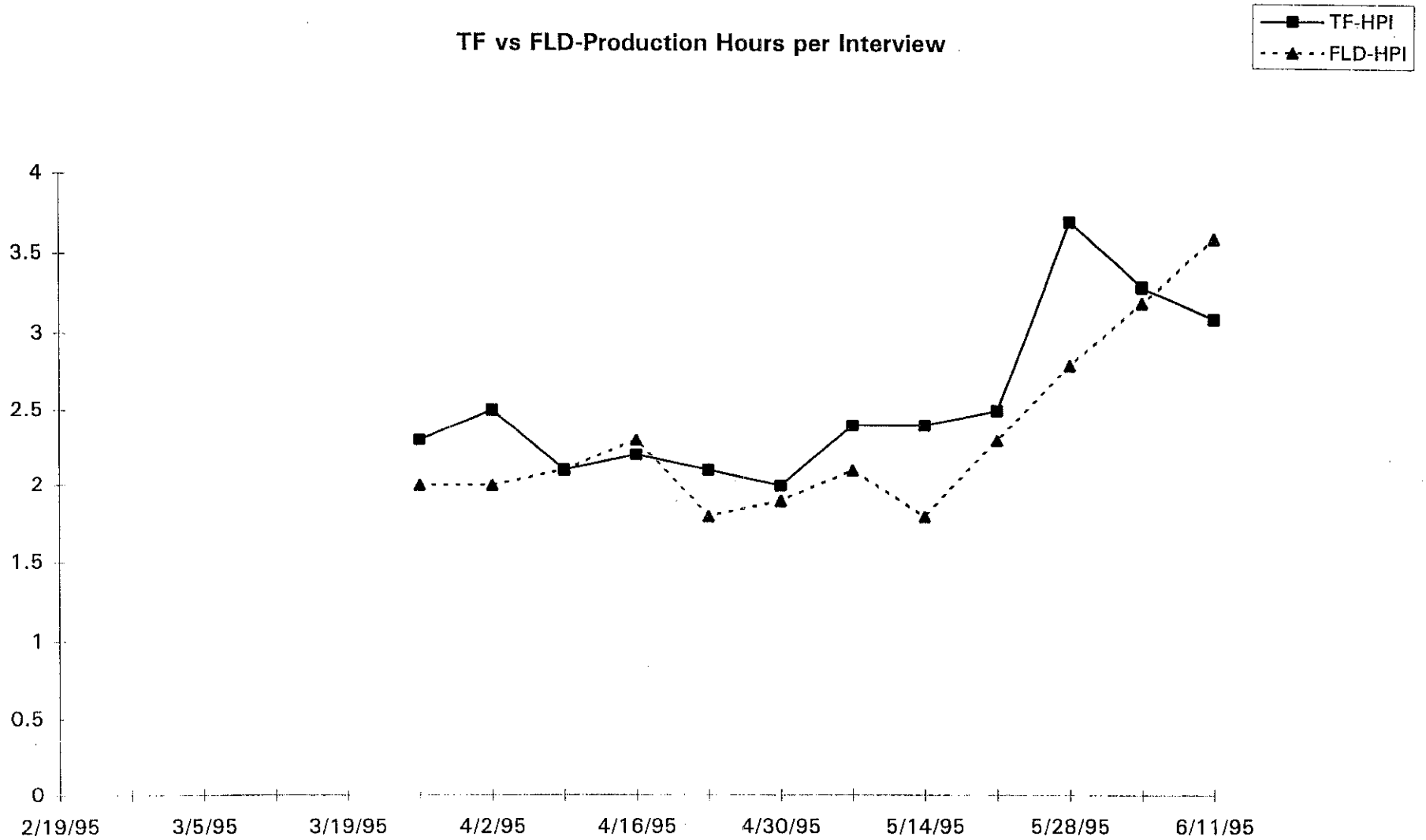
Another measure of cost is the average number of calls necessary to complete interviews. The table below compares the average number of calls needed by decentral and central staffs in the various case status categories.

Table 3. Average Number of Calls Comparison for Core Central and Decentral Cases

MODE	AVERAGE # OF CALLS PER CASE STATUS CATEGORY				
	No Contact	Contact	Resistance	Interview	Non-Interview
Central	7.2	10.6	17.0	6.1	9.3
Decentral	8.4	16.2	12.2	4.8	9.1

The decentral Interviewers average a higher number of calls in the outstanding non-resistance cases, but average a lower number of calls per completed interview.

FIGURE 1. CENTRAL/DECENTRAL PRODUCTION HOURS PER INTERVIEW


TF vs FLD-Production Hours per Interview
TF-HPI
FLD-HPI
4
3.5
3
2.5
2
1.5
1
0.5
0
2/19/95
3/5/95
3/19/95
4/2/95
4/16/95
4/30/95
5/14/95
5/28/95
6/11/95

Interview Length

Interview length affects costs by influencing Interviewer time and efficiency, phone charges, and respondent reaction to the interview. The table below compares the average Core interview (and section) lengths for central and decentral staffs.

Table 4. Average Core Interview and Interview Section Lengths for Central and Decentral Staffs (in minutes)

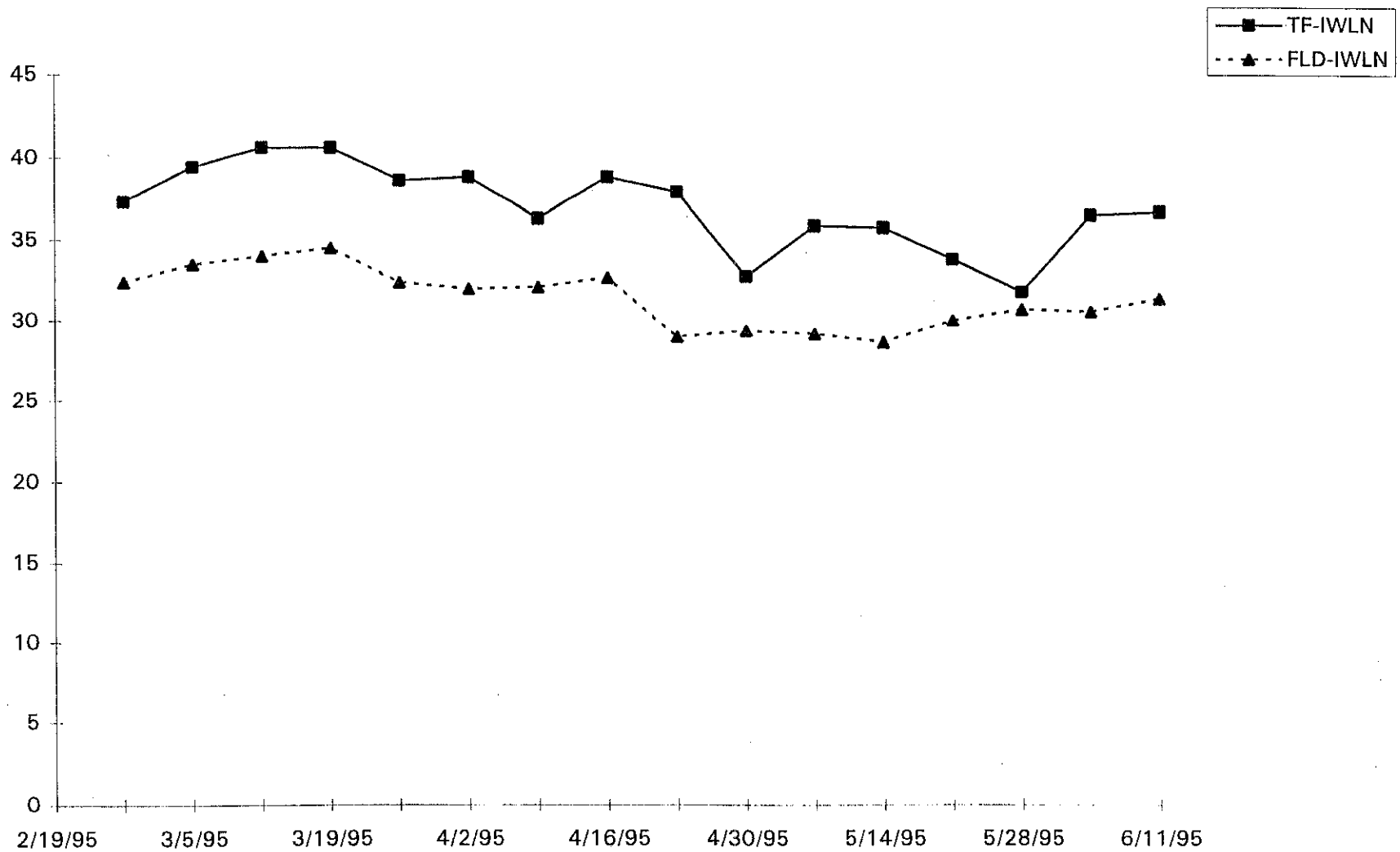
	INTERVIEW SECTION													
MODE	**Tot IW**	CS	A	B-E	F	G	H	HS	PS	J	K	L	M	AD
Central	**37.8**	9.3	3.4	14.1	2.4	8.1	1.7	1.8	3.4	0.7	4.1	7.2	4.1	6.9
Decentral	**31.4**	3.4	3.2	10.1	2.3	6.7	1.5	2.4	3.0	0.8	3.8	6.8	4.4	6.6

As shown in the table above, for almost every section, the decentralized interviews are (in some cases significantly) shorter in length than the centralized interviews. The difference is most noticeable in the Coverscreens, Sections B-E (employment history), and Section G (income). For Section HS (Health Supplement), the decentralized interview length is considerably longer. Overall, the centralized interviews are an average of 6.4 minutes longer than the decentralized interviews.

Figure 2 plots the average Core interview length for central and decentral interviews for the entire production period.

FIGURE 2. CENTRAL/DECENTRAL AVERAGE INTERVIEW LENGTH

TF vs FLD-Average Interview Length



Comparative Costs of Experienced and New Interviewers

In addition to examining the cost differences between the central and decentral interviewing modes, we also wanted to explore the comparative costs of experienced and new Interviewers. For the purpose of this comparison, Interviewers (regardless of mode) are considered "experienced" if they have spent at least one year interviewing on the PSID project.

Hours per Interview
Figure 3 charts the weekly average hours per interview for both experienced and new Interviewers. At the beginning of the project, the experienced Interviewers had a slight advantage (with 2.2 hours per interview, versus the 2.8 hours per interview for new Interviewers), but the new Interviewers needed only a short time to get up to speed and begin producing at a level equal to the experienced Interviewers. By the end of the analysis period, the new Interviewer hours per interview started to decline, due to the experienced Interviewers taking over the most difficult tracking and persuasion cases.

Overall, both experienced and new Interviewers attained the same Core cumulative average hours per interview: 2.3.

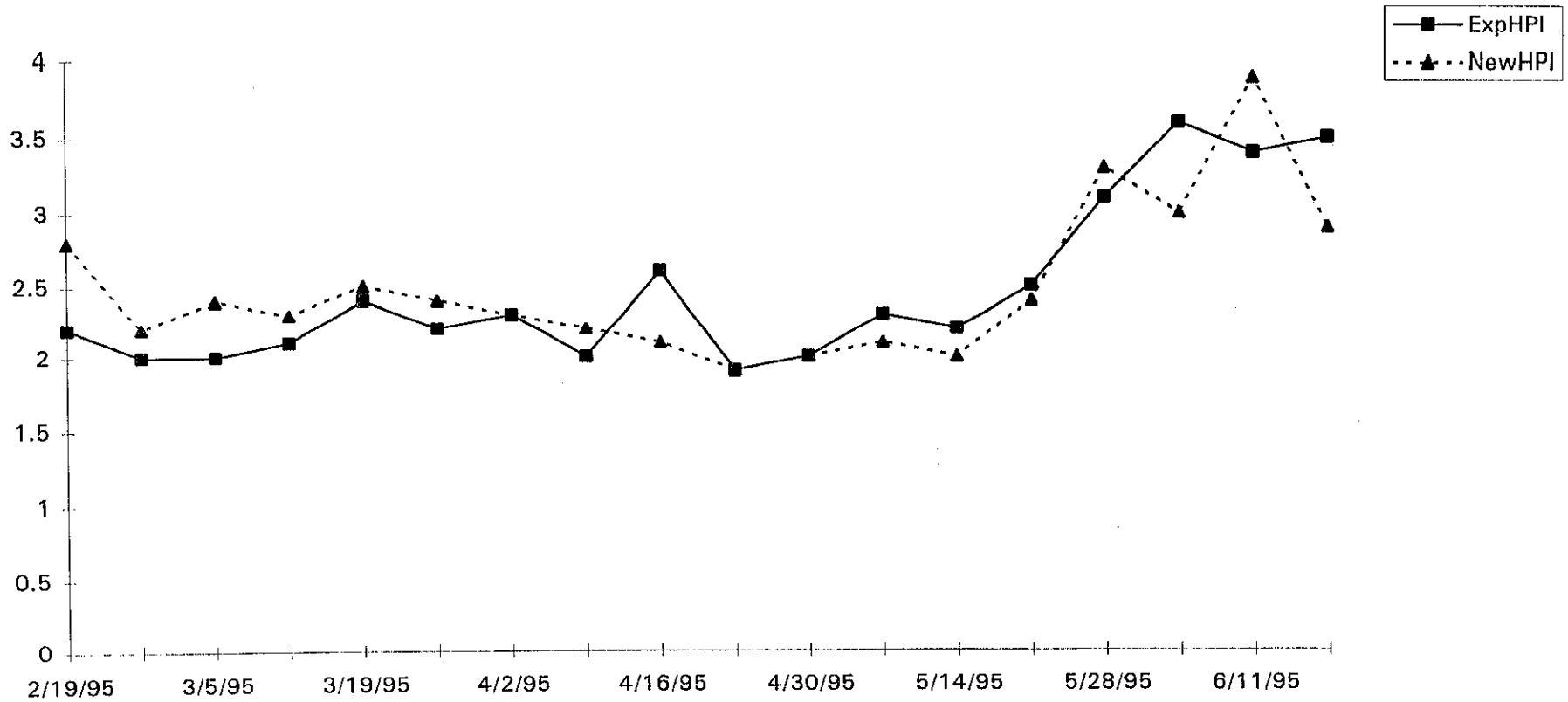
Average Number of Calls
Table 5 shows the average number of calls to attain the various case status categories, by experienced and new Interviewers. As the figures in the table demonstrate, there is no significant difference between the experienced and new staffs.

Table 5. Average Number of Calls Comparison for Core New and Experienced IWers

LEVEL	AVERAGE # OF CALLS PER CASE STATUS CATEGORY				
	No Contact	Contact	Resistance	Interview	Non-Interview
Experienced	11.0	17.7	11.2	5.8	9.4
New	8.8	18.6	12.2	5.1	9.4

FIGURE 3. EXPERIENCED/NEW INTERVIEWER
PRODUCTION HOURS PER INTERVIEW

EXPERIENCED vs NEW IWER Hours per Interview



B. PRODUCTIVITY COMPARISONS

Several productivity measures were examined that compared ability to attain high response rates on the study within the specified data collection period. These measures took into account the ability to effectively estimate and obtain production levels based on anticipated staff efficiency and retention. The measures include:

- Projected and actual Interviewer production hours
- Completed interviews plotted throughout the study period
- Goals, goal adjustment, and goal attainment

Interviewer Production Hours

The following table shows budgeted and actual Interviewer hours for centralized and decentralized Core interviewing. The budgets were prepared for the entire production period and the latest Cost Report covers through 27 July, so some discrepancy is unavoidable. Also, the budgeted amounts do not take into account the majority of Release 5 cases being assigned to the decentralized staff and also the transfer of **all** outstanding Core cases to the decentralized staff by the end of June.

Table 6. Budgeted and Actual Interviewer Production Hours

MODE	BUDGETED CORE HOURS	ACTUAL CORE HOURS*	PERCENT OF BUDGETED HOURS USED
Central	15,278	8,850	58%
Decentral	16,416	17,137	104%

* as of 27 July

The movement of additional cases to the decentralized staff was not a direct function of either staff's performance, but was a result of staff changes in the centralized Telephone Facility and an effort to make the end of production as efficient as possible (by consolidating remaining cases with a fewer number of Interviewers). The decentralized staff was able to accomodate the larger case load by transferring cases to those Interviewers who had already finished their initial sample or who were not working on other projects and had additional time to devote to the PSID. Overall, we estimate that we will have used several thousand

fewer Interviewer hours than were budgeted. Further investigation is needed to determine whether this is one advantage associated with the flexibility of systems.

Completed Interviews

Figure 4 (1995 PSID TelFac vs. Field Production) charts interview completion for both Core and Latino samples by the centralized and decentralized staffs. (Again, for this comparison, we are examining only the Core sample.)

Because of the high proportion of new Interviewers in the centralized setting, production was purposely metered at the beginning of the study, to allow the supervisory staff sufficient time to perform quality control measures. This slower start is clear on the graph, and the centralized staff maintained this production gap throughout the production period.

The separate case release dates are also noted on the chart. Most noteworthy is the final Core release (labeled as Batch 4) where the majority of cases were sent to the decentralized Interviewers, resulting in a dramatic increase in the gap between decentralized and centralized production figures. It also should be noted that regular centralized interviewing stopped at the end of June, so the production count remains stable after that date.

Goals, Goal Adjustments, and Goal Attainment

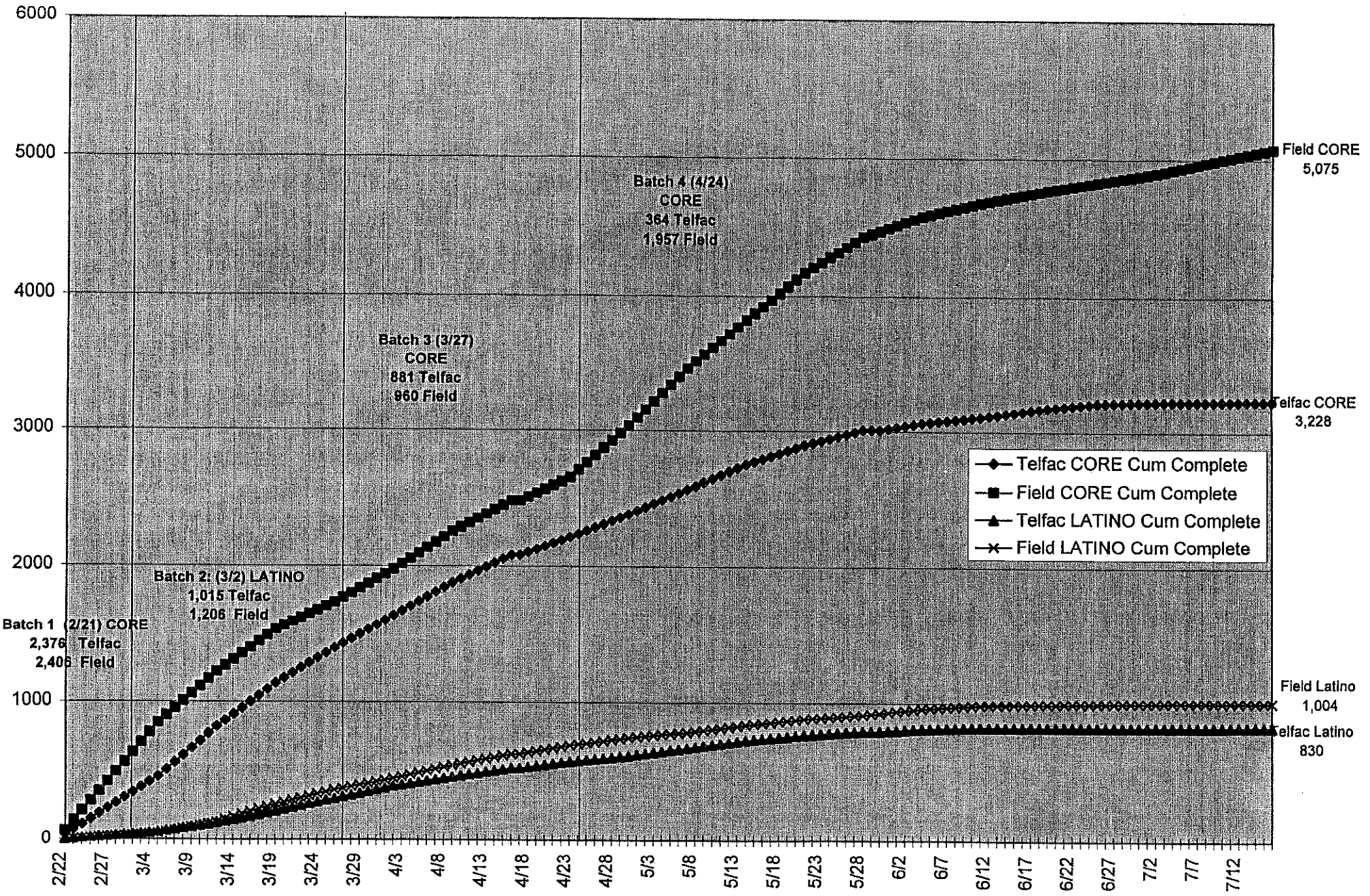
Tables 7 and 8 (Decentralized and Centralized Core Sample Production Goals--By Week) show weekly and cumulative goals and production for both decentralized and centralized interviewing. The goals were devised before production began; the thick line after 6/4 indicates a goal adjustment. These tables do reflect the revision of goals to account for the additional Release 5 sample being sent to the decentralized Interviewers. The "% OF TOTAL COMPLETE" column corresponds to the cumulative goal column, not the actual production.

By the week of 6/4, when the goals were adjusted to bring them into line with actual production, the decentralized staff was 245 interviews behind their goal of 4950 and the centralized staff was 296 interviews behind their goal of 3400. Two weeks later, when all cases were transferred out of the centralized facility, the decentralized staff was 42 interviews behind their goal of 4903 and the centralized staff was 51 interviews behind their goal of 3262.

Figure 5 charts weekly Core response rates for the centralized and decentralized staffs. Response rate is calculated by dividing the number of completed interviews by the total number of cases minus non-sample cases [(IW) / (TOT SAMP - NS)]. Throughout the production period, the weekly decentralized response rate remained higher than the centralized. As of 20 June, the Core **cumulative** response rate for the centralized staff was 89.6% and for the decentralized staff was 89.3% (the lower decentralized response rate is due, in part, to the heavier case load and also to the difficult cases contained in Release 5).

FIGURE 4. CENTRALIZED/DECENTRALIZED PRODUCTION

1995 PSID Telfac vs. Field Production *



* *Source: July 20, 1995 PSID Control File Report.*

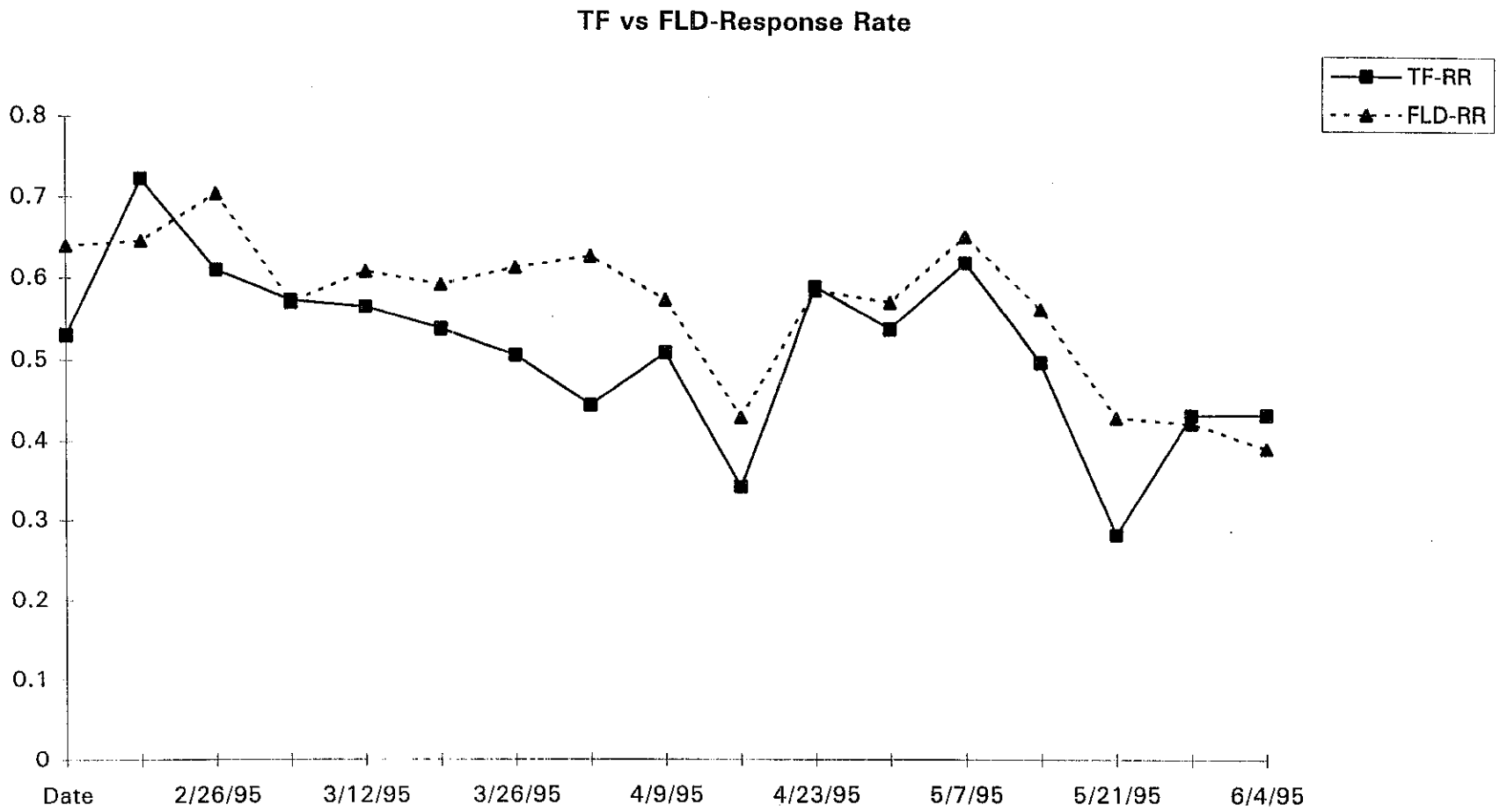
Table 7
1995 PSID Decentralized Core Sample Production Goals--By Week

WEEK OF...	WEEKLY GOAL	ACTUAL WEEKLY COMPLETE	CUMULATIVE GOAL	ACTUAL CUMULATIVE COMPLETE	% OF TOTAL COMPLETE
2/19&2/26	222	850	222	850	5%
3/5	310	367	532	1217	12%
3/12	444	314	976	1531	22%
3/19	444	201	1420	1732	32%
3/26	444	246	1864	1978	42%
4/2	266	284	2130	2262	48%
4/9	374	229	2504	2491	54%
4/16	374	172	2878	2663	60%
4/23	374	370	3252	3033	66%
4/30	374	429	3626	3462	72%
5/7	374	368	4000	3830	78%
5/14	320	346	4320	4176	83%
5/21	262	268	4582	4444	87%
5/28	206	159	4788	4603	90%
6/4	162	102	4950	4705	92%
6/11	110	86	4803	4791	91%
6/18	100	70	4903	4861	93%
6/25	85		4988		95%
7/2	80		5068		96%
7/9	70		5138		97%
7/16	70		5208		99%
7/23	52		5260		100%

Table 8
1995 PSID Centralized Core Sample Production Goals--By Week

WEEK OF...	WEEKLY GOAL	ACTUAL WEEKLY COMPLETE	CUMULATIVE GOAL	ACTUAL CUMULATIVE COMPLETE	% OF TOTAL COMPLETE
2/19&2/26	420	459	420	459	12%
3/5	270	362	690	821	20%
3/12	270	312	960	1133	27%
3/19	270	258	1230	1391	35%
3/26	386	242	1616	1633	46%
4/2	243	245	1859	1878	53%
4/9	249	206	2108	2084	60%
4/16	232	137	2340	2221	67%
4/23	243	170	2583	2391	73%
4/30	297	186	2880	2577	82%
5/7	165	181	3045	2758	87%
5/14	130	138	3175	2896	90%
5/21	95	112	3270	3008	93%
5/28	65	51	3335	3059	95%
6/4	65	45	3400	3104	97%
6/11	79	62	3183	3166	91%
6/18	79	45	3262	3211	93%
6/25	78		3340		96%
7/2	38		3378		97%
7/9	38		3416		98%
7/16	37		3453		99%
7/23	37		3490		100%

FIGURE 5. CENTRALIZED/DECENTRALIZED CORE RESPONSE RATES


TF vs FLD-Response Rate
TF-RR
FLD-RR
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Date
2/26/95
3/12/95
3/26/95
4/9/95
4/23/95
5/7/95
5/21/95
6/4/95

C. QUALITY COMPARISONS

The measures of quality used in this report include:

- Comparisons of refusal and refusal conversion rates
- Measures of Interviewer and Respondent behaviors for interviews completed in central and decentral settings (results of behavior coding)
- Estimated relative error rate for major/minor conceptual errors
- Measure of effectiveness of persuasion letter program in central and decentral settings.

Refusal and Refusal Conversion Rates

Table 9 compares the number of refusals (initial and final) and refusal conversions for central and decentral interviewing (also broken down by experienced and new staff in each location). The number of initial refusal cases is also compared against the total numer of cases in each location.

Table 9. Core Refusal and Refusal Conversion Rates

MODE	INITIAL REFUSALS (% of base)	FINAL REFUSALS	REFUSAL CONVERSION INTERVIEWS	REFUSAL CONVERSION RATE
Central	**72 (2.1%)**	**32**	**39**	**54.2%**
Experienced	49 (2.4%)	18	30	61.2%
New	23 (1.8%)	14	9	39.1%
Decentral	**305 (5.3%)**	**46**	**122**	**39.9%**
Experienced	167 (5.0%)	33	66	39.3%
New	132 (5.7%)	13	56	42.4%

The above table shows that a higher percentage of initial refusals were taken by the decentral Interviewers (5.3% compared to 2.1% for the central Interviewers), and that the centralized staff's refusal conversion rate [(RC IWs) / (INIT REFUSALS)] is considerably higher (54.2% for the centralized staff compared to 39.9% for the decentralized staff). When considering these figures, it is important to note that the majority of cases known to be difficult were assigned to decentralized Interviewers. Please also note that we have not included outstanding and non-sample cases in the chart, which may later increase the decentral refusal conversion rate.

A comparison of experienced and new Interviewers shows a difference between the centralized and decentralized staffs. In the centralized setting, experienced Interviewers took a higher percentage of initial refusals (due, in part, to being assigned the more difficult cases) but had considerably higher refusal conversion success (61.2% compared to 39.1% for the new Interviewers). In the decentralized setting, experienced Interviewers took a lower percentage of initial refusals while the new Interviewers had a slightly higher refusal conversion rate (42.4% compared to 39.3 % for experienced Interviewers).

Figures 6 and 7 chart the **weekly** refusal conversion rates for central and decentral staffs and also experienced and new staffs (central and decentral combined). These charts show dramatic jumps for all staffs across the production period, with no clear patterns occuring.

Behavior Coding Results

Due to a lack of parallel quality control systems in the central and decentral settings, it was decided to conduct a special behavior coding experiement to quantitatively measure both Interviewer and Respondent behaviors. Behavior coding involves listening to taped interviews while using a coding scheme to reflect standard and non-standard behaviors. For this experiment, an equal number of taped interviewers from each mode were used (70), with an equal proportion of cases completed early and later in the production period. No more than two tapes from any one Interviewer were used. Three coders completed all of the coding, listening to the Coverscreens through Section G for each interview.

As mentioned above, both Interviewer and Respondent behaviors were coded at each question. The Interviewer behaviors that we examine in this report are:

- Incomplete question reading (INCOMPLETE)
- Addition of words to the question (ADD WORDS)
- Deletion of words from the question (DEL WORDS)
- Emphasis error in the reading of the question (EMPHASIS)
- Use of a non-standard probe (NON-STD PROBE)
- Failure to probe when required (FAIL PROBE)

FIGURE 6. CENTRALIZED/DECENTRALIZED REFUSAL CONVERSION RATES

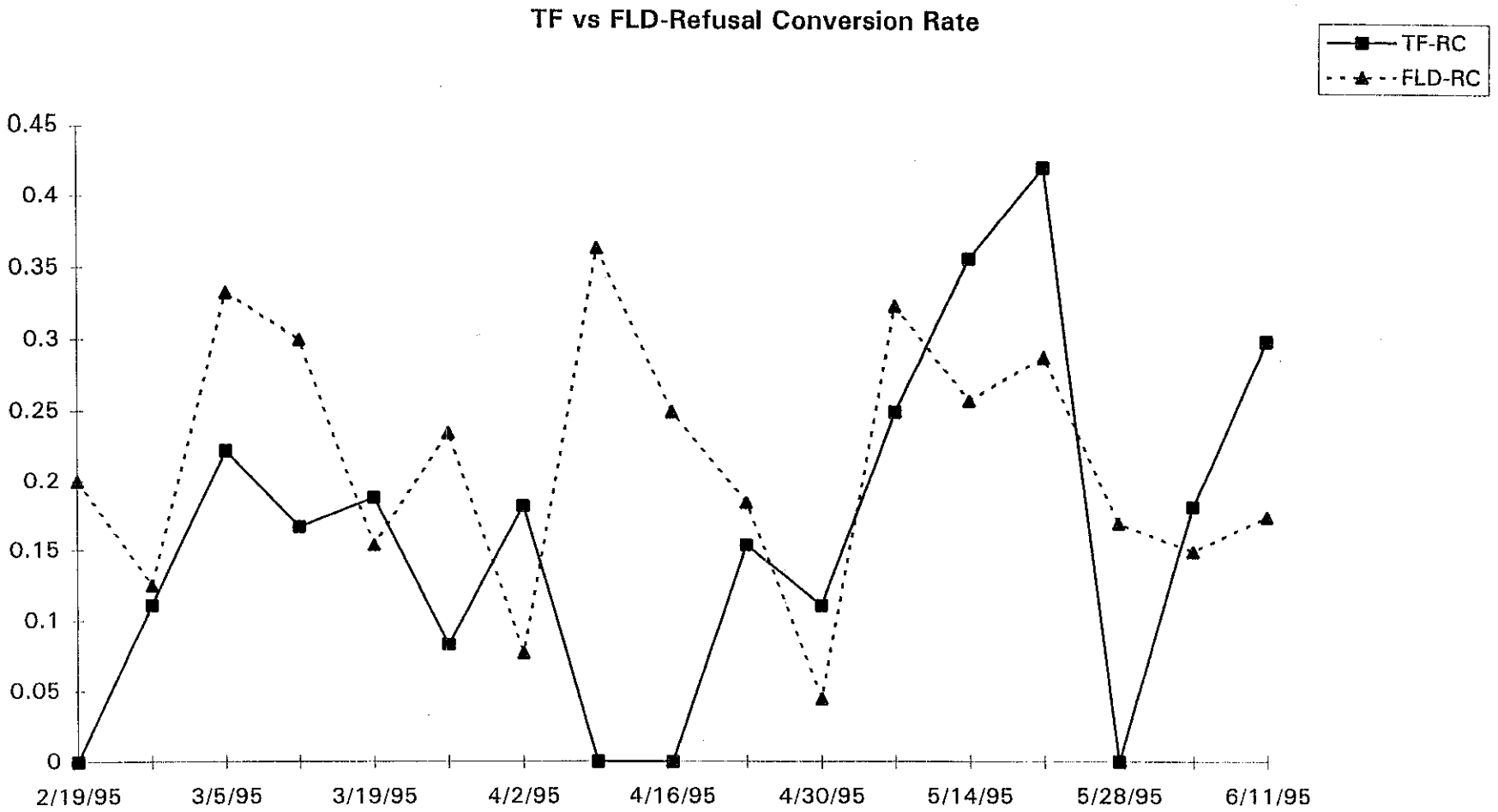
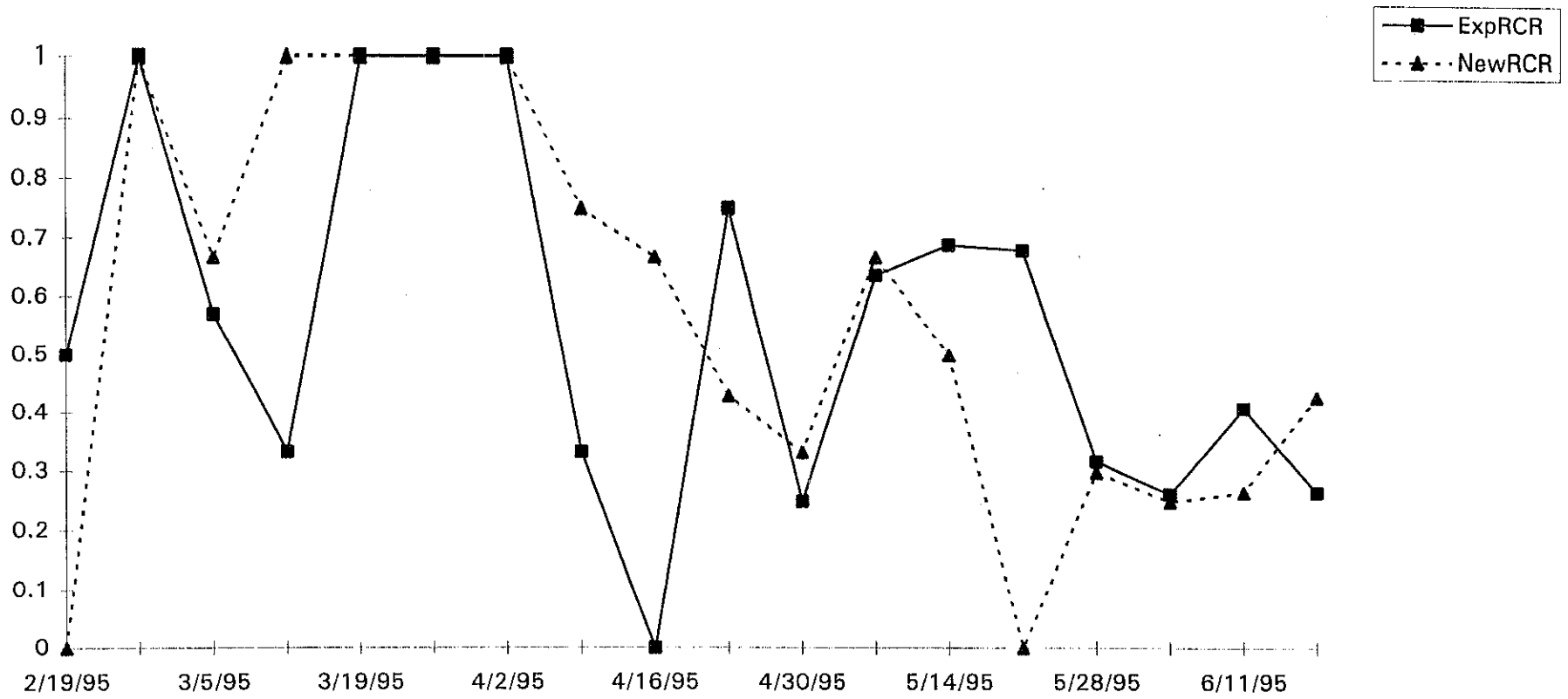

TF vs FLD-Refusal Conversion Rate
TF-RC
FLD-RC
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
2/19/95
3/5/95
3/19/95
4/2/95
4/16/95
4/30/95
5/14/95
5/28/95
6/11/95

FIGURE 7. EXPERIENCED/NEW INTERVIEWERS REFUSAL CONVERSION RATES

EXPERIENCED vs NEW IWER Refusal Conversion Rate



The Respondent behaviors that we examine in this report are:

- Refusal to answer question (REFUSE)
- "Don't know" response given (DONT KNOW)
- Answering outside of the response frame (OUTSIDE R-FRAME)
- Giving more than one answer (MULTIPLE ANS)
- Requesting a definition (DEF REQ)
- Requesting a repeat of the question (RQ REQ)
- Interrupting the Interviewer's question reading (INTERRUPT)

Tables 10 and 11 show the frequency of each of these behaviors in both the centralized and decentralized interviews. The actual count of each code and the percentage of that count compared to the number of overall observations is given. Note that the total number of occurances is higher for the centralized interviews (meaning simply that more codes were entered for the centralized cases, including codes for standard behaviors). Overall, the occurance of errors is extremely low in both modes.

For all non-standard Interviewer behavior codes, the frequency of occurance is higher in the decentralized interviews. Most notably in the use of non-standard probes, the addition of words to questions, and incomplete question readings. On the PSID, the types of errors that are probably most important are those related to probing.

The Respondent behavior codes show much more similarity between the central and decentral interviews, but with slightly higher occurances of answering outside of the response frame and interrupting Interviewers in the decentral interviews.

Tables 10 and 11 give a summary of behaviors occuring in all cases and all questions for each mode; please see Appendix 2 for the full behavior coding results, including a breakdown of behaviors by key questions.

Table 10. Behavior Coding Results--Interviewer Behaviors in Central and Decentral Interviews

MODE	# OF OBS.	INTERVIEWER BEHAVIOR CODES (% of Observations)					
		Incomplete	Add Words	Del Words	Emphasis	Non-Std Probe	Fail Probe
Central	7201	37 (0.5%)	67 (1.0%)	33 (0.5%)	0 (0.0%)	125 (1.7%)	158 (2.2%)
Decentral	6475	204 (3.2%)	272 (4.2%)	173 (2.7%)	3 (0.0%)	370 (5.7%)	249 (3.8%)

Table 11. Behavior Coding Results--Respondent Behaviors in Central and Decentral Interviews

MODE	# OF OBS.	RESPONDENT BEHAVIOR CODES (% of Observations)						
		Refuse	Don't Know	Outside Frame	Mult. Ans.	Def. Req.	Rep. Q. Req.	Interrupt
Central	7201	6 (0.0%)	94 (1.3%)	300 (4.2%)	2 (0.0%)	18 (0.2%)	284 (3.9%)	71 (1.0%)
Decentral	6475	6 (0.1%)	128 (1.9%)	437 (6.7%)	4 (0.0%)	18 (0.3%)	198 (3.1%)	170 (2.6%)

Relative Error Rate for Major/Minor Conceptual Errors

In addition to the quality control measures carried out by Supervisors and Team Leaders in the central and decentral locations, the PSID project staff also listened to a sampling of taped interviews, often combining the review of audio tapes with a review of the actual data entered into cases by the Interviewers. As part of this comparison report, the PSID project staff member responsible for quality control summarized the types and frequencies of errors occuring in interviews completed by centralized and decentralized Interviewers. (See Appendix 3 for a copy of the full report.)

Each case reviewed was rated on a scale of 1 to 4: "1" being used for major problems that potentially would require a missing data callback; "2" being used for major problems with primary data that could be resolved with extensive editing, but without a callback; "3" being used for cases with one or more minor problems that involved few data elements that could easily be edited; and "4" being used to indicate no problems or very minor problems that could be fixed in the edited databases. Table 12 shows the percentage of each error rating found in both central and decentral cases.

Table 12. Central and Decentral Error Rating Percentages

MODE	ERROR RATING			
	1 Callback	2 Major	3 Minor	4 No Prob.
Central	<1%	24%	30%	46%
Decentral	<2%	26%	28%	45%

As shown in the table above, there are no significant differences in the overall percentage error rate profiles. So, the types of error were examined more closely, to see if there are differences in which kinds of major errors are committed (see Table 13). Several common major problems were targetted and the frequency of occurances in the central and decentral locations was examined. For this analysis, only cases with error ratings of "1" or "2" were included.

Table 13. Specific Error Comparisons for Central and Decentral Interviews

PROBLEM DESCRIPTION		MODE (% OF OCCURANCES)	
		Central	Decentral
A.	FU/HU Composition Errors	4%	4%
B.	Relationship to Head Errors	2%	2%
C.	Review of Work Weeks Errors	13%	12%
D.	Main vs. Extra Job Errors	4%	5%
E.	Report of Income Errors	8%	12%
F.	Keying/Encoding Errors	13%	18%
G.	Occupation/Industry	5%	11%
H.	Marginal Note Omitted/Incorrect	2%	5%

For specific major problems A-D, there appears to be no significant difference in the percentage of cases with the error. For problems E-H, though, there appear to be either slight or significantly higher error percentages in the decentral interviews.

The "report of income" errors would include things like missing income from a reported job, recording income for the wrong individual, and double-counting income. "Keying/encoding" errors would include entering an incorrect numeric value, an incorrect response code, or incorrect/misleading responses to open-ended questions. The "occupation/industry" errors are more of a conceptual nature and include Interviewers not probing to obtain sufficiently detailed information (or failing to record all information given by the Respondent). Again, see Appendix 3 for a full copy of the report.

Effectiveness of Persuasion Letter Program

Members of the PSID project staff were responsible for writing all persuasion letters to reluctant Respondents. Due to the physical differences of the central and decentral modes, the methods for Interviewers to request persuasion letters was necessarily slightly different. In the centralized setting, Interviewers completed a paper Letter Request form, which was reviewed by a Supervisor before being forwarded to the letter writing staff. Decentrally, Interviewers gave all letter request information to Supervisors or Team Leaders (over the

telephone), then the Supervisor or Team Leader used an electronic template to send the request to the Field Office (from which it was forwarded to the letter writing staff). The letter writing staff maintained an electronic log of all letter requests, including the Interviewer ID#, the date of the request, the date the letter was sent, and the type of letter sent. For the majority of the production period, letters were sent out within three days of receipt of the request (near the end of the production period, the turn-around time increased to as much as 10 days). Letters other than persuasion letters were requested and sent to Respondents (800#, sympathy, more information), but only persuasion letters were included in this analysis.

Table 14 shows, for both central and decentral staffs, the number of persuasion letters requested and sent, the number of those cases still outstanding, the number of cases not called since the letter was sent, the number of cases resulting in successfully completed interviews, and the number of cases finalized as non-interviews (the majority of these as final refusals). One major concern during the production period was that Interviewers were not calling refusal conversion cases in a timely manner once the letter was sent (therefore minimizing the effectiveness of the letter). This analysis attempts to compare the speed and effectiveness of follow-up between the centralized and decentralized interviewing (and Supervisory) staffs. The data were compiled on 30 June, so the transfer of all outstanding cases to the decentralized staff is not a factor.

There are no overwhelming differences between the two interviewing staffs. Both staffs had approximately 50% of the letter cases still outstanding at the time of the analysis, with no significant difference in the speed of follow-up (once the ineligible cases are excluded). The decentral Interviewers were able to successfully complete a slightly higher percentage of the interviews and also finalized a higher percentage.

Table 14
1995 PSID
SUMMARY OF PERSUASION LETTER REPORT
(run 6/30)

MODE	# OF LETTERS SENT	# OF CASES OUTSTANDING (% OF # SENT)	# OF CASES NOT CALLED SINCE LETTER SENT (% OF # SENT)	# OF CASES COMPLETED (% OF # SENT)	# OF CASES FINALIZED AS NON-IWs (% OF # SENT)
CENTRAL	91	53 (58%)	19* (21%)	18 (20%)	20 (22%)
DECENT.	117	48 (41%)	14** (12%)	33 (28%)	36 (31%)
TOTAL	**208**	**101 (49%)**	**33 (16%)**	**51 (25%)**	**56 (27%)**

* 9 letters had been sent within the last two weeks, so were not yet eligible for calling. Excluding these cases, the cell drops to 11%.

** 3 letters had been sent within the last two weeks, so were not yet eligible for calling. Excluding these cases, the cell drop to 9%.

V. APPENDICES

1. **Central and Decentral Mode Approaches to '95 PSID Data Collection**

2. **Behavior Coding Report**

3. **Comparison of Quality Review Data**

Central and Decentral Approaches to '95 PSID Data Collection

Operational Activity	Purpose	Approach with Field Staff	Approach with Facility Staff
Interviewer Staffing	Recruitment	Written applications screened followed by phone screen followed by FtF interviews and reference checks	Written applications screened followed by phone screen followed by FtF interviews and reference checks
	Compensation	Determined by regional market - Avg = 9.07 with $1.00 Spanish Lang differential	Starting pay at $6.00/hr - Avg = 7.26 with $1.00 Spanish Lang differential
	Retention	Average is about 3 yrs	Average is about 1 yr
Interviewer Training	Initial Training	Five day training program covering GIT, Study Specific, and CAPI	Five day training program covering GIT, Study Specific, and CATI
	Initial Practice	Training session mock interview with immediate Supervisor feedback followed by taped Home Area practice interview with follow-up sup feedback	Monitored practice shift interview with immediate Supervisor feedback followed by live monitoring of first production interviews and sup feedback
	Performance Improvement	Sup holds weekly phone sessions with inwers to discuss adherence to performance contracts including response rate, data recording accuracy, interviewing techniques, and efficiency	Sup holds weekly meeting to discuss adherence to performance contracts including response rate, data recording accuracy, interviewing techniques, and efficiency
	New Skills Development (eg Refusal Conversion Skills)	Conference Calls and Individual Sessions	Workshops, Shift Meetings, and Individual Sessions

Operational Activity	Purpose	Approach with Field Staff	Approach with Facility Staff
Staff Communication	Direction on Original Study Procedures	Procedures Manuals: Interviewer Manual PSID Study Guide CAPI Manual	Procedures Manuals: Interviewer Manual PSID Study Guide CATI Manual
	Updates to Procedures	Conference Calls from Sups/TLs to Inwers	Shift Meetings from Sups/TLs to Inwers
		E-Mail Memos and written Memos	Written Memos
	Debriefings	Conference Calls	Face-to-Face Meetings
	Individual Inwer Progress	Inwer & Sup/TL review of CSMS Report and discuss during weekly phone call	Inwer review of Feedback memo and Sup/TL review of Control Report with joint discussion during weekly meeting
	Interviewer Performance Evaluation	Annual written and verbal (phone) evaluation	Six month written and verbal FtF evaluation
Quality Control	Data Recording	Spot check of all interviews and full check of systematically selected completed interviews	Spot check of all interviews and full check of systematically selected completed interviews
	Interviewing Techniques	Inwer self evaluation and Supervisor review of systematically selected taped interviews	Live monitoring, inwer self evaluation and Supervisor review of systmatically selected taped interviews
	Verification	Verification calls for systematically selected completed interviews	Live monitoring and constant supervisor presence in interviewing environment

Operational Activity	Purpose	Approach with Field Staff	Approach with Facility Staff
Production and Sample Management Control	Daily Production	CSMS update on sample dispositions	Both Control File and SMS update on sample dispositions
	Weekly Reporting	National, regional, primary area, and interviewer summary of production and cost	Facility, supervisor group, and interviewer summary of production and cost
	Case Disposition	CSMS central database is updated with location and disposition info on all cases with each transmission then reconciled against logged coversheet data base	SMS is updated with disposition info then reconciled against logged coversheet data base. Also, Control central database is updated with location and disposition info on all cases each night then reconciled against logged coversheet data base

		Number of Cases	Incomplete SUM	Add Words SUM	Del. Words SUM	Emphasis SUM	Non-Std Probe SUM	Fail Probe SUM	Refuse SUM	Dont Know SUM	Outside R-Frame SUM
Total Sample	LOCATION										
	1	7201	37	67	33	0	125	158	6	94	300
	2	6475	204	272	173	3	370	249	6	128	437
QUEST											
A19	1	64	0	0	0	0	1	0	0	0	6
	2	59	1	1	1	0	4	2	0	0	2
B10	1	44	0	0	0	0	0	4	0	0	5
	2	41	0	2	2	0	4	18	0	0	5
B41	1	12	0	0	0	0	0	1	0	0	1
	2	9	0	0	0	0	0	2	0	0	0
B41b	1	14	0	0	0	0	1	2	0	1	0
	2	9	0	1	0	0	1	5	0	0	1
B78	1	46	0	0	0	0	8	7	0	0	6
	2	41	0	2	0	0	10	11	0	1	12
B9	1	44	0	1	1	0	0	2	0	0	3
	2	40	1	0	0	0	0	9	0	0	1
C11	1	2	0	0	0	0	0	0	0	0	0
	2	4	0	1	0	0	0	4	0	0	1
C33	1	1	0	0	0	0	0	0	0	0	0
	2	3	2	1	0	0	0	1	0	0	0
C33b	1	1	0	0	0	0	0	0	0	0	0
	2	2	0	1	1	0	0	1	0	0	1
C70	1	2	0	0	0	0	1	0	0	1	0
	2	3	0	2	0	0	2	1	0	0	1

(CONTINUED)

		Number of Cases	Incomplete SUM	Add Words SUM	Del. Words SUM	Emphasis SUM	Non-Std Probe SUM	Fail Probe SUM	Refuse SUM	Dont Know SUM	Outside R-Frame SUM
QUEST	LOCATION										
C9	1	2	0	0	0	0	0	0	0	0	0

	2	4	0	0	0	0	0	1	0	0	0
D10	1	24	0	0	0	0	0	2	0	0	5
	2	16	0	2	1	1	1	6	0	0	3
D41	1	1	0	0	0	0	0	0	0	0	0
	2	4	0	0	0	0	1	3	0	0	3
D41b	1	1	0	0	0	0	0	0	0	0	1
	2	4	0	0	0	0	0	1	0	0	0
D78	1	23	0	0	0	0	1	3	0	0	2
	2	17	0	0	0	0	1	5	0	0	3
D9	1	24	0	0	0	0	0	2	0	0	2
	2	17	0	2	1	0	1	7	0	0	3
E11	1	2	0	0	0	0	0	0	1	0	1
	2	1	0	0	0	0	0	0	0	0	0
E70	1	2	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	0
E9	1	2	0	0	0	0	0	0	0	0	0
	2	1	0	0	0	0	0	0	0	0	0
G12	1	61	0	1	1	0	0	2	0	0	1
	2	58	2	3	2	0	2	0	0	0	4
OTHER	1	6767	35	64	30	0	113	131	5	92	266
	2	6082	197	252	163	2	343	170	6	127	397
X1	1	62	2	1	1	0	0	2	0	0	1

(CONTINUED)

		Number of Cases	Incomp-lete	Add Words	Del. Words	Emphas-is	Non-Std Probe	Fail Probe	Refuse	Dont Know	Outside R-Frame
			SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM	SUM
QUEST	LOCATION										
X1	2	59	1	2	2	0	0	2	0	0	0

(CONTINUED)

		Multip-le Ans	Def. Req.	Rep. Q Req.	Interr-upt
		SUM	SUM	SUM	SUM
Total	LOCATION				

Sample	1	2	18	284	71
	2	4	18	198	170
QUEST					
A19	1	0	0	3	0
	2	0	0	7	2
B10	1	0	1	3	0
	2	0	2	2	0
B41	1	0	0	0	0
	2	0	0	0	0
B41b	1	0	0	0	0
	2	0	0	0	0
B78	1	0	2	2	0
	2	0	0	3	0
B9	1	0	0	0	0
	2	0	0	1	1
C11	1	0	0	0	0
	2	0	0	0	0
C33	1	0	0	0	0
	2	0	0	0	1
C33b	1	0	0	0	0
	2	0	0	0	0
C70	1	0	0	0	0
	2	0	0	0	0

		Multiple Ans	Def. Req.	Rep. Q Req.	Interrupt
		SUM	SUM	SUM	SUM
QUEST	LOCATION				
C9	1	0	0	0	0
	2	0	0	0	0
D10	1	0	0	0	0
	2	0	0	1	0
D41	1	0	0	0	0
	2	0	0	0	0

D41b	1	0	0	0	0
	2	0	0	0	0
D78	1	0	0	0	0
	2	0	0	2	0
D9	1	0	0	0	0
	2	0	0	0	0
E11	1	0	0	0	0
	2	0	0	0	0
E70	1	0	0	0	0
	2	0	0	0	0
E9	1	0	0	0	0
	2	0	0	0	0
G12	1	0	0	6	0
	2	0	0	4	1
OTHER	1	2	15	269	71
	2	4	16	178	165
X1	1	0	0	1	0

		Multiple Ans	Def. Req.	Rep. Q Req.	Interrupt
		SUM	SUM	SUM	SUM
QUEST	LOCATION				
X1	2	0	0	0	0

Comparison of Quality Review Data

Sample and Methods

A subset of the interviews already being reviewed by PSID staff for data quality were used for this comparison of Field versus TelFac interviewers. Some 240 interviews taken by TelFac and 140 by Field interviewers were selected, aiming for one of the first five taken by each interviewer and then a one in ten sample thereafter. The interviews were taken by April 30th for the most part, realizing that the error rates would be higher in these early interviews, even for experienced interviewers.

Ninety percent of all Field IWers had at least one case reviewed. Of that 90%, 85% had at least one case reviewed with the use of an audio tape. Ninety-six percent of all TelFac IWers had at least one case reviewed with audio tape.

The reviewers were given instructions, forms and training and their focus was on a set of specific problems that would potentially invalidate the IW or result in significant missing data. However, they recorded every error they observed and completed a form for every case reviewed. Each case was reviewed on AQ-CATI in Review mode from the PSID server using a copy of the datafile, with accompanying audio tape if available.

General Comparison

Each case was then rated on a scale of 1 to 4 in terms of the severity of the problem(s) found. A "1" rating was used for major problems that potentially would require sending the case back for recontact. A "2" rating was used for major problems with primary data that could be resolved with extensive editing, but without a Send Back. A "3" rating was used for cases with one or more minor problems that involved few data elements that could easily be edited in AQ-CATI and/or the EditSys. A "4" indicates no problems, major or minor, were observed or minor problems that can be easily fixed in the edited databases and left unchanged in the master copy of the data file (in AQ).

Percentages at Each Error Rating

	1 Send Back	2 Major	3 Minor	4 No Problem
Field	<2	26	28	45
TelFac	<1	24	30	46

Specific Error Comparisons

I saw no significant differences in the overall percentage error rate profiles above. So I went back to the reviews for some qualitative analyses to see if there are differences in which kinds of major errors are committed. I picked several major problems that the reviewers were targeting and compared the percentage of reviewed cases with that particular problem. I also looked at some more general errors (e.g., keying/typos). For the following analyses, I used only cases with ratings of 1 or 2; counting a case if it had at least one such error recorded.

		Field%	TelFac%
A.	FU/HU Composition Errors (Incorrect Move-In or Out, etc.)	4	4
B.	Relationship to Head Errors (Wrong Head/Wife/"WIFE")	2	2
C.	Review of Work Weeks (Not accounting for all weeks)	12	13
D.	Main vs. Extra Job	5	4
* E.	Report of Income (Missing or in wrong place)	12	8
* F.	Keying/Encoding Errors (Not matching response)	18	13
* G.	Occupation/Industry (failure to probe or record)	11	5
* H.	Marginal Note Omitted/Wrong	5	2

For specific major problems A-D, there appears no significant difference in the percentages of cases with this error. The Report of Income, however, appears significantly (*) better for TelFac. Errors here would include things like missing income from a reported job, recording income for the wrong individual, and double-counting income.

The three general items (F-H below the dashed line) all show significantly lower error rates for TelFac. The keying/encoding errors and marginal note omission can be disasterous, since we cannot run consistency checks that will detect all of them. Differences in errors on Occ/Ind screens and the recording of marginal notes also favor the Telfac. My theory is these are due to a combination of greater keyboard facility and training emphasis on these areas in the TelFac. Keyboard facility, not specifically measured for this comparison, is based on my and others' observations of IWers during training and corroborated by supervisors and Team Leaders. I assume it is related to computer experience and perhaps Field Section has some quantitative data on this.

Failure to probe Occ/Ind is more a conceptual error; interviewing skill and experience usually make a difference here. But the assumption that the Field IWers would do better on this item was not born out. I believe that the increased training and emphasis on thorough Occ/Ind probing in the TelFac during the last two years made the difference. In fact, it was because the Field IWers were all experienced that their training did not include the additional emphasis.

Other Data Quality Information

The review process from which the above comparison data was sampled is only one indicator of data quality. These data were used because they allowed us to draw samples for "fair" comparisons. Two other processes initiate a review of completed cases for problems. One is the review of every Main Family IW that generates a Splitoff IW when persons with Follow Status move and set up their own household. We review all relevant screens and determine whether the Splitoff was generated correctly. An analysis of this data could provide a fair comparison since Splitoffs should occur with the same frequency in both parts of the sample.

By far the major source of reports of problems or errors in IWs is the system of ongoing quality checks that supervisors, team leaders, and IWers themselves carry out. As with the quality reviews used for our comparisons, a sample of each IWers cases are reviewed in part or whole for specific problems, sometimes with the aid of audio tape. A major difference is that feedback and spot training with the IWers (the primary purpose of this system) is more immediate, and therefore, more effective. Because the system operates differently in the Field than in the Telfac,

in terms of sampling, the number of reviewing hours, and the timeliness of feedback, a true comparison of error rates cannot be done using this data. It would be a comparison of the rates at which errors are reported, not committed. But this is also useful information, since early detection and correction is the next best thing to prevention.

Problems reported on the Project Manager Problem Sheet, the Spot-Checking Evaluation Form, or in e-mail message are reviewed by the project manager and then PSID Staff both for case-specific action as well as indications of general problems. A quick review of such reports forwarded to me shows a 2 to 1 ratio, TelFac to Field. This is not unusual given their higher number of supervisor and team leader hours spent reviewing nearly every IW in part or in whole. Furthermore, the TelFac reports include problems ranging in severity from minor to major (like those in our comparisons above) while the Field reports are mostly major problems, with a noticeably high number involving FU/HU Composition or Relationship to Head. Again, we assume the errors occur at the same or similar rates for the two groups of IWers and it is the differences in how many and what kind they detect and report that we are seeing here.

Conclusions

While the overall comparison of percentages of major and minor problems showed no significant differences, the analyses of error rates for a set of specific problem areas showed some small, but significant, differences in data quality favoring the TelFac IWers. We would need further analysis to explain these differences, but I believe they result primarily from, and therefore would be remedied by, experience and training. This is a remedy that could be applied to either group of IWers; resolving any apparent advantage of one group over the other.

I believe the advantage, given that training and experience will reduce errors in both groups, goes to the system that can review a large volume of cases, detect errors, and provide for correcting the data and giving feedback to IWers in a timely manner. As they currently operate, this would be the TelFac. Its more centralized structure provides for closer supervision, monitoring, and on-the-job assistance to the IWers, and better communication with project management and study staff. The system is not perfect and certainly some of these aspects could be achieved with the Field operation, but at present, the Telfac system better serves the data quality concerns.

Recommendations

If I were to recommend an all-or-nothing course, based solely on the above data quality information, I would have to recommend we put all our IWs in the TelFac. However, these decisions are never made based solely on one criterion and I do not recommend an all-or-nothing course. I would recommend we again split the sample putting perhaps 75-80% in the TelFac (5,000-6,000 cases was found to be a manageable number this year). The 25% going to the Field would be those who require/insist on a Field IWer and the related families in their 68ID. We all acknowledge the benefits that accrue to continuity of IWer; a general effect, not specific to Field IWers. Nevertheless, we should aim to reunite long-time respondents with their Field IWer when they have specifically requested this. In addition, we should continue our efforts to match Rs with the same IWers in succeeding years in the Telfac.

In addition, we must make every effort to thoroughly evaluate IWer performance; so that we can retain and retrain the best IWers, both Field and TelFac. Then the burden is on us to evaluate our own performance, to make our training, management, and quality review procedures more effective and responsive to the IWers' and our needs.

Tom Gonzales
PSID