

# **The 1998 Survey of Small Business Finance**

## **METHODOLOGY REPORT**

Submitted to: **The Board of Governors of the Federal Reserve**

Submitted by: **The National Opinion Research Center**

Catherine C. Haggerty

Karen H. Grigorian

Rachel Harter

Amy K. Stewart

April 2001

## **ACKNOWLEDGMENTS**

The authors wish to express their sincere thanks to all those who contributed to the production of this report. First and foremost, we acknowledge the substantive expertise of John D. Wolken, whose insights guided this study from its very beginning through the completion of this report. Marianne Bitler and Alicia Robb at the Federal Reserve Board reviewed early drafts of the report and offered useful suggestions. Ellen Hathaway carefully reviewed and edited the report.

The authors greatly appreciate the dedicated interviewing staff that collected these important data and the respondents who generously gave their time to report financial details about their businesses.

# TABLE OF CONTENTS

I.	Introduction .....	1
	1.1 Summary of Final Outcomes .....	3
	1.2 Organization of the Report .....	4
II.	Background .....	5
III.	Project Staff .....	7
	3.1 Federal Reserve Board Staff .....	7
	3.2 National Opinion Research Center Staff .....	7
IV.	Questionnaire Development and Content .....	9
	4.1 Methodology Study .....	9
	4.2 Conclusions and Recommendations from the Focus Groups and Cognitive Interviews .....	13
	4.3 Expert Consultation .....	16
	4.4 Pretests .....	16
	4.5 CATI Development .....	21
	4.6 CATI Testing .....	22
	4.7 1998 SSBF Screener .....	22
	4.8 1998 SSBF Questionnaire .....	23
	4.9 Questionnaire Length .....	25
	4.10 Questionnaire Improvements .....	25
V.	Data Collection Preparation and Interviewer Training .....	27
	5.1 Respondent Materials Development .....	27
	5.2 Toll Free Telephone Numbers .....	30
	5.3 Letter of Endorsement .....	31
	5.4 Press Release .....	31
	5.5 Electronic Mail Address .....	31
	5.6 Web Sites .....	31
	5.7 Preparation Activities for Screener Data Collection .....	32
	5.7.1 Respondent materials .....	32
	5.7.2 Training Materials .....	32
	5.7.3 Recruiting and Hiring .....	36
	5.7.4 Training .....	37
	5.7.5 Interviewer Certification .....	38
	5.7.6 Mailing Advance Respondent Packages .....	39
	5.8 Preparation Activities for Questionnaire Data Collection .....	39
	5.8.1 Respondent materials .....	39
	5.8.2 Training Materials .....	41
	5.8.3 Recruiting and Hiring .....	49
	5.8.4 Training .....	49
	5.8.5 Interviewer Certification .....	50
	5.8.6 Mailing Advance Worksheet Packages .....	51

## TABLE OF CONTENTS (CONTINUED)

VI.	Data Collection	52
	6.1 Screening	52
	6.1.1 Timing and Schedule	53
	6.1.2 Protocol Changes	53
	6.1.3 Refusal Conversion	54
	6.1.4 Locating	54
	6.1.5 Monitoring	55
	6.1.6 Interviewer Misconduct	55
	6.1.7 Receipt Control Activity	56
	6.1.8 Screening Level of Effort	57
	6.1.9 Completion Rate / Response Rate	58
	6.1.10 Eligibility Rate	58
	6.1.11 Non-response	59
	6.1.12 Non-response Follow-up	59
	6.1.13 Other basic statistics	61
	6.2 Main Data Collection	64
	6.2.1 Timing and Schedule	64
	6.2.2 Special Efforts to Increase Production	64
	6.2.3 Extended Field Period for Data Collection	68
	6.2.4 Refusal Conversion	69
	6.2.5 Locating	69
	6.2.6 Protocol for Working Minority Cases	70
	6.2.7 CATI Changes	71
	6.2.8 Monitoring	71
	6.2.9 Break-offs	72
	6.2.10 Sample Release	73
	6.2.11 Examination of Completed Cases by Interviewer	74
	6.2.12 Respondent Incentives	74
	6.2.13 Level of Effort	76
	6.2.14 Receipt Control	77
	6.2.15 Worksheets	79
	6.2.16 Completion Rate / Response Rate	79
	6.2.17 Other basic statistics	79
VII.	Data Review and Delivery	82
	7.1 Main Data Review	82
	7.1.1 Data Frequency Review	82
	7.1.2 Break-off Protocol	83
	7.1.3 Completeness Criteria	85
	7.2 Data Editing	86
	7.2.1 Identifying Cases for Editing	86
	7.2.2 The On-line Editing Process	87
	7.3 Data Cleaning	89
	7.4 Data Coding	91
	7.4.1 Race Coding	91
	7.4.2 Industry Coding	91
	7.5 Data Recoding Process	91
	7.6 Geo-coding	93
	7.7 Data Deliveries	94

## TABLE OF CONTENTS (CONTINUED)

VIII. Sampling and Weighting Methodology .....	97
8.1 Introduction .....	97
8.2 Objectives .....	97
8.3 The Frame .....	99
8.4 Basic Sample Design .....	100
8.5 Initial Stratification and Sample Selection: Stage 1 .....	101
8.6 Screening and Subsampling: Stage 2 .....	107
8.7 Sample Design Aspects of Interviewing .....	118
8.8 Overview of Weighting Adjustments and Calculations .....	122
8.9 Stage 1 Base Weights .....	122
8.10 Screening Outcomes, Eligibility, and Nonresponse .....	123
8.11 Adjusting Stage 1 Weights for Eligibility .....	127
8.12 Adjusting Stage 1 Weights for Nonresponse .....	130
8.13 Stage 2 Base Weights .....	139
8.14 Adjusting Stage 2 Weights for Eligibility .....	139
8.15 Adjusting Stage 2 Weights for Nonresponse .....	141
8.16 Weight Smoothing .....	149
8.17 Quality Control .....	151
References .....	153
Sampling Appendix – Illustration of Subsampling and Weighting .....	154

## LIST OF TABLES AND FIGURES

Table 4.1 Recruiting Sources for Methodology Study Participants . . . . .	10
Table 4.2 Types of Participating Businesses . . . . .	12
Table 4.3 Mean Number (Range) of Owners of Participating Businesses . . . . .	12
Table 4.4 Mean Number (Range) of Employees of Participating Businesses . . . . .	12
Table 4.5 Age Range of Participating Business Owners . . . . .	13
Table 4.6 Firm Size . . . . .	17
Table 4.7 Pretest Eligibility and Completion Rates . . . . .	18
Table 4.8 Screening Level of Effort . . . . .	19
Table 4.9 Interviewing Level of Effort . . . . .	19
Table 4.10 SSBF Pretest2 Reported Records Used for Complete Cases . . . . .	19
Table 5.1 Training Modules . . . . .	33
Table 5.2 Main Interviewing Training Modules . . . . .	42
Table 5.3 Training Dates, Number of Trainees, and Interviewer Attrition . . . . .	50
Table 6.1 Number of Returned Letters by Week . . . . .	56
Table 6.2 SSBF '98 Screening Level of Effort by Outcome . . . . .	57
Table 6.3 Screener Production, Hours Worked, and Minutes/Screener . . . . .	58
Table 6.4 Final Status of First and Second Half of Screener Sample . . . . .	58
Table 6.5 Number of Firms by Size of Workforce . . . . .	59
Table 6.6 SSBF Final Screening Outcome . . . . .	62
Table 6.7 Level of Effort by Minority Status . . . . .	63
Table 6.8 Number of Eligible Cases by Race/Ethnicity and Firm Type. . . . .	63
Table 6.9 Incentive Trial . . . . .	67
Table 6.10 Status of Post-Incentive as of 5/6/2000 . . . . .	68
Table 6.11 Status of Pre-Incentive as of 5/6/2000 . . . . .	68
Table 6.12 Hardcopy Materials Used to Complete Case Information . . . . .	73
Table 6.13 Release Dates and Sample Size . . . . .	74
Table 6.14 Level of Effort by Case Status and Incentive Status . . . . .	75
Table 6.15 Number of Worksheets and Other Financial Records by Week Returned . . . . .	78
Table 6.16 Sample Status for Main Questionnaire Data Collection . . . . .	79
Table 6.17 Level of Effort by Case Status . . . . .	80
Table 6.18 Level of Effort by Firm Size and Case Status . . . . .	80
Table 6.19 Level of Effort by Industry Category and Case Status . . . . .	81
Table 6.20 Level of Effort by D&B Credit Score and Case Status . . . . .	81
Table 7.1 List of Files Delivered and Delivery Dates . . . . .	96

## LIST OF TABLES AND FIGURES (CONTINUED)

Table 8.1 Derivation of Initial Sample Size .....	98
Table 8.2 Revised Sample Targets .....	99
Table 8.3 Out of Scope SIC Codes .....	100
Table 8.4 Marginal Totals From the Sampling Frame .....	102
Table 8.5 Ideal Sample Sizes for Core Sample of 38,900 .....	103
Table 8.6 Determination of Sampling Rates by Stratum for Stage 1 Screening Sample .....	104
Table 8.7 Estimated Marginal Totals After Screening the First 200 Replicates .....	109
Table 8.8 Expected Stage 2 Stratum Sizes After Screening is Complete .....	110
Table 8.9 Derivation of Stage 2 Sample Sizes .....	113
Table 8.10 Stage 2 Sample Allocation .....	114
Table 8.11 Final Status of Stage 2 By Major Stratum .....	119
Table 8.12 Pre-Interview Race/Ethnicity Compared with Updated Information .....	120
Table 8.13 Pre-Interview Size Class Compared with Updated Information .....	120
Table 8.14 Pre-Interview Census Division Compared with Updated Information .....	120
Table 8.15 Pre-Interview Organization Type Compared with Updated Information .....	121
Table 8.16 Final Accepted Cases By Updated Classification .....	121
Table 8.17 Screener Dispositions and Eligibility Status .....	124
Table 8.18 Subsampling Strata for the 5% Follow-Up Sample .....	125
Table 8.19 Eligibility Results from 5% Follow-Up .....	126
Table 8.20 Screener Eligibility Adjustment Cells .....	129
Table 8.21 Cells for Stage 1 Nonresponse Adjustments Based on Frame Variables .....	132
Table 8.22 Stage 2 Base Weight Adjustments for Replicates Used .....	139
Table 8.23 Outcome Dispositions for the Main Interview .....	140
Table 8.24 Screener Eligibility Status for Ineligible Cases in Stage 2 .....	140
Table 8.25 Credit Score Ranges .....	142
Table 8.26 Organization Type .....	142
Table 8.27 Industry Divisions .....	143
Table 8.28 Nonresponse Adjustment Cells for Stage 2 Weights .....	144
Table 8.29 Review of Cases with Largest Weights .....	150
Figure 1. Stage 1 Sample for Screening .....	154
Figure 2. Interim Stage 1 Weight Adjustments for Stage 2 Sample Allocation .....	157
Figure 3. Setup for Stage 2 Sampling Based on First Completed Replicate Results .....	157
Figure 4. Results of Raking Steps for Stage 2 Sample Allocation .....	158
Figure 5. Allocation to Replicates .....	158
Figure 6. Assigning Eligibles to Stage 2 Replicates as Stage 1 Replicates are Finalized .....	160
Figure 7. Number of Cases Per Replicate By Stratum .....	161
Figure 8. Calculation of Final Weights, Assuming All Stage 1 Replicates Are Used .....	162

## **LIST OF APPENDICES**

- Appendix A. Pretest Debriefing Minutes
- Appendix B. Questionnaire Testing Plan
- Appendix C. Screener
- Appendix D. Questionnaire
- Appendix E. FAQ Brochure
- Appendix F. Greenspan Letter
- Appendix G. Haggerty Letter for Screener
- Appendix H. Haggerty Letter for Main
- Appendix I. Greenspan Speech
- Appendix J. SSBF Previous Rounds Fact Sheet
- Appendix K. Worksheet
- Appendix L. Newspaper Article
- Appendix M. Contact Information Insert
- Appendix N. Folder
- Appendix O. NORC Brochure
- Appendix P. FRB Brochure
- Appendix Q. FRB Book
- Appendix R. SBA Endorsement
- Appendix S. Press Release
- Appendix T. Screener Training Agenda
- Appendix U. Main Training Agenda
- Appendix V. Job Aids
- Appendix W. 5% Follow-Up Questionnaire
- Appendix X. 5% Facesheet
- Appendix Y. # of Calls to Complete – Screener
- Appendix Z. Refusal Conversion Letters
- Appendix AA. Changes to CATI
- Appendix BB. Distribution of Number of Calls to Complete
- Appendix CC. Number of Returned Worksheets
- Appendix DD. List of Break Off Cases
- Appendix EE. List of Bubble Cases
- Appendix FF. List of Accepted Bubble Cases
- Appendix GG. Overview of Data Editing and Cleaning Process
- Appendix HH. File of Graphic Images
- Appendix II. 1998 U.S. Tax Forms



## I. INTRODUCTION

This report documents the methods used to conduct the 1998 Survey of Small Business Finance (SSBF), which began in November 1998 and was completed in January 2001. The target population consists of for-profit businesses with fewer than 500 employees that are either a single establishment or the headquarters of a multiple establishment company, and are not agricultural firms, financial institutions, or government entities. These firms also had to be in business during December 1998. Data were collected for fiscal year-end 1998. The 1998 SSBF is the third survey of this type to be conducted for the Board of Governors of the Federal Reserve System (FRB). Prior surveys collected data for year-end 1987 and year-end 1993. (These earlier surveys were called the National Survey of Small Business Finances (NSSBF)).

The purpose of this survey was to gather information about small business financial behavior and the use of financial services and financial service providers by these firms. The objectives of the survey were to collect information that can inform researchers and policy makers on the availability of credit to small businesses; the location of the sources of financial services; the types of financial services used, including checking accounts, savings accounts, various types of credit, credit cards, trade credit, and equity injections; as well as the firm's recent credit acquisition experiences. The survey also investigated the level of debt held by these firms and their access to credit. Additionally, the survey collected information on firm and owner demographics, as well as the firm's recent income statement and balance sheet. The survey aimed to provide a diverse and general purpose research data base on small businesses' financial relationships and behavior.

The first activity undertaken after contract award was a thorough and extensive questionnaire development period that began with the preparation for and conduct of cognitive interviews and focus groups, as part of a methodology study. The purpose of the methodology study was to uncover possible sources of error in respondent's reporting of financial data and to develop ways to improve the quality of the data through changes in the wording of the questionnaire. Recruitment of respondents and the development of protocols for the methodology study began in December 1998. Cognitive interviews and focus groups were held in January 1999; the report documenting the process and findings was completed and delivered in April 1999.

After the methodology study was completed, questionnaire development continued with two pretests, executed sequentially. In addition to informing questionnaire design, the pretests helped to test the sample drawn from Dun and Bradstreet, allowed a trial run of data collection processes and protocols, and assisted in testing the Computer Assisted Telephone Interview (CATI) instrument. The preparation for the initial pretest began in February 1999. Data collection was started and completed in March 1999. The preparation for the second pretest

began in early April 1999. The second pretest began in late-April and was completed in mid-June 1999.

Throughout and after the period when the methodology study and pretests were in progress, experts in questionnaire construction met to review the questionnaire. The experts were assisted by a Certified Public Accountant who specializes in small business accounting. This group of experts informed many questionnaire changes.

Many activities were underway in the period prior to data collection. These activities included training material preparation, respondent material preparation, the development of a website, obtaining a letter of endorsement, specifying the receipt system and process, developing mail-out protocols, and recruiting, hiring and training interviewers.

Data collection had two phases: a screening phase and an interviewing phase. A sample of about 40,000 businesses was drawn. The sample was selected so that differences in the use of credit and financial services among firms of differing sizes and minority ownership could be measured. The sample was selected from the Dun and Bradstreet master file to include the “target population” definition, including separate strata for African American-, Asian-, and Hispanic-owned firms. A stratified systematic sample for the study was selected according to specifications determined by a NORC statistician and FRB staff. The sample specifications and sample draw were tested with a sample of 1,000 businesses during the pretest phase of the study. A sample of almost 40,000 businesses was drawn for the main screening effort.

Screening data collection began in mid-June 1999 and was completed at the end of September 1999. The screening was designed to verify the name of the business owner and the mailing address of the business, screen the business for eligibility to participate, identify the business legal form or organization, record the fiscal year end date, and most importantly, to determine the ethnic and racial composition of the owners of each small business. Because information on minority ownership was not generally available on any publicly available data set, the requirement for separate minority sampling strata caused the initial screening sample to be much larger than in the prior small business surveys.

The initial contact with respondents consisted of a mailing that included two letters explaining the purpose of the survey and encouraging participation; one letter was from Alan Greenspan and the other was from the project director. The mailing also included a question and answer brochure. Within a few days of receiving the letter, businesses were called and asked to complete the screener. Interviewers first tried to complete the screener with the business owner before attempting to identify and complete the screener with a suitable proxy respondent. The length of time spent screening respondents was twice as long as expected.

Even though NORC did not expect to use all of the eligible sample, after the first half of screening was completed, and then again after the second half of screening was completed, the businesses that met the eligibility criteria were divided into replicates on the basis of the businesses minority status. Replicates were then released to interviewers, as needed, according to a sample release plan developed by project staff.

Interviewing began in mid-September 1999 and ended in early May 2000. The main questionnaire interview collected information on the following: income and expenses; assets and liabilities; loans, equity, checking accounts, credit cards, and other sources of outstanding funding; other financial services such as check clearing, credit card processing, brokerage services, trade credit, and equity injections; and the firm's recent credit acquisition experiences. Prior to calling businesses to complete the 40-minute interview, NORC mailed an information package to the business owner that included letters, brochures, and interesting facts from the prior round of data collection. At the same time the package was mailed to the business owner, a thank-you post card was sent to the proxy respondent if a proxy completed the screener in place of the owner. Within a few days of the mailing, businesses were called by telephone interviewers to answer any questions about the survey and to complete or make an appointment to complete the interview. The time required to complete questionnaire data collection was also twice as long as planned. Much of the increased time was due to an increase from prior experiences in the time and effort required to identify business owners and secure their cooperation. Additionally, cooperation rates declined dramatically from the 1993 experience.

Data preparation and delivery tasks began during the data collection period and were completed at the end of 2000. Throughout data collection, periodic deliveries of the questionnaire data were sent to the FRB. Editing and coding activities spanned the period data were collected and continued for several months after data collection ended. The final data files, code books, and data documentation were sent to the FRB by December 2000.

The procedures for producing weights were developed in collaboration with staff from the FRB. Stage 1 weights were produced after screening was completed. Stage 1 weights included base weights and an adjustment for eligibility and screener non-response. Stage 2 weights were computed in the autumn of 2000. At this point the stage 1 weights were adjusted on the basis of the outcome of additional work performed on a sub-sample of the businesses that did not respond at the screener stage. The stage 2 weights included adjustments for eligibility and non-response to the interview.

### **1.1 Summary of Final Outcomes**

At the close of data collection, 26,998 firms had been contacted and screened, from the 39,240 initially selected. Of the 26,998 firms that were screened, 19,792 met the eligibility criteria for the study. These eligible firms were divided into replicates by

minority type and replicates for all others. The initial plan was to release enough sample so that 6,000 cases were completed with a completion rate of 60 percent. Unfortunately, the 1993 experiences regarding firm eligibility and response rates were not reproduced in the 1998 survey. In particular, it was much more difficult to get firms to participate, despite many improvements in the questionnaire design, materials prepared to gain respondent cooperation, worksheet aids, and other protocols. As a result of these difficulties, both the length of the screening period and the length of the interviewing period were twice the time planned for each. Even with an extended period of data collection, response rates fell from 52 percent complete in the 1993 survey to 33 percent complete in the 1998 survey. The final sample consists of 3,561 firms with completed interviews. Complete details on the preparation for and experiences of data collection, in addition to the survey outcomes, can be found in Chapters V and VI.

## **1.2 Organization of the Report**

To help the reader best understand the survey processes as they were carried out, the chapters in this report are organized to present information in the order in which tasks were undertaken, with the exception of the sampling task, which is presented at the end of the report. Chapter II gives information on the background of the survey. The individuals that contributed to the project and a description of the roles and responsibilities of these individuals are given in Chapter III. Chapter IV includes a description of the methodology study and the recommendations that resulted from this work, details the pretests and the key outcomes of each, describes the questionnaire design process, and reviews the content of the questionnaire. Data collection preparation activities are discussed in Chapter V. Screener and questionnaire data collection processes and problems, as well as the outcomes of each, are presented in Chapter VI. Data preparation and data delivery activities, which included preparing coding, editing and data delivery specifications, the coding and data editing processes, and preliminary and final data deliveries, are documented in Chapter VII. Sample specifications and characteristics, as well as detailed weighting specifications and procedures, are found in Chapter VIII.

## II. BACKGROUND

The 1998 Survey of Small Business Finances (SSBF) marks the third time that the Federal Reserve Board (FRB) has collected information on the finances of small businesses in the United States. While data on small businesses exist from a variety of sources including the Small Business Administration, the Commerce Department, the Treasury Department, and the Labor Department, no other source provides such detailed information on the finances of small businesses. The SSBF not only includes information on small businesses' income, expenses, assets, liabilities, characteristics of the firm and firm owners, but also characterizes small businesses' financial relationships with financial service suppliers for a broad set of financial products and services.

The macroeconomic conditions in the United States in 1998 were quite different from those that existed in 1987 or 1993, the years covered by the previous surveys. The United States was enjoying the seventh year of vigorous expansion during the reference period of data collection for the 1998 SSBF. Unemployment was under five percent, the economy was growing over three percent annually, and productivity was continuing to rise, fueling real wage growth. New businesses had been forming at record rates. Interest rates had remained low, making loans and other credit attractive to businesses.

All this helped to reinforce the fact that small businesses are extremely important to the U.S. economy. According to the Small Business Administration (SBA), 99.7 percent of all employer firms have fewer than 500 employees. These firms employ more than half of the private sector workforce and account for most of the new jobs created between 1992 and 1996. Small businesses account for approximately half of the gross domestic product (GDP). On the other hand, starting and maintaining a small business is difficult, with nearly one million businesses terminating, failing, or going bankrupt each year. In part, the SSBF is conducted to help researchers better understand how such changes in the financial marketplace affect the acquisition and use of financial services by small businesses.

Since the 1993 survey consolidation has continued in the banking industry; insufficient data are currently available to understand the impact of this consolidation on small business lending. In 1997, a report from a conference sponsored by the SBA on "The Impact of Bank Mergers and Acquisitions on Small Business Lending" showed that, for the most part, lending to small businesses had not declined, although the effects of bank mergers and acquisitions varied depending on the nature of the merger or acquisition and the institutions involved. Also, in those bank merger situations where small business lending may have declined due to a merger, this decline may have been partially offset by increased lending to small businesses by non-banking institutions.

Data from the 1998 SSBF will help researchers understand the impact of banking mergers and consolidations, as well as the rise in interstate banking. This data set, together with data collected for 1987 and for 1993, will permit researchers to identify trends in the use of local and nonlocal banks and nonbank institutions, and identify any changes in the types of financial services used by small businesses, such as in credit card and trade credit use. The survey contains an expanded set of information on small businesses' recent borrowing experiences, which can be used to identify segments of the small business sector that have the most difficulty obtaining credit. The data from three points in time, taken together, will lead to analyses such as in Cole and Wolken (1996) and will be more informative.

### **III. PROJECT STAFF**

#### **3.1 Federal Reserve Board Staff**

John Wolken, a Senior Economist at the Federal Reserve Board, was the Contracting Officer's Technical Representative (COTR) for the SSBF project and in that role was responsible for all technical aspects of the project. Dr. Wolken was initially assisted by economist Patrick Lampani who left the project during the pretesting phase and was replaced by Marianne Bitler, also an economist. Economist, Alicia Robb, and research assistants, Emily Rosenberg, Gretchen Christianson, Doug Rohde and Courtney Carter, also, worked on the project. Nicole Meleney, an information systems analyst, worked from the start of the project through the middle of data collection. Michael Hein was the programmer assigned to the SSBF project at the FRB; Gary McQuown filled the position after Mr. Hein left the FRB. Lucy Lucas, a contracting specialist, assisted Dr. Wolken with contracting issues during the period of performance of the contract. Dr. Wolken and his staff were active participants in the project throughout the design, execution, and data delivery phases of the project.

#### **3.2 National Opinion Research Center Staff**

The FRB contracted with the National Opinion Research Center (NORC) to conduct the SSBF project. The SSBF project was led and supported by staff from various departments within NORC. The project was led by Catherine Haggerty, the Project Director, and Karen Grigorian, the Associate Project Director; Ms. Haggerty and Ms. Grigorian shared responsibility for managing all aspects of the project. Amy Stewart was responsible for materials and systems development. Ms. Stewart was assisted by Stephanie Bzdusek during the interviewer training phase of the project. Heather Kwasigroch and Amy Kulekowskis were responsible for tracking the project expenses.

Bernard Dugoni and Lisa Lee, staff from NORC's statistics and methodology group, designed the focus group and cognitive interview protocols and documented the results of the methodology study. Diana Jergovic, from NORC's survey group, lead the focus group discussions.

Rachel Harter was the project's sampling statistician and was also responsible for computing the weights. Rachel was assisted by Steven Pedlow and Javier Porras.

Computing support was led by Michael Crifassi. Phillip Panczuk was the questionnaire programmer and Gregory Wilson programmed the Telephone Number Management System (TNMS).

Rochelle Leslie was the Telephone Center Coordinator. Rochelle was assisted by Margaret Buckhalter, AngerMarie Cheatham, Janel Temple and Michael Rhea.

Angela Herrmann led the mailout and receipt effort with assistance from Crystal Williams. Lyanette Scott, Winifred Mason and Betty Williams were responsible for data editing and coding.

Michael Pergamit, a labor economist and NORC Research Vice President provided corporate oversight in addition to advising on the design of the questionnaire. Norman Bradburn, Senior Vice President for Research, made a significant contribution to the design of the questionnaire by reviewing several iterative drafts of the instrument throughout the seven-month period of design effort and by attending several meetings to review and recommend improvements to the questionnaire.

Steven Knable, Assistant Treasurer and Vice President, Contracts and Grants, and James Casey, Contracts Administrator, provided assistance with contracting issues throughout the period of performance.

Over three summers, the project was assisted by students from Maastricht University in The Netherlands who worked in Chicago as NORC summer interns. They were Wilko Letterie, who helped during the proposal development phase; Mindel van de Laar, who helped during the screening phase; and Man Yi Chan, who helped during the data preparation and delivery phase.

Finally, NORC engaged the services of a small business accountant, Charles Smith. A Certified Public Accountant from Smith and Associates, an accounting firm that specializes in accounting for small businesses, Mr. Smith provided helpful technical guidance throughout the project.



## **IV. QUESTIONNAIRE DEVELOPMENT AND CONTENT**

The SSBF (formerly known as the National Survey of Small Business Finances) was previously conducted for year-end 1987 and year-end 1993. The questionnaires used in the previous surveys were used as the basis for the 1998 SSBF. Two instruments were prepared for 1998. The first was a short screening questionnaire, referred to as the “screener,” that was used to establish contact with the firm, verify contact information (name and address), establish firm eligibility, and collect information on the minority ownership of the business. The second instrument was the main interview questionnaire, referred to as the “questionnaire.” A significant effort was made to improve on the previous instruments and adjust the content to reflect issues that have arisen in the intervening years.

Questionnaire development for the 1998 SSBF screener and questionnaire began even before the contract award in November 1998. Shortly after award of the contract, the FRB provided NORC with a modified version of the 1993 SSBF questionnaire, at which point staff from NORC, staff from the FRB, and the accounting consultant worked together to design the screener and questionnaire for both pretests and ultimately the screener and questionnaire used for the main data collection.

NORC and FRB staff engaged in questionnaire design activities over an eight-month period, between December 1998 and July 1999. These activities began with a methodology study consisting of cognitive interviews and focus groups which were followed by two pretests. The first pretest used a computer assisted telephone interview (CATI) screener and a hard-copy questionnaire; the second pretest used CATI for both screening and questionnaire data collection. Before and after the methodology study and each of the two pretests, NORC design experts and operational staff reviewed the questionnaire wording and format, and recommended changes.

This chapter first reviews the methodology study and the two pretests and then details the final phase of questionnaire development. The chapter ends with a description of the 1998 SSBF questionnaire.

### **4.1 Methodology Study**

NORC conducted a methodological study consisting of focus groups and cognitive interviews to improve the 1998 SSBF survey instrument. The goal was to uncover sources of error in respondents’ reporting of financial data and to develop ways to improve the data through changes in the wording of questionnaire items, including lead-ins and interviewer prompts, and revisions to the worksheet format.

Preparation for the methodology study began in December 1998 with a review of the questionnaire and development of a cognitive interview and focus group protocol. Items in the SSBF questionnaire requiring dollar amount responses were selected for testing in the focus groups and cognitive interviews. In addition, a set of questions about the number of employees in a firm was tested. Follow-up questions (probes) were constructed for each item. These probes were designed to explore the processes respondents go through in answering the questions and to understand any problems they encountered.

Beginning in early December 1998, NORC project staff recruited small business owners and small business accountants in the Chicago area to participate in focus groups and cognitive interviews. NORC staff identified potential participants through six sources as listed in the table 4.1. Table 4.1 shows the number of owners contacted that agreed to participate and the number who actually participated. As the table illustrates, our success in finding participants varied widely across these sources. More than half of the business owners who participated were personal contacts of NORC staff or were NORC vendors. Cognitive interview and focus group participants were offered \$100 with the exception of difficult-to-schedule Hispanic and Asian participants, who were offered up to \$200.

**Table 4.1 Recruiting Sources for Methodology Study Participants**

Participant Source	Focus Group			Cognitive Interview		
	Number Recruited	Number Participated	Percent Participated	Number Recruited	Number Participated	Percent Participated
Personal Contact	16	12	75	14	12	86
NORC Vendor List	7	4	57	NA	NA	NA
Neighborhood Directory	NA	NA	NA	3	1	33
Govcom A8 Database	8	2	25	4	3	75
Yellow Pages	NA	NA	NA	3	3	100
Other Internet Sources	3	1	33	2	2	100
<b>Total</b>	<b>34</b>	<b>19</b>	<b>56</b>	<b>26</b>	<b>21</b>	<b>81</b>

We were successful in recruiting African Americans and Caucasians equally across all sources. It was difficult, however, to recruit Hispanic and Asian participants. All three Hispanic participants were recruited through non-personal sources. One Asian respondent was recruited through the Yellow Pages, while the other was a personal contact of NORC staff.

Overall, the reasons for nonparticipation fell into two categories: level of effort required and sensitivity of information. Particularly among CPAs, level of effort was a primary drawback for participating. A polite, "It's the worst time of year to be contacting me," was a common reply from CPAs feeling the pressure of tax season. A record snowstorm in Chicago during this period also affected participation.

We found that offering the option of completing the interview over the phone or in-person at their place of business increased the likelihood of participation. For example, both Asian business owners who participated in cognitive interviews only did so because NORC project staff came to their establishments. Though the methodology study was an overall success, we believed the challenges faced by the recruiting effort for the methodology study would mirror those we would experience in both the pretests and main study.

NORC conducted four focus groups during the period from January 9 to January 13, 1999, and cognitive interviews were conducted between January 5 and 27, 1999. In all, 19 small business owners participated in focus groups of 4 to 6 persons, and 21 small business owners and accountants who specialize in small businesses completed cognitive interviews; five interviews were conducted by telephone and 16 were conducted in person. Of the cognitive interviews, 15 respondents were the owners of small businesses, three were business owners' accountants, and three were independent accountants. Table 4.2 shows the types of businesses that participated in the focus groups and cognitive interviews.

Prior to both the focus groups and cognitive interviews, the participants were asked to complete a short questionnaire that provided general information about themselves and their businesses. All of the focus group respondents completed the questionnaire, but only 11 of the 21 cognitive interview respondents did so. Ten of the focus group participants were male and 9 were female, while 11 of the cognitive interview respondents were male and 8 were female. Tables 4.3, 4.4 and 4.5 reflect the mean number of owners and the mean number of employees of the participating businesses, and the age of the participating business owners, respectively.

**Table 4.2 Types of Participating Businesses**

Business Type	Focus Group Participants	Cognitive Interview Participants
Sole Proprietors	Social Services Maintenance Retirement Planning Stock Broker Graphic Design Attorney	Restaurant Home Health Art Gallery Graphic Design Retail Apparel Dentist
Partnerships	Law Office Law Office Decorative Painting American Folk Art	Dry Cleaner
C-Corporations	Communication Computers Manufacturing/Sales Manufacture/Distribution Moving Company	Landscaping Real Estate Accounting
S-Corporations	Restaurant Dentist Food Service/ Janitor Supply Translation Service	Bar/Restaurant Travel Agency Engineering Public Relations
Certified Public Accountants	<i>Not Applicable</i>	Dentist Accounting Unknown Attorney

**Table 4.3 Mean Number (Range) of Owners of Participating Businesses**

Sole Proprietors		Partnerships		C-Corporations		S-Corporations		CPAs
Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Cognitive Interview
1 (1-1)	1.5 (1-2)	15.5 (2-45)	Missing	3.8 (1-9)	1.3 (1-2)	2.0 (1-4)	2.3 (1-3)	1.3 (1-2)

**Table 4.4 Mean Number (Range) of Employees of Participating Businesses**

Sole Proprietors		Partnerships		C-Corporations		S-Corporations		CPAs
Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Cognitive Interview
1 (1-2)	1 (1-7)	21.0 (6-38)	Missing	10.7 (3-24)	6 (2-10)	14.7 (9-20)	22.3 (4-50)	4.3 (1-50)

**Table 4.5 Age Range of Participating Business Owners**

Sole Proprietors		Partnerships		C-Corporations		S-Corporations		CPAs
Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Focus Group	Cognitive Interview	Cognitive Interview
36 - 52	36 - 49	35 - 42	Missing	45 - 49	38 - 56	41 - 52	44 - 55	32 - 70

Respondents were asked to complete a worksheet prior to the interview and to bring their financial records (such as tax forms and bank statements) with them to the interview. Twelve of the focus group respondents and five cognitive interview respondents completed the worksheet. The reported time spent completing the worksheet ranged from 15 minutes to 2 hours. The average was 33.8 minutes.

#### **4.2 Conclusions and Recommendations from the Focus Groups and Cognitive Interviews**

The focus group and cognitive interview data show general agreement on the problems respondents face in trying to provide financial data about their businesses. In many cases respondents simply need clarification on what the questions are asking. Defining financial terms and prompting on what to include and exclude in answering a particular question should help respondents provide better answers. Based on the findings from the focus groups and cognitive interviews, the following modifications to the 1998 SSBF instrument and worksheet were recommended:

##### **Checking Accounts and Savings Accounts**

- ▶ For respondents using one checking or savings account for both personal and business finances, provide prompts on how much of the account balance to report as business related.
- ▶ Add prompts or definitions to assist respondents in deciding whether to include margin accounts and trust accounts.

##### **Personal and Business Credit Cards**

- ▶ Use interviewer prompts to help respondents determine whether their credit card(s) or a particular transaction is relevant to the question.
- ▶ Respondents may assume that the use of credit cards only to finance the business (such as to purchase equipment) is of interest for the survey. Change lead-in, the question, or prompts to instruct respondents that business-related cash advances, entertainment costs, and travel expenses are relevant. Respondents should also be told to include charges they make even if these are immediately reimbursed by the company.

- ▶ Prompt respondents to answer for all owners of the business. For businesses with multiple owners, add instructions in the lead-in to include all business owners in responses. In each question, extra emphasis on all owners can be added (e.g., How much did the firm charge to all business owners' personal credit cards?)
- ▶ Clarify whether to include charges on personal credit cards that are used in part or exclusively for business purposes.

### **Lines of Credit**

- ▶ Define the term “line of credit” in the lead-in to this section or in question F7.

### **Capital Leases**

- ▶ Define capital lease and give an example in the lead-in. Most respondents did not understand this concept.
- ▶ Use interviewer prompts to remind respondents that some financial services may be obtained through sources other than banks.

### **Mortgages**

- ▶ In the earlier question on lines of credit, built-in a prompt clarifying the meaning of “home equity loans used for business purposes”.
- ▶ Currently, respondents are instructed to include an asset or liability as business-related if 50% or more of it pertains to business. Modify the instructions to state that the entire value of the asset or liability (not just a percentage of it) should be included in the answer.

### **Most Recent Loan**

- ▶ To simplify the wording of the first question in the most recent loan section (MRL1), consider including the qualifying statement regarding renewals for lines of credit (“In thinking about this question....” in the lead-in to the section.

### **Income and Expenses**

- ▶ Design interviewer prompts (include/do not include) to help respondents determine what kind of “other income” to report.
- ▶ Set-up the CATI instrument to accept the separate dollar amounts for the costs of doing business so that respondents will not need to sum the numbers themselves. [Note: This suggestion was not implemented, in part, due to the experience in prior rounds of the survey.]

### **Assets and Liabilities**

- ▶ Provide a lead-in that tells respondents they will be asked to break down their

total assets and liabilities into categories. Briefly name these categories to give respondents a sense of what to include/exclude in each question.

- ▶ Consider using respondent answers to prior questions about assets or liabilities as a check. If answers appear inconsistent, respondents can be queried. For example, the respondents' cash holdings should not be lower than the amount they reported in their checking and savings accounts; any liabilities (such as capital leases, mortgages) already reported should be included in the liabilities section. Similarly, loans reported earlier should be associated with positive loan amounts on the balance sheet.
- ▶ Include reference period in all questions.
- ▶ Include interviewer prompt that tells respondents to report full amount of an asset or liability as long as 50% or more of it is for business.

### **Worksheets**

- ▶ Provide direct links between worksheet and questionnaire items with a numbering system. Respondents can then be asked for the amount on a specific line of the worksheet. Currently, respondents must listen closely to each question and try to determine which worksheet line to report. Many respondents do not always refer to their worksheets/records even if they have completed them and have them available.
- ▶ Include interviewer prompts that help respondents find information on tax forms.
- ▶ Make clearer that information from sources other than financial institutions may be relevant. Make space on the worksheet and include prompts that remind respondents of leases and loans they may have from other sources.

### **General Recommendations**

- ▶ Respondents were quite strong in their recommendation of the use of dollar amount ranges. We agree that this should be considered where it will not compromise the utility of the data. If exact values are desired, some instruction regarding rounding should be included.
- ▶ Information that is to be combined in a particular way should be spelled out. For example, if an arithmetic average is desired, ask for it. If a portion of a larger value is to be partitioned out, the question (or series of questions) should lead the respondent to the desired partition. Some clarification of the worksheet questions may simplify respondents' answers. For example, some respondents were unclear about the concept of principal. The questions involving this concept could be broken out in the worksheet to ask for the component values needed to accomplish the computation taking the respondent step-by-step.

Additional information about the methodology study can be found in the *Report on the Methodology Study Conducted for the Board of Governors of the Federal Reserve System*.

### **4.3 Expert Consultation**

Norman Bradburn, Bernard Dugoni, and Lisa Lee, NORC survey methodology and questionnaire design experts, and Charles Smith, of Smith and Associates, a CPA firm specializing in small business accounting, reviewed several iterations of the questionnaire and participated in numerous meetings with FRB and other NORC staff to review the questionnaire. These meetings were informed by basic questionnaire construction methods, observations from the methodology study, our experiences during the two pretests, and the bookkeeping and accounting practices of small businesses. The meetings took place over a period of five months and consisted of reviews of question wording and question order. After each meeting changes were made and then tested prior to the next meeting. Many of the recommended changes were made to the questionnaire prior to the pretests and main data collection.

### **4.4 Pretests**

Two pretests were conducted after conclusion of the methodology study during the Expert Consultation phase of questionnaire development. The pretests had several objectives: informing the design of the screening and data collection questionnaires, evaluating the quality of the Dun and Bradstreet sample, testing the respondent materials to see if they encouraged participation and assisted interviewers in answering questions about the study, and testing our processes and protocols for the main data collection effort.

Data collection was conducted for the first pretest between March 10 and April 1, 1999 and for the second pretest between May 25 and June 25, 1999. Both pretests began by screening 500 businesses using a stratified random sample selected from the Dun and Bradstreet master file (this file is described in detail in Chapter VIII.) The sample was drawn according to the following specifications from the NORC statistician:

- ▶ Exclude subsidiaries, branches, and divisions.
- ▶ Exclude the following SIC codes:

0000-0999	Agriculture, fishing, forestry
4311	US postal service
6000-6399	Depository and nondepository institutions, security and



	commodity brokers, insurance carriers
6700-6799	Holding and other investment offices
8600-8699	Membership organizations
9000-9721	Public administration
821103	Public elementary and secondary schools

- ▶ The listings should be selected in equal numbers from "buckets" of two sizes:

1-19 employees, all sites (500 listings)  
20-499 employees, all sites (500 listings)

- ▶ The selection should be random, and listings should be kept as a "deletion file" to avoid duplication with our main sample draw. The specification will be different for the main sample.

The actual pretest sample had the breakdown by employee size category shown in Table 4.6.

**Table 4.6 Firm Size**

Number of Employees	Number in the Sample
0	3
1 - 19	477
20 - 499	487
500+	33
<b>Total</b>	<b>1,000</b>

The programmer was able to build the CATI program ahead of schedule and in time for the first pretest, therefore, the screener was administered via CATI in both pretests. The questionnaire data collection for the first pretest was conducted using a hard-copy version of the questionnaire; complicated consistency and contingency checks that can be easily accommodated in CATI were eliminated to ease the burden of administration. The second pretest was administered using a CATI version of the questionnaire; all the complex skips and checks were in place for the second pretest.

During each pretest, interviewers screened 500 businesses over a four-day period. In the first pretest, interviewers completed the screener with 335 businesses; 231 of the screened businesses were eligible and willing to participate. During the second pretest,

interviewers completed 341 screeners; 252 of those businesses were eligible and willing to participate.

One objective of the pretests was to generate sufficient eligible cases so that 50 completes could be accomplished in a short period of time; therefore, we did not attempt to target the completion rate goal that will be attempted for the main study. Table 4.7 shows the number of completed cases by firm type.

**Table 4.7 Pretest Eligibility and Completion Rates**

Business Type		Screened as Eligible		Completed Questionnaire		Percent Complete	
		Pretest 1	Pretest 2	Pretest 1	Pretest 2	Pretest 1	Pretest 2
C-Corporation		115	107	24	18	0.21	0.17
Partnership		11	9	0	0	0.00	0.00
LLP	as a Partnership	0	4	0	0	0.00	0.00
	as a Corporation	0	2	0	1	0.00	0.50
S-Corporation		60	75	20	16	0.33	0.21
Sole Proprietor		20	107	6	9	0.30	0.30
Unknown		25	25	0	0	??	0.00
<b>Total</b>		<b>231</b>	<b>252</b>	<b>50</b>	<b>44</b>	<b>0.22</b>	<b>0.18</b>

As Table 4.8 shows, the number of minutes per completed screener for the second pretest was less than for the first pretest. However, the amount of time to complete the interviews for the second pretest was significantly longer. The factors contributing to the higher hours per completed case in the second pretest are:

- ▶ Three working days and a weekend in between mailing the package and calling the respondent led to 121 requests for re-mail
- ▶ The Memorial Day holiday occurred
- ▶ Interviewers insisted that respondents complete the worksheet prior to the interview
- ▶ Using the TNMS, SurveyCraft's case management system, for a small sample.

**Table 4.8 Screening Level of Effort**

Pretest 1			Pretest 2		
Interviewer hours	Minutes per attempted case	Minutes per completed case	Interviewer hours	Minutes per attempted case	Minutes per completed case
215	25.8	38.4	179	21.5	31.5

**Table 4.9 Interviewing Level of Effort**

Pretest 1			Pretest 2		
Interviewer hours	Hours per attempted case	Hours per completed case	Interviewer hours	Hours per attempted case	Hours per completed case
302	1.31	6.04	726	2.88	16.50

Two different versions of the worksheet were used for the second pretest; one was a four-page booklet version, while the other was a long two-sided single page sheet. The purpose of this variation was to determine which worksheet format was easiest for respondents to use in the main interview. The booklet version was sent to C-corporations, and the long version was sent to all other firm types. A total of 31 worksheets and one financial statement were returned to NORC. Several of the worksheets were sent to the FRB for review. Most of the data seemed internally consistent, that is, the balance sheet data balanced. Usage of the worksheets for the 44 complete cases by firm type is included in Table 4.10.

**Table 4.10 SSBF Pretest2 Reported Records Used for Complete Cases**

Organizational Firm Type	Reported Records Used	Number Complete	% Reported Using Worksheets†	Worksheet Type
Sole Proprietor	Tax Records	1	88.9%	Long
	Worksheets	8		
LP	Worksheets	1	100.0%	Long
S-Corporation	Tax Records	3	81.3%	Long
	Worksheets	13		
C-Corporation	Tax Records	5	72.2%	Book
	Worksheets	13		
Overall	Tax Records	9	79.5%	Mixed
	Worksheets	35		

† 31 worksheets were returned, 15 booklets and 16 long form.

Pretest interviewers were briefed both before and after each pretest; FRB staff attended the briefings and debriefings. The briefings consisted of a review of the screener, the questionnaire, and the materials sent to respondents and closed with a question and answer period. The debriefings entailed walking through each screener and questionnaire item allowing interviewers to describe the ease or difficulty of administering each item, how respondents reacted, and how interviewers handled the problems they encountered. The pretest interviewers, project staff, and FRB staff suggested many modifications during the debriefing meetings. The suggestions for changes to questionnaire wording and order are included in the documents in Appendix A, as well as process and protocol recommendations. Some of those key recommendations are listed below:

- ▶ If the respondent indicates that s/he has completed the worksheet but refuses to take the time to give the worksheet information over the telephone, preferring instead to mail-in the worksheet, the interviewer should agree to this option, and simply skip to those questions at the end of the interview that are not asked on the worksheet.
- ▶ Project staff should develop a standard protocol that will help the interviewers guide respondents through a naming convention when the respondents refuse to name the financial institution(s) that provided services.
- ▶ If the owner of the firm has changed since the screener was completed, the interviewer should ask the eligibility questions again of the new “main” owner of the firm.
- ▶ Interviewers need to be sure that the proxy identified by the owner is knowledgeable about the firm’s finances and has access to the businesses financial records. If a proxy takes over prior to Section D, the CATI screens must indicate that the demographic questions reference the owner.
- ▶ When an institution has merged since the time the loan was granted, we should capture the name of the institution at the time the loan was granted.
- ▶ SIC abbreviations were hard to understand. The interviewers asked if the description could be more complete. A second, more detailed SIC description was added to the CATI questionnaire.
- ▶ Interviewers asked for a simple rule to help determine whether or not a business is in operation. Two suggestions were made: If the businesses filed a tax return or

generated expenses or income during December 1998.

- ▶ If the screener was completed by a proxy, send the proxy a postcard thank-you at the time the worksheet package is mailed to the business owner. Interviewers believed that in many cases the proxy would provide much of the questionnaire data and that those proxy respondents who specifically requested the worksheet package be sent to them may feel we ignored their request. Sending a thank-you postcard at the same time the worksheet package is mailed may help in these situations.

Details about the two pretest instruments, respondent materials, survey processes and protocols, and the minutes from the debriefings are included in *The 1998 Survey of Small Business Finances Pretest I Report* and *The 1998 Survey of Small Business Finances Pretest II Report*. These reports were delivered to the FRB in April and July 1999, respectively.

#### **4.5 CATI Development**

Programming for the 1998 SSBF screener and questionnaire began in early February and continued into September 1999. The CATI programmer attended meetings where the questionnaire content was reviewed and changes discussed. The programmer asked questions about presentation, response categories, allowable ranges, and consistency and contingency checks. The desired specifications were noted in a hard-copy questionnaire and then documented electronically in the programmer's log. The specifications were then translated into SurveyCraft code (the programming language used for the CATI instrument). As changes were made to the questionnaire, this process was repeated until the questionnaire was deemed final.

The screener was the first instrument to be completed. Several changes were made to the screener between the initial and second pretest, but since the screener was short, and the changes were few and straightforward, additional work for the programmer was minimal.

Programming for the questionnaire was much more complex and time-consuming than for the screener. This was not only because the instrument design was inherently complex but also because many changes were made to the instrument between the time the programming was begun in February 1999 using a draft instrument and the start of the second pretest. (The project's very tight time schedule made it essential for the programming of the instrument to begin before the instrument itself was in final form.)

In particular, the complexity of the questionnaire required the programming of many consistency and contingency checks. The existence of these checks, plus the ongoing changes to the questionnaire during the programming phase, necessitated a comprehensive testing process to ensure that the instrument met the questionnaire specifications. NORC recommends that for future surveys budgetary provisions be made for more comprehensive testing.

#### **4.6 CATI Testing**

CATI testing began in March, before the first pretest, and continued into September, prior to beginning the Main data collection. The testing protocol followed these general steps:

- ▶ Step 1: The programmer tested the questionnaire to be sure that it was performing according to specifications.
- ▶ Step 2: Project and production staff were assigned specific sections of the instrument to test; each potential path within a subsection was systematically reviewed and tested. Additionally, question text was reviewed to be sure that it matched the hard-copy questionnaire. One person was responsible for maintaining a log of error. Errors were reported to the programmer on a flow basis.
- ▶ Step 3: The programmer made corrections to the text and code and released updated versions of the program for further testing.
- ▶ Step 4: After iterative testing cycles, the instrument was released for data collection.

A detailed test plan can be found in Appendix B.

FRB staff played a key role during the testing phase of the questionnaire design process. NORC sent the FRB three laptop computers on which the CATI program has been loaded. As updates were made to the CATI instrument, NORC posted new versions of the instrument on a web site for download and testing by the FRB.

#### **4.7 1998 SSBF Screener**

The 1998 SSBF screener accomplished the following:

- ▶ Confirmed the name and address of the business and name of the business owner

- ▶ Screened for the appropriate size and type of business
- ▶ Asked the race and ethnicity of the business owner
- ▶ Determined the fiscal year end date of the business

The average time to administer the screener was 4.49 minutes.

Few changes were made to the screener instrument; the screener questions used in 1998 closely resemble those used in 1993. The following changes were made:

- ▶ Added questions to capture the firm's physical address if the mailing address was different from the physical address.
- ▶ Revised the question asking the number of employees in the business. The 1993 SSBF data indicated that 10.4 percent of the small businesses interviewed were firms with no employees. The FRB felt that business owners may have failed to include themselves when asked to report the number of employees of the firm. Therefore, the FRB carefully crafted new questions so that owners were first asked how many owners worked for the firm and then how many employees other than owners worked for the firm. Of the businesses that were interviewed for the 1998 SSBF, 0 percent were firms with zero workers (nine businesses reported zero workers, including owners). The FRB believes that the low percentage was due to the improved screener questions.
- ▶ Added a question to capture the date at which the firm expected to file its 1998 tax return.

The screener questionnaire can be found in Appendix C.

#### **4.8 1998 SSBF Questionnaire**

The 1998 SSBF questionnaire included the following sections:

##### **Section I: Characteristics of the Firm**

- ▶ *Screener.* Questions in this section were asked of those businesses that had been screened by a proxy or firms that indicate an ownership change had occurred since the initial screening; the questions were asked again to be certain the business was eligible to participate in the study.
- ▶ *Organization Demographics.* This section confirms/collects the principal activity of the firm from which a standard industry classification can be made. This section also captures the fiscal year end date; this information helps to put the firm's financial data in context. The business is also classified into one of four major groups: sole proprietorship, partnership, S-Corporation, or C-Corporation.

This section establishes whether there is more than one owner/stockholder of the firm.

- ▶ *Personal Characteristics of the Owner(s)*. This section collects basic information about the principal owner of the firm such as race, sex, age, highest level of education, and number of years managing/owning a business. Sole proprietors and single-owner firms were asked one set of questions, while multiple owner firms were asked a different set of questions.
- ▶ *Firm Demographics*. This section collects information about number of sites and the geographic region served by the business, how the business uses computers, and the single most important problem facing the business.
- ▶ *Records*. This section includes one question that asks the respondent what records s/he will be using during the interview. Because later sections of the questionnaire capture detailed financial information about the firm, we encourage the owner to have business financial records available throughout the interview, especially in sections III and IV. The materials mailed to the respondent in advance of the telephone interview included a worksheet that could be filled-out in advance to make the interview flow more smoothly. Interviewers reported that respondents sometimes used more records than they indicated at this question.

## **Section II: Sources of Financial Services**

- ▶ *Use of Deposit Services*. The name of the financial institution for each savings and checking account held by the business at 1998 fiscal year end was captured along with the dollar amount in each account. In this and the other subsections of Section II, the interviewer asked the respondent to give an estimate if the respondent could not state an exact dollar amount.
- ▶ *Use of Credit and Financing*. Information about the use of credit cards, lines of credit, leases, mortgages, and loans was captured, along with the names of the institutions from which the firm received these services or held these loans.
- ▶ *Most Recent Loan*. Information about the most recent loan that was approved and denied was captured, along with information about the loan(s), including the name of the institution at which the business applied for the loan(s).
- ▶ *Use of Other Financial Services*. This section simply asks if the firm uses transaction services, cash management, and services related to credit, trusts, and brokerages. For each service used, the name of the associated financial institution(s) was captured.
- ▶ *Relationship with Financial Institutions*. The most important source of financial services for the business was determined and the characteristics of up to eight financial institutions were collected.
- ▶ *Trade Credit*. This section captured information about the use of trade credit.



- ▶ *New Equity Investments in the Firm.* This section captured information about additional equity capital and the primary use of the capital.

### **Section III: Income and Expenses**

- ▶ *Income and Expenses.* Detailed information was collected about the firm's income and expenses during the 1998 fiscal year. To be sure that dollar amounts were entered correctly in sections III and IV the dollar value was displayed on the computer screen, in narrative form, and the interviewer read the amount back to the respondent. When respondents could not report a precise amount, they were asked to give an estimate. If the respondent could not, or refused, to give an estimate, the interviewer read a series of dollar ranges and asked the respondent to select a range that most closely fit the amount.

### **Section IV: Balance Sheet**

- ▶ *Assets.* This section asked about the assets of the firm.
- ▶ *Liabilities and Equity.* This section asked for the liabilities and equity of the firm.
- ▶ *Credit History.* This section asked about the credit history of the firm and its principal owner.

## **4.9 Questionnaire Length**

The time to administer partially and fully completed questionnaires ranged from 7.52 to 178.95 minutes with an average time of 40.39 minutes. The time to administer the questionnaire steadily decreased throughout the data collection period, with a reported aggregate time, after two weeks of data collection, of 43.11 minutes.

## **4.10 Questionnaire Improvements**

Many question construction (language) changes were made to the SSBF questionnaire prior to the pretest and main data collection. Most of the changes fell into two categories: adding statements prior to asking the question, so that the question would be better understood; and including definitions in the question text, so that all respondents heard the definition of particular terms, not only those who requested a definition. Other changes included switching the order in which information was presented in the question and simplifying the language.

The question order was changed for certain sections from prior surveys. The most-recent-loan section was moved to precede the institution information, so that all institutions for which information was to be collected were identified. Also, information was collected for the first time on the frequency of credit applications and the characteristics of the institution from which the firm was most recently denied credit. Information on credit cards was asked prior to lines of credit to help respondents separate

these two categories, and detail on both personal and business credit cards was asked for the first time.

Information on the firm's profit, total assets, total liabilities, and total equity were asked only when the respondent could not provide dollar amounts for each of the more detailed elements of the income statement and balance sheet. This change reduced respondent burden somewhat, helped speed up the interview, and eliminated much redundant and inconsistent information that had been problematic in earlier surveys. Lastly, questions were added to collect information on the personal wealth of the business owner.

During questionnaire data frequency reviews and while monitoring interviews in progress, several flaws in the CATI program were identified. The CATI program was adjusted immediately and updated programs were released to the interviewers. The changes made to the questionnaire after data collection began are described in Chapter VI of this report. The questionnaire can be found in Appendix D.

## **V. DATA COLLECTION PREPARATION AND INTERVIEWER TRAINING**

This chapter describes the activities in which NORC and FRB staff engaged in preparation for screener and questionnaire data collection. These activities included materials development; creating a web site; obtaining an endorsement; developing a press release; recruiting, hiring and training staff; and mailing materials to respondents. Identifying the information most important to share with respondents, and presenting this information in a simple, attractive, and professional form was considered critical to the success of the study. NORC and FRB staff developed informational materials that explained the study and attempted to convince respondents to participate. This information was presented in letters, brochures, pamphlets, the reproduction of a speech and research articles, and posted on several different web sites. NORC worked closely with FRB staff on the design of an exclusive web site directed toward the interests of survey respondents. The FRB requested and obtained a letter of endorsement from the Small Business Administration; this letter was used to help gain the cooperation of reluctant respondents. The FRB wrote a press release that was sent to newspapers across the country.

Identifying the right staff to collect these complex data and providing good training was also considered critical to the success of the survey. NORC staff spent a significant amount of time in preparation for interviewer training. Staff from NORC's Human Resource's department and Survey Operations Center recruited and hired interviewers, developed training materials, and conducted interviewer training in preparation for both the screener and questionnaire data collection.

Finally, mailing the materials to respondents prior to screener and questionnaire data collection and planning the receipt and storage of returned worksheets and financial records were also undertaken during this period.

In this chapter we first describe the activities in which staff engaged to support both screener and questionnaire data collection. Then we describe those activities associated with the preparation of screener data collection and close the chapter by describing the activities associated with questionnaire data collection.

### **5.1 Respondent Materials Development**

Soon after contract award NORC and FRB staff discussed the importance of developing a strategy that would assist us in gaining the cooperation of business owners. One of the strategies we discussed was designing materials to send to respondents, in advance of our telephone contacts, that would help to convince respondents to participate in the survey.

We also discussed the fact that we needed two sets of materials: one set designed to convince respondents to participate in a five-minute screener and a more elaborate set designed to convince respondents to participate in a 45-minute interview requesting detailed financial information. The information needed to state the goals of the survey and how the data would be used in an easily understandable form, and also needed to have a polished, business-like appearance. The materials needed to be eye-catching without being flashy.

First we decided on a standard logo that would be used on our materials. The logo NORC suggested, and the FRB approved, was simply the letters S S B F in blue, with each letter separated by a small red diamond, underlined by a red bar with the annotation “The 1998 Survey of Small Business Finances” printed underneath in blue letters. A Word Perfect graphic image of the logo can be found in Appendix HH in the File of Graphic Images. This logo was placed on documents with a white background, giving the logo a patriotic look.

The next step was both to identify the information that we thought would best explain the study and attempt to convince respondents to participate and then to select the appropriate form in which to present the information. The following materials were prepared for the survey:

- ▶ *Frequently Asked Questions Brochure.* This brochure was reproduced in the form of an 8.5 x 11 inch tri-fold. In addition to stating those basic questions and answers that we believed would explain what the survey was about, why the survey was being conducted, and who was funding and executing the survey, information about where respondents could get additional information was also included. This brochure can be found in Appendix E.
- ▶ *Letter from Alan Greenspan.* This letter requested the business owners’ help by participating in the survey, explained why the survey was being conducted, stated that the business was randomly selected, and assured that the information provided would be held in the strictest confidence. This letter was reproduced on the Chairman’s stationery. A copy of this letter can be found in Appendix F.
- ▶ *Letter from the Project Director (for the Screener).* This letter requests the business owner to participate in a short telephone interview for the Federal Reserve Board. A copy of this letter appears in Appendix G.

- ▶ *Letter from the Project Director (for the Questionnaire).* This letter enumerates the contents of the package sent in advance of the call to screen businesses, explains the importance of completing the worksheet in advance of the telephone interview, and tells the respondent how to get more information about the survey. A copy of this letter can be found in Appendix H.
  
- ▶ *Speech Reprint.* FRB Chairman Alan Greenspan mentioned the 1998 Survey of Small Business Finances in a speech he made at the Federal Reserve System Research Conference on Business Access to Capital & Credit in Arlington, VA on March 9, 1999. This text of this speech was taken from the FRB web page and made into a pamphlet, with the remarks made about the survey highlighted and in bold print. The text was printed with blue ink on 17 x 11 inch heavy paper that was folded in half, like a booklet. A copy of the reprint of this speech can be found in Appendix I.
  
- ▶ *Important Facts.* This brochure explained the purposes of the study and presented interesting facts from the prior rounds of data collection. The brochure had a die-cut and was printed on 14 x 11 inch heavy paper with an American Bald Eagle superimposed on the document. A copy of this brochure appears in Appendix J.
  
- ▶ *Customized Worksheet.* NORC spent a considerable amount of time on the design of a worksheet that was sent to respondents. Respondents were asked to complete the worksheet prior to the interview and to use the worksheet as an aid during the interview. The new design was a two-sided form, printed on 17 x 11 inch paper, requesting financial record data on one side and financial services and sources of financing on the other side. Each possible business type (e.g. sole proprietor, LLC, etc.) had a unique worksheet with the appropriate reference to lines in the tax return to assist respondents in looking up the data. Also printed on the worksheet sent to each business was that business's unique identifier; this enabled NORC to match the worksheet data with the questionnaire data when the worksheets were returned to NORC. More details on the development of the worksheet can be found in section 5.8.1.1 of this report. A copy of one of the worksheets can be found in Appendix K.
  
- ▶ *Article Reprint.* A reprint of a newspaper article that points out the differences in lending to small businesses owned by Caucasian business owners and those owned by minority business owners that was sent to minority business owners only. A copy of this article can be found in Appendix L.

- ▶ *Information Insert.* A one-page insert indicating the address of the NORC and FRB websites, in addition to the project's 1-800 number. A copy of this insert can be found in Appendix M.
- ▶ *Folder.* A folder in which the letters, brochures, pamphlets, and inserts are contained in the pockets on the inside of the folder. The folder was made with glossy card stock and had a simple red, white and blue geometric design. The SSBF logo and the SSBF web site address was on the front cover and the toll-free telephone number was printed on the back cover. A copy of this folder can be found in Appendix N.

NORC and FRB staff also identified additional information already in print that would be helpful to include with the respondent mailings. The following additional materials were used to help gain respondent cooperation:

- ▶ *A brochure describing NORC.* A copy of this brochure can be found in Appendix O.
- ▶ *A brochure describing the Federal Reserve Board.* A copy of this brochure can be found in Appendix P.
- ▶ *A detailed booklet that describes the Federal Reserve System.* A copy of this booklet can be found in Appendix Q.

## **5.2 Toll Free Telephone Numbers**

Toll-free telephone numbers were established for respondents to inquire about the study or to fax support materials such as worksheets or tax forms. The telephone number established for voice communication was mentioned in letters and brochures, and respondents were invited to call the toll-free number if they had any questions. This number terminated at our production facility and was answered by staff especially trained to answer questions or route the call appropriately. This number also had a voice-mail option that invited callers to leave a message. Calls received after hours were returned the next business morning.

The toll-free fax number was used to receive worksheets and financial documents; the fax facilitated both the distribution and receipt of worksheets, tax forms, and other respondent financial information. Additionally, the fax was used to send information to businesses; businesses that said they did not receive our worksheet mailing were sent the FedEx tracking confirmation data so that they could locate the package by speaking with

the person who signed the Federal Express receipt. Occasionally, a business asked that we send a letter or other information via fax, which we did, as soon as the request was made.

### **5.3 Letter of Endorsement**

NORC and FRB staff felt that an endorsement from an organization respected by most businesses could help persuade business owners to participate in the survey. The COTR asked the Small Business Administration (SBA) to endorse the survey, and they did so. A copy of the endorsement letter can be found in Appendix R.

### **5.4 Press Release**

NORC and FRB staff also felt that announcements of the survey in the news media could help persuade business owners to participate. The FRB prepared a press release that was sent to major news media organizations across the country. A copy of the press release appears in Appendix S.

### **5.5 Electronic Mail Address**

NORC established an E-mail address for respondents to use, if that was the business's preferred method of communicating questions or information. The account received approximately 15 inquiries; most of the respondents wanted to know when the current study's data would be available and where data from past rounds could be found. We expect that use of E-mail communication will increase in future rounds of the SSBF.

### **5.6 Web Sites**

Many business owners use the Internet for both business and personal reasons. Therefore, NORC and FRB staff thought that a site with information about the project could be helpful in gaining respondent cooperation, in addition to giving respondents another means of obtaining information about the study. At the SSBF website established by NORC (<http://www.norc.uchicago.edu/ssbf/homepage.htm>) respondents could learn more about the study by reading the answers to frequently asked questions. Respondents could also obtain extra copies of the various worksheets and read the press release about the study. The site had links to the FRB site and the SBA site.

Word Perfect graphic images for both of the screens for the 1998 SSBF from the NORC site can be found in Appendix HH in the File of Graphic Images.

The FRB also prepared a page on their WEB site devoted to the 1998 SSBF survey (<http://www.federalreserve.gov/ssbf/>). The FRB page explained the purpose of the study, posted answers to frequently asked questions, posted reports of previous rounds of the study, and included the remarks from Chairman Greenspan and others about the SSBF. This page also had a link to the FRB home page and a link to the NORC site.

A Word Perfect graphic image for both of the screens for the 1998 SSBF from the Federal Reserve Board WEB site NORC site can be found in Appendix HH in the File of Graphic Images.

## **5.7 Preparation Activities for Screener Data Collection**

**5.7.1 Respondent materials.** As already detailed, NORC and FRB staff developed a set of materials to be mailed as an informational packet to the respondents in advance of the screening call. The materials included a letter from the FRB Chairman Greenspan, a letter from the NORC project director, and a brochure with answers to the most frequently asked questions about the study. These materials, which were tested during two pretests, were designed to give respondents information about the study and its sponsors, to answer questions they might have, and to convince them that their participation was critical to the success of the study.

**5.7.2 Training Materials.** NORC developed materials for the interviewer training session used to prepare interviewers to screen businesses. The materials included an agenda (shown in Appendix T), Trainers' Guide, Interviewer Manual, mock interview scenarios, and job aids. The agenda is typically the first document prepared for training as it forms the basis from which all other training documents are prepared. The agenda is organized in such a way that each topic builds upon the information presented in the previous topic. The material presented in both the Interviewer Manual and the Trainers' Guide follow the outline of the agenda.

NORC staff prepared a Trainers' Guide for the screener training. The *1998 Survey of Small Business Finances Trainers' Guide for Main Screener (June 1999)* consists of modules that describe in detail the concepts and materials covered during the training. The module instructions outline learning goals for that section, the presentation mode used (e.g., lecture, round-robin mock interview), the materials needed, and an explanation of the importance of each



topic.

The emphasis of the training was to familiarize the interviewers with the purposes of the study, how to gain cooperation, the importance of reassuring respondents about the privacy of their responses, and the eligibility criteria. Concerns specifically addressed were the importance of the survey eligibility criteria (e.g., whether the firm was in business during 1998), and employment, race, and ethnicity definitions. Interviewers were instructed on the protocol necessary to identify the appropriate respondent – one of the principal firm owners – and relay to the respondent the confidentiality of any information provided to the interviewers.

The following training modules described in Table 5.1 were developed for screener training:

**Table 5.1 Training Modules**

Day	Module Number	Module Description
Day 1	Module 1	Welcome/Introduction. The training opens with introductions of both FRB and NORC staff.
	Module 2	Study Background and Purpose. This is a presentation of the history of the FRB and the background and objectives of the SSBF survey.
	Module 3	Screener Review. This module explains the interviewers' tasks during the screener call, introduces the eligibility criteria, and demonstrates the use of the screener.
	Module 4	Eligibility Criteria. This module explains in detail the eligibility criteria that are critical in determining which firms are included in the study. In order to be eligible, the firm must be: in operation in 1998, headquarters of the business, non-subsiary, for-profit, not owned by a governmental entity, and under 500 employees.
	Module 5	Confidentiality. This module explains the importance of confidentiality. Interviewers are also reminded of the pledge of confidentiality they made at the time they were hired.
	Module 6	Introduction Scenarios. This discusses the importance of the first 30 seconds of the initial contact with the respondent. Interviewers then engage in round-robin role-play exercises with the introduction.
	Module 7	Mock Screeners. Using scripted mock screeners, interviewers participate in round-robin role-play exercises with the screener.
	Module 8	Review Advance Materials. This module reviews the materials included in the advance letter to respondents.
	Module 9	Gaining Cooperation. In this module we review the answers to commonly asked questions and talk about ways to gain cooperation.

Day	Module Number	Module Description
	Module 10	Mock Screeners. Using scripted mock screeners, interviewers participate in round-robin role-play exercises with the screener.
	Module 11	Eligibility Exercises. This module add the practice of gaining cooperation into exercises developed to help interviewers identify ineligible companies.
	Module 12	Review Eligibility Exercises. NORC and FRB staff answer questions regarding eligibility.
	Module 13	Review Homework Assignment and Agenda for Following Day.
Day 2	Module 14	Review Completed Homework Assignment.
	Module 15	TNMS/CATI Overview. The TNMS delivers the cases at appointed times or according to programmed algorithms and also stores the cases when they are not being worked by the interviewers. The CATI is the actual on-line questionnaire. This module explains the functions of the data collection software packages.
	Module 16	TNMS/CATI Exercises. In this module, interviewers are given login instructions and an exercise that tests their knowledge of the TNMS and CATI.
	Module 17	Mock Screeners. During this module, interviewers practice administering the screener using the screener on-line. Interviewers also practice gaining cooperation.
	Module 18	TNMS Dispositions. In this module, interviewers learn the case dispositions and complete an exercise that require them to categorize various case outcomes.
	Module 19	Administrative Tasks. Administrative details are explained in this module.
	Module 20	Mock Screeners. These mocks are completed in interviewer pairs with one interviewer role-playing as the respondent and another role-playing as the interviewer.
	Module 21	In-class Quiz. Interviewers are given a quiz to test their understanding of information presented in the training; interviewers complete the quiz independently. The answers to the quiz are reviewed after everyone has had sufficient time to complete the quiz.
	Module 22	Question and Answer Period and Wrap-up. This module closes the training session by answering any questions the interviewers have and explains the certification process.

This guide helped to ensure that the trainees in all of the various screener training sessions received the same information.

NORC prepared a manual for interviewers that described the study and explained the concepts the interviewers needed to know (primarily the eligibility criteria) to properly administer the screener instrument and collect good screener data. The *1998 Survey of Small Business Finances Telephone Interviewer Manual for Main Screener (June 1999)* was distributed to interviewers at the start of the training so they could become familiar with the contents and make notes in the manual

during the training session. The manual became a handy reference guide after the training was completed. The interviewer manual included the following chapters and appendices:

- Chapter 1: Study Background. This section provides information on the previous iterations of SSBF, its purposes and goals, as well as background on the FRB.
- Chapter 2: Confidentiality. The interviewers are informed of the importance of confidentiality and their role in the confidentiality procedures of NORC.
- Chapter 3: SSBF Eligibility Requirements and Screener. This chapter outline the eligibility requirements that a firm must meet in order to be considered for the main sample. It also describes the advance materials sent to respondents and the organization of the screener.
- Chapter 4: Telephone Number Management System (TNMS) and Computer Assisted Telephone Interviewing (CATI). The TNMS and CATI are integral to the implementation of the interview. This chapter briefs the interviewers on the functions of these systems. The TNMS organizes the cases in the order in which they will be called by the interviewers. It facilitates randomness and also keeps track of appointments to make sure that respondents are called back at the pre-arranged days and times. The CATI is a software program that enables interviewers to have visual and keyboard access to the questionnaire while on the phone with respondents. It then captures the data to be cleaned and analyzed in later phases.

Appendix A: NORC's Commitment to Confidentiality

Appendix B: Package of Advance Materials

Appendix C: CATI Screener with Question-by-Question Answers

Appendix D: TNMS Technical Supplement

Appendix E: Daily Progress Chart

Appendix F: Job Aids

To facilitate an understanding of the screener and the variety of possible

screening outcomes, NORC prepared mock scenarios to be used during group round-robin and duo-mock practice sessions. The initial scenarios were straightforward situations, and the subsequent scenarios were less typical and more difficult situations.

NORC also prepared job aids to assist the interviewers both during training and while interviewing respondents. Some of the job aids were designed to be at the interviewers' work stations to provide easy access to key project information, such as the 1-800 number, the name of the sponsoring agency, and the frequently asked questions and answers. Other job aids were available in both hardcopy (included in the interviewer manual) and electronic form (included in CATI), such as the glossary of terms and the question-by-question specifications. The job aids were distributed at the training; copies of the job aids were also included in the interviewer manual.

**5.7.3 Recruiting and Hiring.** NORC began recruiting interviewers for the screening effort in early May 1999. In general, recruiting sufficient numbers of qualified interviewers for both the screening and interviewing phases of this project was challenging. It required creative and sustained initiatives over an extended time frame, with allowance made for attrition during and after training. For the screening phase, NORC invited 87 candidates to the initial training, 72 of these accepted and attended the initial session, including approximately half experienced NORC interviewers and half who were new. Of this number, 66 passed the certification exam at the conclusion of training.

Advertisements were placed in local newspapers, notices were placed at community organization and state unemployment offices, and notices were posted at several area colleges. Additionally, announcements about the job openings were made during several public service spots on a local radio station.

NORC wanted to be able to accommodate Asian-language speaking respondents, so to recruit Asian- language-speaking interviewer candidates, notices were posted at Asian community organizations and churches. NORC also ran an advertisements in Korean and Chinese newspapers, as well as posting notices with Asian student organizations at local colleges.

The advertisements instructed interested candidates to call NORC for further information. When candidates called they spoke to the recruiter, who invited those that seemed to have the proper qualifications to attend one of the several

general information sessions offered. At the sessions, recruiters measured verbal and quantitative ability, and engaged each candidate in a one-on-one interview. Candidates were told to arrive early to complete the application. The information sessions were held in groups of up to 15 interviewer candidates; the information session introduced NORC, provided a description of the project, and described the role of an interviewer. Then interviewers took two 20-minute standardized tests that evaluated verbal and quantitative skills. Following the tests, candidates were invited to sit at a workstation so that their keyboard skills could be evaluated; this exercise also gave candidates an opportunity to see what a CATI interview looked like. The session ended with one-on-one meetings that included an oral reading (which included the screener instrument) and a job interview.

**5.7.4 Training.** NORC held supervisor and interviewer training sessions to prepare staff to supervise interviewers and to screen respondents.

**5.7.4.1 Supervisor Training.** Two weeks prior to the interviewer training, NORC held a training session for supervisors who were assigned to oversee the screening work. The training was led by project staff and the two Telephone Center staff who had been involved in the project since the execution of the first pretest. All of the supervisors assigned to the SSBF project had assisted with testing the CATI instrument and had some familiarity with the screener instrument prior to the training. To train the supervisors, project staff used the materials that were prepared for the screener training; this was an opportunity to both test the training materials and train the supervisors.

**5.7.4.2 Interviewer Training.** The FRB staff joined the NORC team of managers, supervisors, and programmers to train the screening interviewers. The initial training session was held on June 14 and 15, 1999. The COTR presented information about the mission of the FRB, presented historical information about the SSBF project, and shared interesting findings from data collected in prior rounds of the study.

The training agenda covered a wide range of topics in both lecture and interactive sessions.

Practice sessions allowed the interviewers to administer the screening instrument using mock eligibility scenarios. The mock interviews were

practice scenarios using various pre-loaded respondent information giving the interviewers a chance to path through all portions of the screener interview, as well as the opportunity to practice navigating some of the myriad situations that occur while on the telephone with a respondent. These mock scenarios started out as straightforward exchanges between interviewers and trainers, so that the interviewers could learn the basic screening task. The mocks became increasingly more complex as the training progressed to expose the interviewers to many of the challenges that could be presented by respondents.

At the close of the initial day of training, the trainees were assigned a specific written study task to be completed at home; this assignment was reviewed at the beginning of the following day. The training concluded with an interviewer certification module, in which each interviewer must pass a comprehensive test to become a certified SSBF interviewer before working with live respondents in the actual study.

NORC conducted two additional trainings. In late June a special training was held for two bilingual English/Spanish speakers. They began interviewing in early July. The bilingual interviewing pool included one Cantonese speaker, three Spanish speakers, and one Korean speaker. NORC conducted an attrition training on August 23 and 24. At that time six interviewers were trained and added to the screening team.

**5.7.5 Interviewer Certification.** After screener training and before interviewers were allowed to begin screening respondents, each interviewer was required to demonstrate that he or she understood the eligibility requirements and could successfully:

- ▶ Gain cooperation from resistant respondents,
- ▶ Record verbatim responses accurately, and
- ▶ Perform a CATI screener interview.

In order to test the interviewers' understanding of eligibility requirements, interviewers completed a written test that challenged their understanding of the eligibility criteria by presenting various scenarios and asking whether the company in each scenario was eligible, ineligible, or if there was not enough information to determine eligibility.

The gaining-cooperation exercises were scripted scenarios. Each interviewer had to demonstrate the ability to address a respondent's concerns, answer questions appropriately, overcome several different roadblocks, and persuade the respondent to participate.

The CATI certification screener was a scripted mock that tested probing skills, verbatim recording, using the CATI system, and coding responses. For the initial training, the interviewer certification process was completed on June 23, 1999.

**5.7.6 Mailing Advance Respondent Packages.** The screening sample consisted of 39,240 businesses purchased from Dun and Bradstreet. A subset of these addresses included in the sample purchased from D&B were incomplete, therefore, NORC attempted to obtain a more complete address prior to the screener mailout. Because we were unable to obtain addresses for 13 businesses the advance letter was mailed to 39,227 businesses. The sampling statistician divided the sample into 393 random replicates, so that the sample could be fielded in small batches.

The mailings, sent via U.S. first-class mail, were staged to coincide with the initial telephone screening contact, so that the letter was received about one week prior to the initial call to the business. The mailings began on June 12 and ended on July 31, 1999. As stated earlier, the advance mailing included a personalized letter from the project director, a letter from the chairman of the Federal Reserve Board, and a question and answer brochure.

Of the 39,227 businesses to which NORC mailed a letter, 3,836 letters were returned: 1,071 included new address information and 2,785 were returned without new address information. NORC prepared new mailing envelopes for those letters returned with an address update and remailed letters to those businesses. Businesses for which a letter was returned without an address update were routed to interviewers especially trained to locate businesses. Various means such as directory assistance, internet searches, and contacting community organizations such as the local Chambers of Commerce were utilized to attempt to locate such firms.

## **5.8 Preparation Activities for Questionnaire Data Collection**

**5.8.1 Respondent materials.** The NORC and FRB staff developed a set of materials to be mailed to respondents in advance of the telephone interview.

These materials were described in detail in section 5.1 and included the following:

- ▶ Letter from Alan Greenspan encouraging participation;
- ▶ Letter from the project director encouraging participation;
- ▶ Brochure about NORC;
- ▶ Brochure about the Federal Reserve Board;
- ▶ Copy of the speech Alan Greenspan made at the Federal Reserve System Research Conference on Business Access to Capital & Credit in Arlington, Virginia, on March 9, 1999;
- ▶ Important Facts Brochure detailing interesting information from prior rounds of data collection;
- ▶ Worksheet customized by firm type and labeled with a case identifier;
- ▶ Reprint of an article that points out the differences in lending to small businesses owned by Caucasian business owners and those owned by minority business owners (only sent to minority business owners only); and
- ▶ One-page insert indicating the address of the NORC and FRB web sites and the project's 1-800 number.

**5.8.1.1 Worksheet Development.** The redesign of the worksheet used for the 1993 SSBF was a major undertaking. The design of the 1993 worksheet had been a multi-page booklet with the reference to tax return schedules and lines for each type of business form included on the worksheet. The business owner could look for the reference for their particular business form to identify the appropriate figures on the tax documents to provide accurate data. For some questions the business owner needed to look for the question sub-part that applied to their particular business form and answer just the questions that applied.

NORC thought that the 1993 worksheet design required too much of the respondents. NORC and FRB staff decided to create a unique worksheet for each business type, therefore, ten different worksheets were created -- different forms for organizational types and fiscal year end dates. Furthermore, NORC and FRB staff felt that a single page worksheet was preferable to a booklet. The amount of information included in the worksheet was more than could be included on two sides of an 8.5 x 11 inch page. Therefore, NORC used an 11 x 17 inch page and put the questions capturing financial data on one side and the services and



providing institutions on the other side. This design was tested during the pretest and was well-received by respondents.

**5.8.1.2 Mailout and Receipt Protocol.** A personalized letter and a worksheet customized by firm type and labeled with a case identifier were enclosed in the advance package. Additionally, for the minority-owned firms, a reprint of an article was included that points out the differences in lending practices to small businesses owned by Caucasians and those owned by minority individuals. All of the SSBF generic materials were carefully organized in a pocket folder and placed into an overnight Federal Express envelope with an appropriate air bill. Given the level of customization, this mailing was particularly labor intensive.

The assembly of the generic materials into the pocket folders was completed in advance of the mailings; a subset of the folders included the article to be included in each of the folders sent to a minority business and those folders were stored separately from the folders prepared for the non-minority cases. Each week the mailout center prepared a new batch of cases to be mailed by generating the personalized letter, the customized worksheet, and the air-bill for each firm in those replicates that were scheduled to be mailed that week. A small team of staff assembled the packages and examined them to be sure that each of the customized pieces belonged to the same business.

Packages returned as undeliverable and completed worksheets sent to NORC by respondents were carefully receipted. The receipt of an undeliverable package triggered the notification of the telephone center so that locating could begin on that case. Worksheet returns were stored by case identifier so that they could be easily retrieved when respondents refused to answer those questions for which they already provided data.

**5.8.2 Training Materials.** NORC developed materials for the interviewer training session held to prepare for administration of the questionnaire. The materials included an Agenda (Appendix U), a Trainers' Guide, an update to the Interviewer Manual, mock interview scenarios, and job aids. Just like the screener training agenda, the questionnaire training agenda is organized in such a way that each topic builds upon the information presented in the previous topic. The material presented in both the Interviewer Manual and the Trainers' Guide follow the outline of the agenda.

The questionnaire training had to provide a greater level of detail about business finance than was presented at the screener training. While preparing the materials for the questionnaire training, NORC was careful to focus on those issues that would present the greatest challenge to interviewers (e.g., learning financial concepts and understanding the various financial documents and recording decimals and large numbers), as well as the organization of the questionnaire and the path variations that would be encountered by interviewers.

NORC staff prepared a Trainers' Guide for the main training. The *1998 Survey of Small Business Finances Trainers' Guide for Main Interviewing (September 1999)* describes in detail the concepts and materials covered during the training. The Trainers' Guide consists of modules that follow the items listed in the Training Agenda. The module instructions outline learning goals for that section, the presentation mode used (e.g. lecture, round-robin mock interview), the materials needed, and an explanation of the importance of each topic. The training modules listed in Table 5.2 were developed for Main Interview training:

**Table 5.2 Main Interviewing Training Modules**

Day	Module Number	Module Description
Day 1	Module 1	Welcome/Introduction. All project staff, clients, interviewers, and telephone center managers introduce themselves.
	Module 2	Study Background and Purpose. The COTR and the FRB staff review the history of the SSBF, the completion of the screening effort, the investigative goals of the study, and how the interviewers' role fit into the big picture. They also presented background on the FRB.
	Module 3	Questionnaire Review. The trainers go through the primary sections of the telephone interview as well as the question-by-question notes that accompany each part.
	Module 4	Eligibility Criteria and Sample Composition. The eligibility standards set forth in the screener are re-emphasized, both for those who had and those who had not worked on the screener.
	Module 5	Review Advance Materials. The contents of the advance mailing packet are distributed and explained, as are several job aids and forms, such as Re-mail Request Forms, that are pertinent to the mailing.
	Module 6	Detailed Review of Questionnaire Section I: Characteristics of the Firm – Questionnaire Subsections A, B, C, D, and N. The trainers lead the interviewers through these portions of the interview step-by-step, taking care to indicate particular complexities.

Day	Module Number	Module Description
	Module 7	Detailed Review of Questionnaire Section II: Sources of Financial Services (Roster Building) – Subsections E, F, MRL, G, and H. A crucial part of the interview is to capture information on the firm’s financial services and products, and at which financial institutions they are provided. This module familiarizes interviewers with this roster-building process.
	Module 8	Detailed Review of Trade Credit - Subsection L. This section familiarizes interviewers with questions about the firm’s trade credit and suppliers.
	Module 9	Detailed Review of New Equity Investments - Subsection M. Trainers review the details of questions about new equity investments for a firm.
	Module 10	Wrap-Up: Assign Homework/Review Agenda for Day 2. This is primarily an administrative session where trainers answer questions from the interviewers for that day, distribute the daily assignment, and review time and place logistics for the next training day.
Day 2	Module 11	Review of Homework. This module resolves any issues from the previous night’s homework assignment.
	Module 12	Overview of Income and Expenses and Balance Sheet. The trainers outline the particular aspects of the Balance Sheet, Income, and Expenses sections.
	Module 13	Detailed Review of Balance Sheet - Subsections P, R, and S. This module familiarizes interviewers with the balance sheet part of the interview. More specifically, recording decimals and large dollar amounts correctly, rounding numbers, and learning financial concepts are topics covered.
	Module 14	Detailed Review of Credit History - Subsection U. The trainers take the interviewers through questions on the credit history of both the firm and the owner.
	Module 15	Gaining Cooperation. The trainers use lecture, role play, and round-table discussions to review basic principles of gaining cooperation and refusal avoidance. This entails use of several job aids, such as the SSBF Frequently Asked Questions.
	Module 16	Confidentiality. This module, standard in training, allows the trainers to emphasize the importance of confidentiality for respondents and to outline the procedures that NORC has in place to preserve it.
	Module 17	Mock Interview: Sole Proprietorship. This module is the first in a series of mock interviews where the trainers take the interviewers through practice interviews with scripted responses that path through the instrument in various ways.
	Module 18	Mock Interview: C-Corporation. This mock interview focuses on questions and pathing that are particular to C-Corporations.
	Module 19	Wrap-Up: Assign Homework/Review Agenda for Day 3. Like the Wrap-Up module for day 1, trainers answer outstanding questions from the day, explain the homework assignment, and review the agenda for the coming day.

Day	Module Number	Module Description
Day 3	Module 20	Review of Homework. The homework module for this day facilitates discussion about the interviewers' assignment from the previous night, as well as the substantive Sections III and IV of the questionnaire.
	Module 21	Break-offs at Section P. This module covers the procedure for handling break-off interviews when the worksheet has already been mailed or when the respondent agrees to mail in the completed worksheet after the interview.
	Module 22	Mock Interview: Partnership. As in the other mock interviews, the trainers took the interviewers through the instrument. This time, they follow the particular paths and skips associated with partnerships.
	Module 23	Administrative Tasks. This module explains staffing, toll-free phone and fax numbers for the project, time cards, attendance policies, and production tracking.
	Module 24	Mock Interview: S-Corporation. This mock covers substantive issues and questionnaire paths that are related to S-Corporations.
	Module 25	Mock Interview: C-Corporation. This is the second mock that prepares interviewers for the particular issues associated with C-Corporations.
	Module 26	Mock Interview: S-Corporation. This module runs through another mock interview for S-Corporations.
	Module 27	TNMS/CATI Overview. The trainers review the basic functions of the TNMS and CATI software.
	Module 28	TNMS/CATI Exercises. This module consists of exercises where the interviewers recall their login procedures and TNMS/CATI functions.
	Module 29	Certification. In this module, trainers take interviewers through a series of evaluations to ensure that they are prepared for actual telephone interviews with actual respondents.

The three-day training was not enough for most of the interviewers. All of the interviewers were invited to take an additional day to practice before the certification interview with a supervisor, and almost all of the interviewers took the extra time. In the future, NORC recommends that the substantive training span four days, with the last day devoted to supervised practice sessions.

For the interviewer training, NORC updated and expanded the interviewer manual prepared for the screener training. The updated manual contained the basic information shared with screening interviewers, explained the financial concepts and intricacies of the instrument, and more fully detailed the various paths that each type of firm could travel through the instrument. The updated *1998 Survey of Small Business Finances Telephone Interviewer Manual (June*

1999) was distributed to interviewers at the start of the training so that they could become familiar with the contents and annotate the manual during the training. The manual became a handy reference guide after training was completed. The interviewer manual included the following chapters:

- Chapter 1: Study Background. This section provides information on the previous iterations of SSBF, its purposes and goals, and background on the FRB.
- Chapter 2: Confidentiality. The interviewers are informed of the importance of confidentiality, and their role in the confidentiality procedures of NORC.
- Chapter 3: SSBF Eligibility Requirements and Screener. This chapter outlines the eligibility requirements that a firm must meet in order to be considered for the main sample.
- Chapter 4: TNMS and CATI. The TNMS and CATI are integral to the implementation of the interview. The interviewers are briefed on the functions of these systems.
- Chapter 5: Conducting the Interview. This section contains instructions and helpful tips regarding the administration of the SSBF interview.
- Chapter 6: Gaining Cooperation. The interviewers learn techniques and engage in role play that teaches them strategies for refusal avoidance and gaining cooperation.
- Chapter 7: Locating. This skill is outlined as a series of steps in the process by which interviewers track down respondents in the sample.
- Chapter 8: Administrative Specifications. Interviewer's administrative tasks are reviewed as a reference for the procedures and policies related to payment, attendance, telephone protocol, etc.
- Chapter 9: Main Interview Introduction. The interviewers learn to broach the interview with the respondents and answer questions.
- Chapter 10: Main Interview Review. This chapter covers the items specific to the SSBF interview, the main sections, and special issues.
- Chapter 11: When Bad Things Happen to Good Interviewers. Due to the nature, complexity, and length of the SSBF interview, there is margin for "bad things to happen." This chapter provides ways to cope and work around such situations.
- Chapter 12: Worksheet Mailing. Prior to the telephone interview, a mailout package of various materials is sent to respondents. This chapter explains each item in detail.
- Chapter 13: Main CATI Mechanisms and Features. Interviewers need to be

extremely familiar with the CATI, its function buttons, pathing, and general characteristics. This section outlines information that enables the interviewers to get comfortable with the software.

**Glossary:** This section of the manual contains terms and definitions, many of them related to financial concepts, that are included in the interview.

**Appendix A:** NORC's Commitment to Confidentiality

**Appendix B:** Package of Advance Materials

**Appendix C:** CATI Screener with Question-by-Question Answers

**Appendix D:** TNMS Technical Supplement

**Appendix E:** Daily Progress Chart

**Appendix F:** Job Aids

To facilitate an understanding of the main interview, particularly the financial services roster and balance sheet sections, as well as the variety of possible firm paths and outcomes, NORC prepared mock scenarios that were used during group round-robin and duo-mock practice sessions of training. As in the screener training, the initial scenarios consisted of straightforward situations and the subsequent scenarios consisted of increasingly more difficult respondents or businesses with many institutions and multiple services.

NORC also prepared job aids to assist the interviewers both during training and while interviewing respondents. The job aids were developed with both general interviewing and specific project experience in mind. For example, NORC has learned through previous studies that aids giving standard information such as the 1-800 number, the name of the sponsoring agency, and the OMB Standards for Race and Ethnicity Classification are very helpful. As well as including some of the same information as those prepared for the screener training, the job aids provided information specific to the main interview, such as the CATI functions for the main interview. The FRB brought previous experience to bear in informing our development of job-aids that instructed interviewers with procedures to follow if a respondent breaks-off in the middle of the interview or recalls an additional service / financial institution later in the interview. Other job aids available in both hardcopy (in the interviewer manual) and electronic form (in CATI) were the glossary of terms and the question-by-question specifications.

The job aids were distributed at the training; copies of the job aids were also included in the interviewer manual.

NORC staff prepared 13 job aids for the data collection. Copies of all job aids are shown in Appendix V. They include:

1. *Answers to Interviewers' Most Frequently Asked Questions.* This document was created as a first-line resource for interviewers. It addresses inquiries such as: "What is the purpose of the Survey of Small Business Finances?" , " Who is doing the survey?" , "Will my answers be confidential?" , "Why was I chosen to participate?" , "What kinds of questions will be asked?" , " Why do you need my financial information?" , "Am I required to participate and answer every question?" , " What will I get out of this?" , and "Is additional information available?" Many interviewers used the answers in this job aid in gaining cooperation from the respondents.
2. *SSBF Brochure.* A brochure on the SSBF and results from past iterations of the study provided background on the study and uses of the data.
3. *Fax/Re-mail Request Forms.* These forms were utilized internally to keep track of respondent requests, and fulfillment of those requests, for additional mailout materials. Re-mail occurred primarily in cases where the respondent had misplaced the material or it had been lost in transit.
4. *Eligibility Questions.* A reprint of the eligibility criteria was provided whether or not the interviewer had worked on the screening effort. In this way, all interviewers were familiar with the criteria, and they could make sure that the firms they were interviewing were eligible respondents.
5. *Generic Institutions.* This aid assisted interviewers in addressing situations where the respondent would not divulge the name of their financial institution(s). The instructions standardized the way in which such cases were recorded, thus facilitating downstream analysis.
6. *Break-Off Questionnaire.* This document includes the most important financial questions from Sections P, R, and S, as well as all of the Section U questions. Interviewers utilized it in cases where the respondent had already completed and/or mailed in their worksheet and did not want to answer the same questions again over the telephone. This form was match it against the incoming worksheets in break-off case processing.

7. *FedEx Tracking.* The tracking forms utilized internally to note efforts to locate mailout packages that had been lost in transit or to find out which person in the respondent's firm had signed for the package. This tracking and locating effort minimized material re-mails.
8. *Proxy Job Aid.* This aid defined appropriate respondents and differentiated them from inappropriate respondents. Only those respondents who were knowledgeable about the firms' finances qualified as respondents. Ideally, this person would be an owner. However, when an owner was not present, this aid helped the interviewer to find an appropriate proxy.
9. *Race Coding Aid.* This aid, patterned on the 1996 U.S. Census race coding frame, was designed to provide a quick reference for the demographic sections of the instrument. It was also used to eliminate coding error and to decrease time spent back-coding to correct such errors.
10. *Extra Services Job Aid.* This aid assisted interviewer in correcting answers changed by respondents later in the interview. This situation occurred frequently in past iterations of the SSBF.
11. *Tax Forms.* Appropriate tax forms for each firm type were distributed to the interviewers, even though specific instructions were provided, both on the worksheets and by the CATI interview, on where to find the answers on the tax forms that correspond to each question in the instrument. A copy of the 1998 U.S. Tax Forms can be found in Appendix II.

NORC staff also prepared some job aids regarding refusal conversion and worksheet mailing specifications:

12. *Refusal Conversion Form.* Field interviewers (the field effort is described in Chapter VI) logged their interaction with respondents on this form, which were then sent to the telephone center to cross-check with the contacts occurring there.
13. *Worksheet Mailing Specifications.* These aids were distributed to all field staff, as they were performing initial mailings and re-mailings of materials independent of the central mailout center. The form assisted interviewers in mailing the correct materials, particularly worksheets, to any given



respondent, and standardized the mailing process throughout the field and the central mailout center.

**5.8.3 Recruiting and Hiring.** All of the interviewers invited to the initial interviewer training for the main data collection had participated in the screener training and had conducted screening interviews. At subsequent trainings we had a mixture of experienced and new interviewing staff. To identify potential candidates for the main interview, the recruiter used the same resources and processes that were used for screener recruiting, with one exception: the oral reading during the one-on-one interviews with potential candidates included subsections of the main interview, to be sure that interviewers could follow instructions and properly pronounce the financial terms included in the instrument. This exercise also gave the candidates an opportunity to experience the task that they would be performing.

**5.8.4 Training.** NORC conducted two interviewer trainings in September 1999. The first training was scheduled to be held between September 7 and September 9; the training actually spanned five days, concluding on Monday, September 13. The second training, scheduled to be held between September 28 and September 30, also spanned five days, concluding on Monday, October 4. For both trainings, the initial day took place in the Harris School classrooms. All subsequent training days were held at NORC's Lake Park facility. The initial training included 24 interviewers while the second training included 21 interviewers. All interviewers attending the initial training successfully completed the training and certification protocol. All 21 interviewers attending the second training completed the training, but only 17 successfully completed the certification protocol.

A special training session for interviewers assigned to evening and weekend shifts was held from Thursday, October 8, through Monday, October 11. An additional training for daytime staff was held from October 19 through 21. Table 5.3 summarizes this information.

**Table 5.3 Training Dates, Number of Trainees, and Interviewer Attrition**

Training Session	Training Dates	Number of Interviewers Attending Training	Number of Interviewers Certified	Number of Interviewers Retained through the end of data collection
1	9/7 thru 9/13	25	21	14
2	9/28 thru 10/4	21	18	8
3 (evening)	10/8 thru 10/11	6	6	2
4	10/19 thru 10/21	11	5	3
5	11/15 thru 11/19	19	13	4
<b>OVERALL</b>	<b>24 days</b>	<b>82</b>	<b>50</b>	<b>31</b>

**5.8.5 Interviewer Certification.** After interviewer training and before interviewers were allowed to begin contacting respondents and administering the questionnaire, each interviewer was required to demonstrate that he or she understood the eligibility requirements and could successfully:

- ▶ Gain cooperation from resistant respondents,
- ▶ Record verbatim responses accurately, and
- ▶ Administer an interview using CATI.

The gaining-cooperation exercises were scripted scenarios. Interviewers had to demonstrate the ability to address a respondent's concerns, answer questions appropriately, and persuade the respondent to participate.

In order to test the interviewer's understanding of the material covered in training, the interviewers completed a written test that challenged their understanding of the following:

- ▶ Characteristics of an appropriate proxy,
- ▶ Eligibility criteria,
- ▶ Contents of the advance worksheet package,
- ▶ Procedures to follow if a new firm is suspected of being ineligible,
- ▶ Distinction among certain business types,
- ▶ Definition of various key financial terms,

- ▶ Ways in which respondents can obtain additional information about the study,
- ▶ Sample characteristics, and
- ▶ CATI functions.

Additionally, at the end of the written exam the interviewers' ability to accurately record a large number was tested by asking interviewers to write down the numbers dictated by a supervisor.

The CATI certification interview included three scripted mocks that covered the following scenarios:

- ▶ Breaking-off a case because a firm was ineligible firm (ineligible SIC code) after a proxy had provided information in the screener,
- ▶ Respondent refuses to give institution names, and the interviewer must use a job aid to identify financial institutions with generic names; this type of interview has a break-off at section P, and the interviewer must switch to the off-line questionnaire,
- ▶ Business is a C-Corporation using multiple institutions and multiple services.

These mock interviews all tested probing skills, verbatim recording, using the CATI system, and coding responses.

**5.8.6 Mailing Advance Worksheet Packages.** The sampling statistician divided the sample into replicates so that the sample could be fielded in small batches. Prior to mailing the advance worksheet packages, the project and mail-out center staff worked out a schedule of mailings that would allow all cases mailed to be contacted within a week of having received the package. The worksheet packages were sent via Federal Express two-day delivery. The mailings began on September 15, 1999, and ended on March 17, 2000.

On September 29, 1999, replicates 415 (n=103), 416 (n=97) and 705 (n=50) were mailed but not subsequently released for the interviewers to work. This error was discussed with FRB staff. NORC and FRB staff decided to re-mail the worksheet package and then contact these businesses. On January 17, 2000, the worksheet package was re-mailed to these three replicates and these businesses contacted to complete the interview.

## VI. DATA COLLECTION

Data collection occurred in two phases: a screening phase and an interviewing phase. Prior to screening, a sample of about 40,000 businesses was selected from the Dun and Bradstreet master file, as described in Chapter VIII. The sample was selected so that differences in the use of credit and financial services among firms of differing sizes and minority ownership could be measured. Screener data collection began in mid-June 1999 and was completed at the end of September 1999. Prior to the screening telephone contact, businesses were sent a mailing that explained the study and encouraged participation. As expected, the screener took 4.49 minutes to administer, on average, once the owner or suitable proxy was identified and agreed to answer the screener questions. However, the total time spent contacting businesses and gaining respondent cooperation was twice as long as planned.

Following screening, but before interviewing began, eligible businesses were divided into replicates by minority type and all others. This process is described in Chapter VIII. Telephone supervisors released sample to interviewers in a steady flow. Interviewing began mid-September 1999 and ended in early May 2000. Businesses were first sent a package of materials that further explained the study and included a worksheet for the respondent to complete prior to the telephone contact. The average interview length was actually five minutes shorter, on average, than budgeted. However, as in the screening phase, the total time spent contacting businesses and gaining respondent cooperation was twice as long as planned.

This chapter describes the data collection processes and protocols, details the problems encountered, and describes the outcomes of screening and interviewing.

### 6.1 Screening

The screener verified the name of the business owner and the mailing address of the business, determined eligibility to participate, identified the legal form of organization, recorded the fiscal year end date, and most importantly, determined the ethnic and racial composition of the owners of each small business. Before the initial call to screen the business, NORC sent a mailing that included two letters explaining the purpose of the survey and encouraging participation; one letter was from Alan Greenspan and the other was from the project director (as described in Chapter V). The mailing also included a question and answer brochure. Within a few days of receiving the letter, businesses were called and asked to complete the screener. Interviewers first tried to complete the screener with the business owner before attempting to identify and complete the screener with a suitable proxy respondent.

**6.1.1 Timing and Schedule.** At the time of contract award screening was scheduled to begin in mid-May 1999. During the initial weeks of the period of performance of the contract, NORC and FRB staff reviewed and modified the schedule so that more time was devoted to planning, development and review of materials and protocols. Additional time was also needed to obtain the sample from Dun and Bradstreet. Screener data collection was rescheduled to begin on June 17, 1999, after completion of the screener data collection for both pretests, so that experiences from the pretest could inform instrument development and protocol design. As interviewing staff successfully completed the certification process, they began screening. Sixty-six interviewers from the initial group of trainees were screening businesses by June 24, 1999.

The screener data collection was planned as a two-month activity. However, the screening did not end until September 30, 1999. There were a few system problems system problems experienced early in screener data collection but the major contributing factor to the lengthy time period was the total time spent per completed screener; the total time spent contacting businesses and gaining respondent cooperation was much longer than anticipated. The projected total time spent per completed case actually was almost twice the budgeted amount of time: 20 minutes budgeted vs. 39.2 minutes actual.

**6.1.2 Protocol Changes.** Two protocol changes were made during screener data collection: the number of calls attempted to reach the business owner before accepting an interview with a proxy respondent and selecting the sample at two points in time instead of once, at the end of screener data collection.

- ▶ *Calls to owner.* Shortly after screener data collection began, the COTR requested that interviewers make up to four calls to reach the business owner before relying on a proxy to complete the screener, instead of leaving the number of calls at the discretion of the interviewer. Three weeks after making this change, the telephone center requested that the number of calls be changed from four to three, as the interviewers felt that four calls were too many. The COTR approved reducing the number of calls to the owner from four to three.
- ▶ *Sample selection.* Initially, NORC planned to screen all respondents, then select the sample, and conduct interviews. Due to the longer than anticipated screener data collection period, NORC and the FRB decided to select the sample at two separate points in time: the first selection would

take place after the initial half of the screening had been completed, and the second selection would take place after the second half of the screening had been completed. This revised plan had several drawbacks, such as managing two processes concurrently and complicating the sample selection process, but it allowed NORC to keep both training and data collection for the main questionnaire as close to the scheduled dates as possible.

**6.1.3 Refusal Conversion.** Interviewers with strong refusal conversion skills were selected to call those businesses that initially refused to be screened. The basic refusal conversion strategy was to call the business again, determine why the business would not participate, address any concerns, and then attempt to administer the screener. In some cases, the interviewer thought the business was refusing to respond because it may have been ineligible. For instance, interviewers recorded statements such as, “this may be a not-for-profit” or “sounds like a government office.” For cases like these, refusal converters asked the appropriate eligibility screening question immediately after confirming that they had the correct business, so that they did not waste time convincing an ineligible business to complete a screener.

Often, by the time a refusal converter called the business, the advance letter had been discarded. When the owner or a proxy requested that another letter be sent to help the business owner decide whether or not to participate, the interviewer asked if the letter could be faxed, instead of mailed, to the business. For those businesses that did not have a fax machine or preferred to receive the letter via U.S. mail, the letter was mailed. Requests to either fax or re-mail a letter were fulfilled within 24 hours. Businesses that received the letter via fax were called same day that the letter was faxed. Businesses that were mailed a letter were called within four days of the letter being mailed.

Of the 39,240 businesses in the screener sample, 5,982 (15.2%) refused to participate at some point during screener data collection. Refusal conversion specialists were able to convince 2,683 (45%) of these businesses to reconsider and complete a screener.

**6.1.4 Locating.** Unlocatable cases were identified in two ways: through mail that was returned as “undeliverable as addressed” and as bad telephone numbers. Locators attempted to find a telephone number and new address for those businesses from which we received an undeliverable letter without an address

update, and for those businesses with bad telephone numbers. Locators called directory assistance in the area the business was last known to reside and asked for a telephone number for that business. Locators also conducted Internet searches and called the local Chamber of Commerce to locate the business.

Many of the cases that were initially unlocatable, but subsequently were found, turned out to be ineligible. Of those cases that were unlocatable and subsequently found, 66 became completed screeners. There were 3,413 cases finalized as locating problems at the close of screener data collection.

**6.1.5 Monitoring.** NORC supervisory and project staff monitored screening activity throughout the course of screener data collection. During this effort, NORC supervisors conducted 2,370 sessions of monitoring interviewer screening activity. Monitoring included observing interviewer time spent gaining cooperation, contacting, locating, and refusal conversion, as well as conducting the screener.

The kinds of problems observed while monitoring screener data collection included:

- ▶ Not using surnames when addressing office staff or the business owner; interviewers were expected to always use surnames except when requested otherwise.
- ▶ Not verifying the spelling of the name of the business owner or the street and city names; getting accurate spelling of any new information about the business was essential.
- ▶ Not probing for fiscal year end date when respondents reported that date as April 15; this date informed the type of worksheet that we would send to the business. Therefore, this was critical information.

Observations made while monitoring were shared with the interviewers during one-on-one and group meetings.

**6.1.6 Interviewer misconduct.** During a regular monitoring session of the screening effort, we discovered that one of the interviewers had not made all the calls (and/or asked all the questions) he or she claimed to have done. The interviewer was terminated immediately and the COTR was both appraised of the incident and the plan to remedy the situation. This plan involved contacting all respondents screened by this interviewer to determine if the business had been

called by the interviewer, and if so, to reask a subset of the questions as a check that all of the questions had been asked. If the interviewer had falsified the contact, the business would be screened at the time of the re-contact. The COTR agreed to the plan, and the plan was implemented.

Of the 482 screeners completed by that interviewer, 406 were screened again; we were not able to reach 76 of the cases. The 76 cases that could not be reached were changed from a complete to a not-screened outcome and the associated data were removed from the CATI. Forty-two of the 482 cases were associated with the Screener 1 sample, the first half of the sample that had been completed and closed out in early August for interviewing in September. Of these 42 screeners completed, 19 had been selected for the main interview sample. We were able to screen all of the 19 cases again and learned that they were all eligible.

**6.1.7 Receipt Control Activity.** Throughout screener data collection, advance letters were returned as undeliverable. The following table shows the number of letters returned during this period.

**Table 6.1 Number of Returned Letters by Week**

<b>Week Ending</b>	<b>Advance Letter Return</b>
6/19/99	0
6/26/99	0
7/3/99	0
7/10/99	0
7/17/99	687
7/24/99	200
7/31/99	1190
8/7/99	1542
8/14/99	161
8/21/99	48
8/28/99	28
9/4/99	0
9/11/99	0
9/18/99	0
9/25/99	0
10/2/99	0
<b>Total</b>	<b>3856</b>



**6.1.8 Screening Level of Effort.** NORC obtained 39,240 cases for the screening effort. However, 13 cases overlapped with the cases worked in the pretest effort. Therefore, NORC actually released and worked 39,227 cases during the screening phase of the project.

The total number of calls made to the 39,227 cases was 269,567, for an average of 6.9 calls per case. The details of this level of effort information by final outcome disposition can be found in Table 6.2 below.

**Table 6.2 SSBF '98 Screening Level of Effort by Outcome**

Final Disposition	Description	Number of Cases	Number of Calls	Average Calls Per Case
00	Overlap Case with Pretest; Not Released	13	NA	NA
11	Final Non-contact; unconfirmed phone number	680	15,211	22.4
12	Final Locating Problem	3,413	13,944	4.1
21	Final Unavailable During Field Period	1,339	28,027	20.9
22	Final Incapacitated Respondent	7	48	6.9
23	Final Non-contact; phone number confirmed	782	18,030	23.1
24	Final Language Barrier	53	328	6.2
31	Final Refusal	5,913	47,147	8.0
32	Final Hostile Refusal	40	236	5.9
64	Complete; DK Eligible per Proxy	120	810	6.8
66	Complete; RF Eligible per Proxy	13	120	9.2
67	Complete; Eligible per Owner	13,355	71,682	5.4
68	Complete; Eligible per Proxy	6,294	40,470	6.4
71	Complete; Ineligible per Owner	1,957	9,688	5.0
72	Complete; Ineligible per Proxy	2,280	11,177	4.9
73	Complete; DK Eligible per Owner	161	853	5.3
75	Complete; RF Eligible per Owner	34	157	4.6
79	Complete; Ineligible Tax Filing for Partnership	31	127	4.1
81	Intro Ineligible; not in business in Dec 98	614	3,326	5.4
82	Intro Ineligible; not headquarters/main office	806	2,899	3.6
83	Intro Ineligible; subsidiary	177	709	4.0
84	Intro Ineligible; not-for-profit	612	2,346	3.8

**Table 6.2 SSBF '98 Screening Level of Effort by Outcome**

Final Disposition	Description	Number of Cases	Number of Calls	Average Calls Per Case
85	Intro Ineligible; government owned	231	689	3.0
87	Intro Ineligible; 500 or more employees	161	629	3.9
88	Intro Ineligible; ineligible business type	154	914	5.9
<b>Overall</b>		<b>39,240</b>	<b>269,567</b>	<b>6.9</b>

**6.1.9 Completion Rate / Response Rate.** When screening ended on September 30, 1999, NORC interviewers had completed 26,998 screeners in 17,658.3 hours, for an average of 39.2 minutes per completed screener. Table 6.3 reflects the total production for screener data collection:

**Table 6.3 Screener Production, Hours Worked, and Minutes/Screener**

Completed Screeners	Hours Worked	Minutes/Screener
??	??	<b>ERR</b>

Interviewers completed screeners with 69% of the sample; each half of the screener sample reached the same level of completeness. The final status of the screening sample was as follows:

**Table 6.4 Final Status of First and Second Half of Screener Sample**

Version	Replicates	Complete	Intro Ineligible	Total Complete	% Complete
SCR1	1-200	12,268	1,555	13,823	69.15
SCR2	201-393	11,975	1,200	13,175	68.50
<b>Total</b>	<b>1-393</b>	<b>24,243</b>	<b>2,755</b>	<b>26,998</b>	<b>68.83</b>

**6.1.10 Eligibility Rate.** Of the 26,998 firms screened, 19,782 (73%) met the eligibility criteria and were included in the pool of cases from which the sample for the main data collection would be drawn.

A breakdown of the size of the eligible firms can be found in Table 6.5. A much greater number of firms were in the smallest size class (0-4 workers), with fewer firms in each subsequent larger size class, with the one exception. The very largest size class (100-499) has more cases in it than does the next smallest size class (50-99).

**Table 6.5 Number of Firms by Size of Workforce**

Number of Workers	0-4 Workers	5-9 Workers	10-19 Workers	20-49 Workers	50-99 Workers	100-499 Workers	Eligible Unknown Size	Don't Know/Refused
Number of Firms	9,875	3,664	1,916	1,758	1,195	1,310	23	41

**6.1.11 Non-response.** Of the 39,227 cases released for screening, NORC was unable to screen 12,227 cases. Many contact attempts were made on these problematic cases and the average number of calls per case for these non-responders was 10.1. Almost twice as many calls were made to non-responding businesses as were made to screened businesses (average 10.1 calls versus average 5.4 calls). Despite considerable effort, some business were never located, and some businesses simply refused to cooperate. Other businesses were never available for screening during the field period, (such as seasonal businesses), and the telephone at some businesses was always answered by a machine.

**6.1.12 Non-response Follow-up.** In late December 1999 we launched a follow-up effort to determine the eligibility of the non-responding businesses. We expected data collection for this task to be completed by the end of January 2000; instead we completed data collection on March 31. The field period was extended to allow extraordinary locating efforts and extra time to encourage businesses to disclose the information we sought.

The first steps taken to accomplish the follow-up effort were the selection of a sample and development of an instrument and data collection protocol. Our sampling statistician selected a proportionately stratified random sample of 5% of the non-responding businesses from the screener sample. The sampling procedures are described in Chapter VIII. Concurrently a questionnaire and a data collection protocol were developed.

We anticipated that some of these businesses might be more inclined to give us the screening information if we mailed the screener questions. Therefore, the questionnaire was designed for administration by telephone interviewers or as a self-administered form that we could mail or send via fax to the selected business, on request. A copy of the questionnaire used for the 5% follow-up sample can be found in Appendix W.

We used three different modes for collecting data from the 5% follow-up sample: mail, telephone and E-mail. Each method is described below.

**Mail.** We sent a package containing the following items to those firms that were Final Non-Contact Phone-Number Confirmed (23), Final Refusal (31), and Final Hostile Refusal (32):

- ▶ A letter requesting that the business owner complete a brief questionnaire
- ▶ A brochure describing NORC
- ▶ The Federal Reserve System Purposes and Functions (only to those businesses that refused to be screened)
- ▶ An SSBF FAQ sheet (modified)
- ▶ An abbreviated screener designed to determine eligibility
- ▶ Instructions for E-mail response
- ▶ A \$2 bill as a token of our appreciation
- ▶ Business return envelope

There were two versions of the letter: one for businesses that refused to be interviewed and another for the businesses we could never reach. For the mailings, we modeled our approach on the *Mail and Telephone Surveys: Total Design Method* (Dillman, 1978 ) using the following protocol:

- ▶ Mail the package.
- ▶ Two weeks later, mail a postcard reminder.
- ▶ Two weeks later, re-mail entire package without the \$2.00.
- ▶ Two weeks later, make telephone contact to collect the abbreviated screener information over the telephone.

**Telephone.** For those cases that were locating problems with unconfirmed telephone numbers, we immediately began locating efforts via telephone. We also made telephone contact, with cases that could not be completed during the original field period, language barrier cases, and cases where an owner provided answers to some but not all of the key eligibility questions.

In all telephone contacts, the interviewer utilized a face sheet designed for each case in the subsample. Included on the face sheet were the questions on the abbreviated questionnaire that had been mailed to the businesses to facilitate data collection during telephone contacts with respondents. A sample of the face sheet can be found in Appendix X.

***E-mail.*** In addition to the hard copy and telephone options, the mailout packet included instructions on how to respond to the screener via e-mail.

NORC captured the outcome of the cases worked by all three modes in a simple Paradox database, using a graphical interface to data enter the results. The database included cases that did and did not respond. For purposes of weighting and other sample analysis, this database was delivered to the statistician for applying the results of the 5% subsample to the balance of the unscreened sample. The results for the 5% subsample are described in chapter VIII.

**6.1.13 Other basic statistics.** NORC completed screeners with 26,998 businesses. One-fifth (21%) of the cases were completed with just one telephone call. Almost half (47%) of the completed cases required three or fewer calls. Most of the cases (85%) required 10 or fewer calls. The average number of calls to complete a case across all complete dispositions is 5.4. The distribution of number of calls to complete the screener can be found in Appendix Y.

A more detailed breakdown by result code for screened and unscreened cases by level of effort, or average number of calls per case, is shown in Table 6.6. The data indicates that final “non-contact” cases and undeliverables received the highest average number of calls per case (about 22 calls), while the other categories of cases received under 10 calls per case on average.

**TABLE 6.6 SSBF FINAL SCREENING OUTCOME**

<b>Final Screener Disp Code</b>	<b>Final Screener Disposition Outcome Description</b>	<b>Case Status</b>	<b>Number of Cases</b>	<b>Percent of Cases</b>	<b>Average Number of Calls</b>
00	Overlap Case with Pretest; Not Released	Ineligible	13	0.03%	NA
11	Final Non-contact; unconfirmed phone number	Unknown	680	1.73%	22.4
12	Final Locating Problem	Unknown	3413	8.70%	4.1
21	Final Unavailable During Field Period	NIR	1339	3.41%	20.9
22	Final Incapacitated Respondent	NIR	7	0.02%	6.9
23	Final Non-contact; phone number confirmed	NIR	782	1.99%	23.1
24	Final Language Barrier	NIR	53	0.14%	6.2
31	Final Refusal	NIR	5913	15.07%	8.0
32	Final Hostile Refusal	NIR	40	0.10%	5.9
64	Complete; DK Eligible per Proxy	Complete	120	0.31%	6.8
66	Complete; RF Eligible per Proxy	Complete	13	0.03%	9.2
67	Complete; Eligible per Owner	Complete	13355	34.03%	5.4
68	Complete; Eligible per Proxy	Complete	6294	16.04%	6.4
71	Complete; Ineligible per Owner	Ineligible	1957	4.99%	5.0
72	Complete; Ineligible per Proxy	Ineligible	2280	5.81%	4.9
73	Complete; DK Eligible per Owner	Ineligible	161	0.41%	5.3
75	Complete; RF Eligible per Owner	Ineligible	34	0.09%	4.6
79	Complete; Ineligible Tax Filing for Partnership	Ineligible	31	0.08%	4.1
81	Intro Ineligible; not in business in Dec 98	Ineligible	614	1.56%	5.4
82	Intro Ineligible; not headquarters/main office	Ineligible	806	2.05%	3.6
83	Intro Ineligible; subsidiary	Ineligible	177	0.45%	4.0
84	Intro Ineligible; not-for-profit	Ineligible	612	1.56%	3.8
85	Intro Ineligible; government owned	Ineligible	231	0.59%	3.0
87	Intro Ineligible; 500 or more employees	Ineligible	161	0.41%	3.9
88	Intro Ineligible; ineligible business type	Ineligible	154	0.39%	5.9
<b>Total</b>			<b>39240</b>	<b>100.00%</b>	<b>6.9</b>

The level of effort for each racial or ethnic group is displayed in Table 6.7. The overall number of calls per eligible case was about six; this number is fairly consistent for all racial or ethnic groups. The only exception was “Black and

Asian, Non-Hispanic;” that category received about 3 calls per eligible case on average.

**Table 6.7 Level of Effort by Minority Status**

Racial/Ethnic Group	Number of Eligible Cases	Average Calls per Eligible Case
White, Non-Hispanic	16,891	5.63
Hispanic & Any Race	1,017	6.51
Asian, Non-Hispanic & Non-Black	800	6.13
Black, Non-Hispanic & Non-Asian	619	6.24
Black & Asian, Non-Hispanic	4	3.25
Native American	175	5.53
Don't Know / Refused	276	5.95
<b>Total</b>	<b>19,782</b>	<b>5.72</b>

Table 6.8 indicates the number of eligible cases in each cell by race/ethnicity and firm type.

**Table 6.8 Number of Eligible Cases by Race/Ethnicity and Firm Type.**

Racial/ Ethnic Group	Number of Eligible Cases											
	Sole Prop	Partner -ship	LLP as Partner	LLP as C-Corp	S-Corp	C-Corp	LLC as Partner	LLC as C-Corp	LLC as Sole Prop	DK	RF	Total
White, Non-Hispanic	6,901	1,073	323	244	3,455	4,420	105	18	6	346	0	16,891
Hispanic & Any Race	553	66	15	12	134	211	2	0	0	24	0	1,017
Asian, Non-Hispanic & Non-Black	399	44	6	12	110	207	3	0	0	19	0	800
Black, Non-Hispanic & Non-Asian	364	42	6	4	71	120	1	0	0	11	0	619
Black & Asian, Non-Hispanic	1	2	0	0	0	1	0	0	0	0	0	4

Native American	103	7	1	0	32	32	0	0	0	0	0	175
Don't Know Refused	120	15	1	4	27	79	1	0	0	26	3	276
<b>Total</b>	<b>8,441</b>	<b>1,249</b>	<b>352</b>	<b>276</b>	<b>3,829</b>	<b>5,070</b>	<b>112</b>	<b>18</b>	<b>6</b>	<b>426</b>	<b>3</b>	<b>19,782</b>

## 6.2 Main Data Collection

The main questionnaire interview collected information on the following: income and expenses; assets and liabilities; loans, equity, checking accounts, credit cards, and other sources of outstanding funding; other financial services such as check clearing, credit card processing, brokerage services, trade credit, equity injections; and recent credit acquisition experiences. Prior to calling businesses to complete the 40-minute interview, NORC mailed an information package to the business owner that included letters, brochures, and interesting facts from the prior round of data collection. If a proxy had completed the screener in place of the owner, a thank-you post card was sent to the proxy respondent at the same time the package was mailed to the business owner. Within a few days of the mailing, businesses were called by telephone interviewers to answer any questions about the survey and to complete or make an appointment to complete the interview. Like the screener interviews the time required to complete questionnaire data collection was about twice as long as planned.

**6.2.1 Timing and Schedule.** The field period for questionnaire data collection was scheduled to start in mid-August 1999 and end on December 31, 1999. The actual first day of data collection was September 20, 1999; data collection ended on May 6, 2000.

**6.2.2 Special Efforts to Increase Production.** From the beginning of main questionnaire data collection, we made continuing efforts to increase production. Special initiatives included (listed in chronological order):

- ▶ Adding more interviewer staff.
- ▶ Interviewer meetings.
- ▶ Special meetings.
- ▶ Interviewer bonus plan.
- ▶ Assistance from field staff.
- ▶ Respondent incentives.



Each of these initiatives is discussed below:

***Adding more interviewers.*** Table 6.18 reflects the production for each week of data collection. The budgeted hours per completed case was 4.7, but the overall actual hours per completed case was 6.6. Early in the field period we could see that additional staff was needed to reach the desired goal. As described in Chapter VI, between September 15 and November 18, NORC completed a total of five training sessions, three more than originally planned.

***Interviewer meetings.*** NORC supervisory staff conducted weekly meetings with the interviewers to give them an update on production, to discuss strategies to gain cooperation, and to share observations from monitoring sessions, as well as address administrative matters. The meetings were typically held on Friday because that was the least productive day of the week. The meeting time was scheduled during the period when interviewer schedules overlapped, so all interviewers had an opportunity to attend.

During these meetings interviewers shared the experiences they were having in gaining the cooperation of gatekeepers and business owners. Group leaders acknowledged the difficulty of the interview and focused on the successful contacts that became completed interviews. Each interviewer was encouraged to describe strategies that led to completed interviews. Staff were encouraged to continue to try new phrases and other strategies to gain respondent cooperation.

Each supervisor was assigned 7 - 10 interviewers to mentor. Each week every interviewer had a one-on-one session with their supervisor in which observations from the monitoring sessions were shared and the individual's production was compared to the group average. During this session interviewers also practiced gaining-cooperation skills in mock situations.

***Special meetings.*** In other meetings senior staff from NORC spoke with the interviewers about the importance of the SSBF project to the company, and the importance of the interviewers to the success of the project.

***Interviewer bonus plan.*** The initial bonus plan was for a one-time bonus for a given end date, with set parameters. Interviewers were told the following:

“To be eligible for the bonus, the team must work 85% of the scheduled hours, come within 10% of the equivalent of an average of 7 cases per week per person, and meet the per case completeness criteria (no more than three consecutive

missing values per completed case).”

This bonus plan was too ambitious, given that no one reached the goal and no one received the bonus payment. Therefore, the initial plan was discontinued in January 2000 and another bonus plan instituted in March, was based solely on consistently high attendance rates by interviewers. This plan was based on the premise that there was ample virgin sample for release so that higher attendance rates by the interviewers should result in higher production rates. The plan was comprised of two main elements:

- ▶ Interviewers who worked all scheduled hours for a week merited a \$10 payment.
- ▶ Two consecutive 100% attendance weeks merited a premium payment.

This bonus plan resulted in approximately \$2,175. in bonuses paid.

*Assistance from field staff.* Another method NORC tried to boost production was adding a small cadre of field staff to gain the cooperation of respondents who had decided not to participate. This effort began in mid-December 1999 as a trial, with six field interviewers. The field interviewers were trained in late December as gaining-cooperation specialists. Project staff prepared hard-copy face sheets, a form with basic information about each case (such as business name, address, telephone number, and name of the business owner), for each of the 750 cases selected, and sent them to the field interviewers along with some of the materials that were used for telephone interviewer training and respondent mailings. The cases were divided among the six field staff - first by geographic location, then to even out the work load. The call records accompanied the face sheets for the assigned cases. The call notes for these cases were also included in the assignments sent to the field staff to aid in conversion efforts.

A special protocol was established for interaction between field and telephone center staff. When a field interviewer convinced a respondent to participate, the field interviewer first completed a refusal conversion form for their own records. The second step was to call the 1-800 number at the Lake Park Telephone Center to speak with the field liaison. Third, the field liaison at the telephone center completed and filed a copy of the refusal conversion form, entered the information about the appointment into TNMS/CATI, and completed an appointment card.

If the new converted respondent that had just been converted and wanted to

proceed with the interview immediately, the field liaison would transfer the field interviewer through to an available telephone interviewer.

Project staff held five meetings with the field staff, between December 16, 1999 and January 21, 2000. The field interviewers thought it best to wait until after the Christmas/New Year holiday before starting work on the cases. After Christmas field staff were very slow to begin working the cases and they indicated that the cases they did call were very uncooperative. On January 21<sup>st</sup> the refusal conversion trial was terminated. None of the cases worked by field staff had turned into a completed case, and the field staff were very frustrated with the task. The experience of the field interviewers validated the experience of the telephone interviewers; respondents were uncooperative.

Field interviewers were instructed to return all project materials, including folders and the enclosed documents, refusal conversion letters, extra worksheets, forms for conversion documentation, and FRB books to the NORC mail-out center. They were instructed likewise to return face sheets and any other documentation of pending action on cases to the telephone center. Also they reported their final accumulation of hours and expenses to the supervisory staff for record keeping. Other materials could be discarded, as they decided.

**Respondent Incentives.** In January 2000, NORC suggested using monetary incentives to encourage survey participation. The FRB recommended a Pre- and Post-Incentive Trial. NORC examined the status of the sample at that time and reported that there were 5,665 cases released in 1999, of which 1,643 were not complete or finalized and had had 14 or few calls. Of the cases that were not complete, 399 had some data.

NORC recommended a post-incentive payment of \$20 and selected 77 cases for the post-incentive trial. NORC also recommended a pre-incentive of \$20 to cases that we had not yet contacted. The FRB requested that we identify a control group for each of these two trial groups. Table 6.10 shows number of cases assigned to each group.

**Table 6.9 Incentive Trial**

Pre-Incentive		Post-Incentive	
Control	Treatment	Control	Treatment
1100	200	77	77

Tables 6.10 and 6.11 depict the final results of the incentive trial.

**Table 6.10 Status of Post-Incentive as of 5/6/2000**

Strata	Control			Treatment		
	# Complete	Total Cases	% Complete	# Complete	Total Cases	% Complete
Hispanic	2	6	33.3%	3	13	23.1%
Asian	2	9	22.2%	1	9	11.1%
Afr. Am.	5	11	45.5%	2	7	28.6%
Other 0-19	3	30	10.0%	7	33	21.2%
Other 20-49	0	8	0.0%	2	7	28.6%
Other 50-99	1	8	12.5%	1	4	25.0%
Other 100-499	0	5	0.0%	0	4	0.0%
<b>Overall</b>	<b>13</b>	<b>77</b>	<b>16.9%</b>	<b>16</b>	<b>77</b>	<b>20.8%</b>

**Table 6.11 Status of Pre-Incentive as of 5/6/2000**

Strata	Control			Treatment		
	# Complete	Total Cases	% Complete	# Complete	Total Cases	% Complete
Hispanic	7	25	28.0%	7	25	28.0%
Asian	18	65	27.7%	17	35	48.6%
Afr. Am.	8	25	32.0%	11	25	44.0%
Other 0-19	200	630	31.7%	22	69	31.9%
Other 20-49	31	87	35.6%	4	14	28.6%
Other 50-99	41	132	31.1%	4	18	22.2%
Other 100-499	34	136	25.0%	2	14	14.3%
<b>Overall</b>	<b>339</b>	<b>1100</b>	<b>30.8%</b>	<b>67</b>	<b>200</b>	<b>33.5%</b>

The results of the use of incentives is discussed in subsection 6.2.12 of this chapter.

**6.2.3 Extended Field Period for Data Collection.** In January, 2000 NORC and FRB staff met to review the current status of the project and consider possible changes in strategy which might boost the pace and level of production, which was lagging. A decision was made to extend the field period and offer incentives

to the respondents. A lower production target (n=3,250 completed cases) was agreed upon. (Details of the revised implementation plan are described later in this chapter.)

**6.2.4 Refusal Conversion.** Telephone center staff carefully reviewed cases that refused to participate and classified them into one of the following categories:

- ▶ Company has a blanket policy against participation.
- ▶ Taxes or government restrictions or regulations have harmed the business.
- ▶ Owner or suitable proxy is too busy.
- ▶ Government already has the information we are seeking.
- ▶ Information we are seeking is too intrusive.
- ▶ Business owner does not trust the government.

NORC and FRB staff drafted three refusal conversion letters. One letter addressed the reason the FRB is collecting these data even though another branch of the federal government (the IRS) already has the data. This letter was sent on FRB letterhead to businesses that complained about the intrusive nature of the data we were collecting, those that said they already reported these data to the government, and those that had a blanket policy against participation. The second letter addressed the FRB's need to talk with all types of businesses, even those that were too busy to participate. The third letter addressed anti-government sentiment. The second two letters were sent on NORC letterhead. A copy of each of the refusal conversion letters can be found in Appendix Z.

Of 11,053 cases released, 5,162 were referred for refusal conversion at some point during data collection. These were very difficult cases to convert; of the 5,162 refusals, 285 cases were converted successfully and completed an interview.

**6.2.5 Locating.** Of the 11,053 cases released, 419 cases were referred for locating. A case was identified as needing locating in two ways: the advance package was returned as undeliverable as addressed or the telephone was no longer in service. Locators took the following steps to try to locate the business:

- ▶ Called Directory Assistance to get a new business telephone number.
- ▶ Conducted searches to look for similar businesses in the area of the last known location of the business so that we could contact them to ask if they knew the present whereabouts of the missing company.

- ▶ Conducted on-line searches for local Chamber of Commerce offices or called these offices to see if the business was listed.
- ▶ Called the County Tax Assessor.
- ▶ Called the State licensing bureau.
- ▶ Called Directory Assistance to get the residential telephone number of the business owner.

Of those 419 cases, an interview was completed with 42 cases. In most of the remaining cases the unlocatable business appeared to have closed its operations and was not in business elsewhere, and we were unable to locate the firm's owner.

**6.2.6 Protocol for Working Minority Cases.** Because NORC anticipated that gaining cooperation from minority firms (African American, Asian, and Hispanic) might prove especially difficult, NORC took special steps, in particular:

- ▶ Offered bilingual capability in Korean, Mandarin, Cantonese and Spanish languages. Businesses with owners who required a foreign language speaking interviewer in one of these languages were accommodated.
- ▶ Included in the package of materials mailed to minority owned firms, a research article about minority business' access to credit. In addition to mailing this article to African American, Asian, and Hispanic owned businesses, NORC also sent it to firms that were 50% African American, Asian and Hispanic owned. Additionally, the article was sent to businesses owned by Native Americans.
- ▶ Assigned interviewers particularly adept at gaining cooperation to the minority cases. As interviewing progressed NORC continually assessed how interviewers were performing and reassigned interviewers as needed.
- ▶ Separated the three minority group businesses in unique locations in the TNMS so that these cases could be more easily targeted for special treatment when necessary.
- ▶ Tracked the strata and quantity of cases worked by each interviewer. About halfway through the interview period, NORC determined that only those interviewers assigned to the Hispanic queue had a disproportional number of completed cases.

- ▶ Made more call attempts to minority owned businesses than non-minority businesses.

As an appendix to this report we have included a table that reflects the cases completed by the interviewers assigned to the minority queues. The interviewers also were assigned to cases in other strata. The supervisory staff attempted to balance the assignment of cases worked by the minority interviewers, however, the bilingual Spanish language interviewers completed more cases in the Hispanic queue than in the non-minority queues.

**6.2.7 CATI Changes.** Changes were made to the CATI program after data collection began. The following list lists the types of changes and an example of each:

- ▶ *Change in allowable responses.* When reviewing the frequencies we noticed that the interviewers sometimes filled the four digit extension of the zip code or the telephone number field with zeros instead of indicating “refused” or “don’t know.” The programmer changed this field so that a response of all zeros was not allowed.
- ▶ *Text changes.* The interviewers reported that some respondents found it unsettling when asked “Is the mailing address still . . .” indicating that we had their mailing address. Instead we changed the text to read “Is the mailing address . . .”
- ▶ *Skip logic.* Sometimes the demographic questions in Section C were being skipped in error for one of the firm types; the question logic was corrected.
- ▶ *Flag construction.* The code written to construct flags after trust services and brokerage services was in error and was corrected.

All the changes made to the CATI instrument after data collection began are documented in Appendix AA.

**6.2.8 Monitoring.** Telephone supervisory and project staff monitored interviewing activity throughout the course of main data collection, conducting 3,764 sessions. In addition to monitoring interviews in progress, such activity included observing interviewer time spent gaining cooperation, contacting,

locating, and converting refusals.

NORC facilitated a linkage with FRB staff, so that they could monitor interviewing in real-time from their offices in Washington, DC. The FRB used laptops loaded with specialized software to engage in both aural and visual monitoring of the interactions between the respondent and interviewer as interviews were in progress. The FRB documented their observations in an Excel spread sheet; this database was updated frequently and sent to NORC.

The problems observed during monitoring included the following:

- ▶ Failure to read back spelling for institution names, addresses, etc.
- ▶ Failure to read back dollar amounts.
- ▶ Failure to read text exactly as it appeared on the screen.
- ▶ Recording zeros in the ZIP code field when the respondent didn't know the ZIP code.
- ▶ Recording artificial fillers, such as 11111, 00000, 99999 for institution addresses.
- ▶ Not adequately answering questions about the background/purpose of the study or the confidentiality policy.
- ▶ Not adequately describing financial terms.
- ▶ Not using the worksheet line and tax form line references when necessary.
- ▶ Need to improve pacing of interview / voice quality.
- ▶ Need better mastery of the data collection protocols, such as recording fake institutions and using break-off questionnaires.
- ▶ Need to review definitions of financial concepts.
- ▶ Need to review recording of large numbers.
- ▶ Recording unnecessary margin notes.
- ▶ Need to review meaning of certain bank titles, e.g., "national" in the bank name means a commercial bank.

Overall, interviewers did a good job handling difficult respondents and administering a long and complex questionnaire. Feedback from both in-house and FRB monitoring sessions was shared with the interviewing staff during one-on-one and weekly meetings.

**6.2.9 Break-offs.** Some respondents who completed the worksheet prior to our call objected to answering again the questions covered on the worksheet. When this happened, interviewers were instructed to complete a “break-off”



questionnaire in hard copy that collected information not included on the worksheet. The partial CATI data for 313 cases was supplemented with break-off questionnaire and/or completed worksheet data. A more complete description of this phase of data collection is found in Chapter VII.

Table 6.12 shows the number of cases for which NORC obtained hard copy materials (break-off questionnaires and/or completed worksheets), and which of those materials were used in completing cases.

**Table 6.12 Hardcopy Materials Used to Complete Case Information**

<b>Worksheet Status</b>	<b>Break-off Quex Status</b>	<b>Hard Copy Materials Use Status</b>	<b>Number of Cases</b>
No Worksheet	No Break-off Quex	Hard Copy Documents not available	9,006
	Break-off Quex	Hard Copy Documents Not Used	13
		Break-Off Quex Used	100
<b>Total Number of Cases without a Worksheet</b>			<b>9,119</b>
Complete Worksheet	No Break-off Quex	Hard Copy Documents Not Used	1,712
		Worksheet Used	12
	Break-off Quex	Hard Copy Documents Not Used	16
		Break-Off Quex Used	168
<b>Total Number of Cases with a Complete Worksheet</b>			<b>1,908</b>
Partial/Tax Forms	No Break-off Quex	Hard Copy Documents Not Used	30
	Break-off Quex	Break-Off Quex/Partial Worksheet or Tax Forms Used	2
<b>Total Number of Cases with a Partial Worksheet or Tax Form</b>			<b>32</b>

**6.2.10 Sample Release.** Rather than release the entire sample to interviewers at the beginning of data collection, cases were released in batches throughout the data collection period. As cases were completed, or finalized in some other way, new batches were released to be worked. In February 2000, the decision was made to release an additional 3,500 cases in an effort to achieve new targeted completion goals. Due to replicate composition and size, NORC released a total of 11,053 cases. Table 6.13 shows the release date and number of cases released.

**Table 6.13 Release Dates and Sample Size**

Release Date	# of Cases Released
9/15/99	310
9/17/99	356
9/24/99	991
9/29/99	856
10/1/99	699
10/6/99	898
10/12/99	900
11/19/99	148
11/26/99	501
1/17/00	600
1/31/00	1300
2/28/00	447
3/2/00	449
3/7/00	1297
3/20/00	1301
<b>Overall</b>	<b>11053</b>

**6.2.11 Examination of Completed Cases by Interviewer.** In the first week of March 2000, NORC tabulated interviewer association with cases by census division, urbanicity, state, SIC code, and firm type, as well as minority status, sex, and education level of the owner. The tabulation showed that, except for those interviewers assigned to the Hispanic queue, no single interviewer handled a disproportionate number of respondents in any of the categories examined.

**6.2.12 Respondent Incentives.** In February 2000 the FRB agreed to fund three different incentive treatments:

1. A \$20 pre-incentive to each business in cases released on or after 2/28/2000;
2. A post-incentive for businesses that have provided some, but not all of the data (partial completes); and
3. A post-incentive to business owners that need to pay their accountants to provide the data to us.

At that time, 3,449 additional cases were going to be released in 2000, for a total sample size of 11,053. NORC mailed a \$20 check to all new cases with the worksheet mailing package.

The overall level of effort for the pre- and post-incentives was measured by the average number of telephone calls per case. Table 6.14 shows that the average number of calls necessary per case was highest for those cases that received the post-incentive, but lower for those cases that received no incentive and those cases that received a pre-incentive.

**Table 6.14 Level of Effort by Case Status and Incentive Status**

Case Status	No Incentive			Pre-Incentive			Post-Incentive		
	# of Cases	Average Calls	% Complete	# of Cases	Average Calls	% Complete	# of Cases	Average Calls	% Complete
Complete	2369	15	32.5	1111	10	31.8	16	30	20.8
Partial	870	26	11.9	206	15	5.9	10	43	13.0
Non-interview	4043	22	55.5	2177	14	62.3	51	31	66.2
<b>Overall</b>	<b>7282</b>	<b>21</b>	<b>100</b>	<b>3494</b>	<b>13</b>	<b>100</b>	<b>77</b>	<b>35</b>	<b>100</b>

Note: The 200 pre-incentive cases from the incentive trial are not reflected in this table.

The completion rate for the no-incentive group is comparable to the rate for the pre-incentive group. However, we do see a reduced level of effort expended on those cases that received a pre-incentive when compared to those businesses that did not receive any type of incentive. Therefore, while incentives did not increase the response rate, they cut the time needed to reach that rate.

When we compare the post-incentive group with the pre-incentive and no-incentive groups we see that the level of effort is much greater but the completion rate is much lower. We expected the post-incentive group to require more effort than the pre-incentive group, but were surprised to see that the post-incentive group required more calls than the group receiving no incentives. We were also surprised to see that the post-incentive group had a completion rate significantly lower than the no-incentive group.

**6.2.12.1 Discretionary Incentives.** In addition to the pre- and post-incentives listed above, NORC and FRB jointly agreed upon additional incentive measures. These included a discretionary fund to be used in

certain interview situations.

- ▶ *Accountant Time Reimbursement.* Businesses that were not willing to pay their accountant's to look up and report the data we sought were offered reimbursement for the accountant's time, provided the fee was \$125 or less.
- ▶ *Partial Data.* For cases without worksheets for which we had some data and had placed in a pending disposition, NORC offered the business up to \$100 to complete the interview. There were about 225 cases in this category. We did not recommend contacting cases with partial data that were final refusals; there were about 330 cases of that type. NORC also offered up to \$100 to any business that had sent in a complete worksheet, regardless of any other case status information. There were about 40 cases in this category.

**6.2.12.2 Incentives Offered.** NORC offered an incentive of \$20 to 3,571 businesses and offered various amounts of discretionary incentives to 250 businesses. Of the 3,571 businesses to which NORC mailed a \$20.00 incentive, 3,494 businesses were sent the check in advance of the initial contact to request a telephone interview. Of the 77 businesses that were promised an incentive if they completed the interview, only 16 actually completed the interview and were sent a check for \$20.00. Of the 250 cases offered discretionary incentives, three accepted \$50 and completed the interview, one accepted \$55 and completed the interview, and 11 accepted \$100 and completed the interview.

**6.2.13 Level of Effort.** NORC and FRB staff examined the level of effort on all cases, both finalized and pending, as measured by the number of calls for each case. During the month of March, NORC and the FRB agreed to suspend effort on cases that had received more than 30 calls.<sup>1</sup>

---

<sup>1</sup>Effects of the Level of Effort on the Distributions of Completed Responses: An Experiment Using the 1998 Survey of Small Business Finances. Marianne P. Bitler. International Conference on Establishment Surveys, June 17-21, 2000.

The 1998 Survey of Small Business Finances: Sampling and Level of Effort Associated with Gaining Cooperation from Minority-Owned Businesses. Haggerty et.al. International Conference on Establishment Surveys, June 17-21, 2000.

Overall, NORC completed interviews with 3,777 cases. About 25% of the cases were completed with five or fewer calls; just over half of the cases (53%) required 10 calls before the interview was completed. Thus, 85% of the cases were completed with fewer than 25 calls. The distribution of number of calls to complete an interview can be found in Appendix BB.

**6.2.14 Receipt Control.** Receipt Control (RC) protocols and systems were housed at the production center to facilitate tracking and storage of hard copy materials, such as worksheets, tax forms, or other financial records, as they were received from the respondents. In addition, the RC process also kept track of returned mail, which helped with sample management by updating the contact information of the firms in the sample, so that the maximum number of respondents could be contacted.

At the interviewer training, and during group and one-on-one interviewer meetings, interviewers were instructed to encourage respondents to complete and return the worksheet, both during contact calls and at the time of the interview. Forty-seven percent of all cases with any data, and 50 percent of the cases with enough data to be considered complete, returned a worksheet, as compared to only 12 percent of those businesses interviewed for the 1993 survey. The number of returned worksheets by case outcome is presented in Appendix CC.

Table 6.15 shows the number of returned worksheets and other financial records by the week in which they were returned.

**Table 6.15 Number of Worksheets and Other Financial Records by Week Returned**

<b>Week Ending</b>	<b>Complete Worksheet Received</b>	<b>Complete Tax Form Received</b>	<b>Worksheet Return Refused/Blank</b>	<b>Total Returns</b>
10/2/99	0	0	0	0
10/9/99	77	0	41	118
10/16/99	76	0	45	121
10/23/99	99	0	42	141
10/30/99	41	0	16	57
11/6/99	83	0	6	89
11/13/99	52	0	0	52
11/20/99	0	0	0	0
11/27/99	0	0	0	0
12/4/99	191	0	10	201
12/11/99	25	0	3	28
12/18/99	44	0	13	57
12/25/99	0	0	0	0
1/1/00	31	0	12	43
1/8/00	39	0	0	39
1/15/00	11	0	3	14
1/22/00	21	0	12	33
1/29/00	0	0	0	0
2/5/00	48	0	63	111
2/12/00	64	0	0	64
2/19/00	30	21	59	110
2/26/00	47	1	4	52
3/4/00	7	0	0	7
3/11/00	37	2	0	39
3/18/00	55	0	65	120
3/25/00	150	3	115	268
4/1/00	134	2	60	196
4/8/00	237	1	71	309
4/15/00	112	1	0	113
4/22/00	68	0	76	144
4/29/00	0	0	12	12
5/6/00	85	0	11	96
5/13/00	29	1	4	34
5/20/00	0	0	0	0
5/27/00	4	0	0	4
6/3/00	10	0	5	15
6/10/00	4	0	0	4
6/17/00	0	0	0	0
6/24/00	0	0	0	0
7/1/00	0	0	0	0
7/8/00	0	0	0	0
<b>Total</b>	<b>1911</b>	<b>32</b>	<b>748</b>	<b>2691</b>

**6.2.15 Worksheets.** NORC performed an analysis of a subsample of the balance sheet and income data included in the receipted worksheets to determine the extent of their completeness. We found an exceptionally high rate of completeness. The results indicate that 60% of the cases sampled provided full disclosure (all questions answered), 78% of the cases sampled provided more than 90% disclosure, and 92.5% of the cases sampled provided more than 75% disclosure. Only 7.5% of the cases sampled provided less than 75% disclosure of the income statement and balance sheet data on the worksheets.

**6.2.16 Completion Rate / Response Rate.** Table 6.16 shows that of the 11,053 cases released, 3,561 were eligible completes, for a completeness rate of 33%. Of the 11,053 cases released, 4,990 were refusals, for a refusal rate of 45%. The production statistics for the main interview data collection effort, at the time data collection ended, are shown in Table 6.16.

**Table 6.16 Case Release and Sample Status for Main Questionnaire Data Collection**

Cases Released	Completes		Final Refusal	Other Final	Total Finalized	% Final
	Eligible	Ineligible				
11053	3684	308	4990	416	9398	84.60%

*Note: After editing and applying the FRB completeness criteria, the total number of completed cases was reduced from 3,684 to 3,561. To distinguish completed cases from partially completed non-respondents, NORC applied the following criteria approved by the FRB: A completed case was one with 75% item response overall, a 75% item response for Sections P, R and S (the income statement and balance sheet), a 75% item response for Section U (credit history), and a 75% item response for the entrance questions in Sections E, F and MRL (sources of financial services and most recent loans). For cases that came very close to meeting these criteria the FRB applied slightly looser criteria of a 75% (plus or minus one percent) item response on the balance sheet and substantial item response for the entrance questions in Sections E, F and G. For these cases, failure to meet the criterion for section U was not fatal.*

**6.2.17 Other basic statistics.** There are two other basic statistics measured, questionnaire length and level of effort.

Regarding questionnaire length, the average time spent administering this questionnaire was 40.22 minutes for all cases completed through the end of data collection (May 6, 2000).

Regarding level of effort, interviewers worked 24,401 hours and completed 3,561 cases making the overall hours per completed case 6.85.

The level of effort, as measured by the number of telephone calls per case, shows

that an average of 13 calls was required for those cases that eventually completed an interview; this broad category required the least level of effort (see the following table). For the cases in which we did not complete an interview, 19 calls, on average, were made. The table also shows that for those cases in which we completed only part of the interview, we made an average of 24 calls.

**Table 6.17 Level of Effort by Case Status**

Case Status	Number of Cases	Percentage of Cases	Number of Calls	Average # of Calls
Complete	3,561	32%	47,476	13
Partial	1,106	10%	26,254	24
Non-Interview	6,386	58%	123,943	19
<b>Overall</b>	<b>11,053</b>	<b>100%</b>	<b>197,673</b>	<b>19</b>

We also examined the level of effort required, on average, for various sample strata. The following table indicates that the lowest average overall number of calls per case was for the smallest firm size class, even though all the size classes are relatively close. In general, a greater average number of calls was required for partial cases compared to complete cases, and the next greatest level of effort was spent on those cases for which no data were collected (non-interview cases). For cases for which we completed an interview, we experienced the least average number of calls per case.

**Table 6.18 Level of Effort by Firm Size and Case Status**

Case Status	Firm Size - Number of Workers											
	0-4		5-9		10-19		20-49		50-99		100-499	
	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls
Complete	1,857	13	574	14	251	12	362	14	267	14	250	16
Partials	568	34	175	30	103	30	88	28	85	25	87	25
Non- Interview	3,177	19	1,058	19	542	19	633	19	498	21	473	21
<b>Overall</b>	<b>5,602</b>	<b>22</b>	<b>1,807</b>	<b>21</b>	<b>896</b>	<b>20</b>	<b>1,083</b>	<b>20</b>	<b>850</b>	<b>20</b>	<b>810</b>	<b>21</b>

When we examined the level of effort by industry type, we found that the average number of calls per case was greatest for cases in the Mining and Construction SIC class, although all categories hover in the 18-22 calls per case range. The highest average number of calls per case appears in the partial case category, with non-interviews and complete cases experiencing fewer average number of calls per case.



**Table 6.19 Level of Effort by Industry Category and Case Status**

Case Status	Industry Category											
	Mining & Construction		Manufacturing		Transportation, Communication & Utilities		Wholesale & Retail Trade		Finance, Insurance, & Real Estate		Service Industries	
	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls	# Cases	Avg. # Calls
Complete	397	15	366	14	126	14	1,016	13	209	13	1,447	13
Partials	119	27	115	20	57	24	366	24	66	22	383	24
Non- Interview	829	23	563	20	270	21	1,928	18	340	18	2,456	19
<b>Overall</b>	<b>1,345</b>	<b>22</b>	<b>1,044</b>	<b>18</b>	<b>453</b>	<b>20</b>	<b>3,310</b>	<b>18</b>	<b>615</b>	<b>18</b>	<b>4,286</b>	<b>19</b>

Finally, we examined the level of effort by the Dun and Bradstreet credit score of the firm. The following table shows that the highest average number of calls per case occurred for the cases in the 0-25% credit score range; businesses in this range are considered high credit risk firms. The lowest average number of calls per case occurred for cases in the 51-75% credit score range. We also see that the higher average number of calls per case is required for partial cases, with non-interview cases and completes having fewer average calls per case.

**Table 6.20 Level of Effort by D&B Credit Score and Case Status**

Case Status	D&B Credit Score							
	0-25%		26-50%		51-75%		76-100%	
	# of Cases	Avg. # of Calls	# of Cases	Avg. # of Calls	# of Cases	Avg. # of Calls	# of Cases	Avg. # of Calls
Complete	886	15	883	13	786	13	2,557	14
Partials	327	26	281	24	220	23	822	25
Non-Interview	1,680	22	1,652	19	1,480	18	4,780	20
<b>Overall</b>	<b>2,893</b>	<b>21</b>	<b>2,816</b>	<b>19</b>	<b>2,486</b>	<b>18</b>	<b>8,159</b>	<b>20</b>

## VII. DATA REVIEW AND DELIVERY

This chapter describes the processes used to review and prepare the questionnaire data for delivery. NORC delivered the questionnaire data in two ways:

1. **Main data:** These files are unedited and reflect only automated, systematic recoding steps; these files reflect the data as it was originally captured.
2. **Edited data:** These files reflect editing and cleaning measures, as well as the automated, systematic recoding steps applied to the main data.

Each data set was delivered in cumulative batches as additional interviews were completed.

### 7.1 MAIN DATA REVIEW

In general, NORC did not manipulate or change the information found in the main data files. It was the intention of both the FRB and NORC to have the main data directly reflect the respondent's information as it was captured during the interview. The minimal processing applied to the main data prior to delivery is described below.

**7.1.1 Data Frequency Review.** Prior to each delivery of the main interview data, NORC reviewed the data frequencies. NORC had three primary objectives when reviewing the main data frequencies:

1. To identify problems in the intricate CATI program and its skip patterns that are independent of any respondent or interviewer error that may have occurred.
2. To identify key areas in the questionnaire that required targeted editing attention, such as incomplete addresses, especially ZIP codes.
3. To pinpoint invalid data that would need to be corrected during editing.
4. To identify problems caused by faulty CATI logic.

As a result of this data frequency review, NORC was able to initiate corrections to both the data and the CATI program prior to the editing phase. The following problems were discovered in the frequency review and subsequently corrected:

1. A coding error occurred for sex of sole proprietor owners and others (single owned corporations): women (the minority group) were to be flagged as 1 and men as 2. The programmer, however, used an alternative coding for sex. This problem was rectified in a post processing recoding step.
2. LLP firms were not pathing correctly through the owner's demographic sections. The program was modified and the LLP demographic data was corrected in editing. If the skip instructions had worked properly, all LLPs filing as a corporation would have gone to the owner demographic questions for corporations; however, LLPs were asked both the sole proprietor and partnership demographic questions instead. Because the error was caught early in the field period, the main CATI program was corrected so that subsequent LLP firms would path correctly.
3. Both the institution names and Subsection G flags were not being exported. Also errors were found in the programming logic used to construct some of the service flags for Subsection G. The export program and the "behind the scenes" flag programming were subsequently modified.
4. Fifteen cases that had been screened by a proxy in which the eligibility questions had not been asked of the owner during the main questionnaire, as was required. This occurred for two reasons: (1) two cases were pre-loaded with an incorrect screener eligibility flag (cases associated with the problem screening interviewer), and (2) 13 cases followed an incorrect skip pattern at SKIP1 due to a CATI error which was corrected during the field period. Retrieval was subsequently completed on all 15 cases.
5. A CATI problem at SKIP1 was identified. If SKIP1 had worked properly, any case screened by a proxy would go from SKIP1 directly to the eligibility questions. However, SKIP1 sent all cases with data at Question A1\_1 to Question A4\_2 without a check on owner/proxy screening status. As a result, 20 cases were asked Question A4\_2 inappropriately before the problem was corrected. Because the error was caught during the field period, it was corrected.

**7.1.2 Break-off Protocol.** SSBF Main Interviewers attempted to complete all interviews by telephone, however, some respondents did not cooperate entirely.

Despite refusal conversion attempts and multiple contacts, some respondents who had broke off the interview before completing the final section of the questionnaire never would have completed the questionnaire. In these instances, interviewers stopped the interview, suspended the case, and filed it as an incomplete in the system.

These break-off cases occurred for a number of different reasons. Some respondents became impatient with the length of the interview. Other respondents who had completed their worksheet prior to the telephone interview found the questions to be redundant and refused to answer the questions again. The FRB and NORC had anticipated these break-off scenarios and prepared a protocol to follow under these circumstances. The break-off protocol was as follows:

1. The interviewer administered a hard copy break-off questionnaire over the telephone. The break-off questionnaire collected the firm's sales, total assets, total liabilities, and all credit history data. Before ending the call, interviewers encouraged the respondents to mail in their completed worksheet or a copy of the firm's income tax records. The interviewers also verified that the respondent still had the business reply envelope (BRE) in which to return such materials or provided the mailing address if the BRE had been thrown away.
2. The data from the break-off questionnaire was reviewed. If the respondent had mailed the completed worksheet or other financial record information, that data was entered as well. For every break-off case, project staff reviewed the partial CATI data captured on-line, the data captured in the break-off questionnaire, and, if available, the data provided on the worksheet or income tax form.
3. If it appeared that the collective data for a break-off case was consistent and logical, the project staff entered data from the break-off questionnaire into the CATI instrument, along with the worksheet or income tax form information. If any item on the worksheet had not been previously captured in the CATI, it was added. Any items not covered by the break-off questionnaire, worksheet, or income tax form that had not been answered by the respondent were coded as "refused." If the staff believed that additional data could be captured by locating addresses or calling the respondent again, such cases were referred back to the telephone

interviewers to retrieve those particular data.

Approximately 300 cases were completed according to this break-off protocol. Of these, 113 were processed with only a break-off questionnaire, while 186 were processed with both a break-off questionnaire and a worksheet (or other financial documentation). A list of all cases completed in this manner, annotated with the question at which the interview broke off and the processing date, can be found in Appendix DD.

Once the supplemental data for a break-off case was entered, the complete case was filed in the system. At this point, the case was subject to the triple standards of the re-coding process, FRB completeness criteria, and the editing process.

**7.1.3 Completeness Criteria.** Any finalized case, whether finished through telephone interview or break-off processing, was subject to the completeness criteria set forth by the FRB in their memo of September 3, 1999. This memo outlines the various considerations and contingencies regarding this final standard. A case had to satisfy four criteria in order to be considered complete:

1. At least a 75 percent item response for the entire questionnaire, where as a general rule “exceptions” are valid responses and “don’t knows” and “refuses” are missing responses. The completeness ratio can be calculated according to the following equation:

$$\text{Completeness ratio} = \frac{\text{number of questions with valid responses [not DK, RF, ".", or " "]} }{\text{number of questions asked [not "." or " " ]}}$$

2. At least an item response of 7 out of 9 or 10 of the key financial service entrance questions, which include E1, E4, F7, F16, F20, F27, F33, F39, and F42. It also includes the entrance sequence of Subsection MRL (MRL1, MRL2.1, MRL3, and MRL3.2), which was combined into one item.
3. At least a 75 percent item response rate for the questions on profit, total assets, total liabilities, and total equity (Subsections P, R, and S).
4. At least a 75 percent item response rate for questions on credit history (questions 1-4 in Subsection U).

Initially, if any of the four criteria were not met, the case failed the FRB completeness criteria standard. Later, however, it was decided that a case could fail the fourth criterion (Subsection U questions) and still pass if criteria 1, 2 and 3 were met.

Cases that failed this final FRB standard, or that failed the other criteria by a small margin, were individually evaluated to see if they were close to passing or were “on the bubble.” NORC presented these “bubble” cases to the FRB for final review. The FRB evaluated the cases and identified those which they would accept as complete and those that they would not. A list of all of the cases that were processed in this manner was kept in a tracking database and submitted to the FRB along with the delivered data. This database is shown in Appendix EE.

There were 291 finalized cases that failed the FRB completeness criteria. Of these, 191 (66%) were break-off cases. After “bubble” case review, 138 of the 291 cases were presented to the FRB for possible acceptance. The FRB accepted 76 of these cases as complete; this number represented about 25% of the finalized cases that failed the original FRB completeness criteria. These are listed in Appendix FF.

## **7.2 DATA EDITING**

The cleaning and organizing of the data is a required phase before data analysis can proceed. NORC undertook a number of measures to ensure the quality and accuracy of the data collected for the 1998 SSBF. Appendix GG gives an overview of the cleaning and editing steps.

**7.2.1 Identifying Cases for Editing.** All editing work was done on a copy of the original raw data, referred to as the main data. To accomplish this, NORC copied the main CATI data for all complete cases into a duplicate CATI system called the “Shadow” CATI. For editing purposes, the definition of complete is quantitative and means the case reached the end of the CATI instrument and was filed in the system. This definition of complete is entirely separate from the FRB’s completeness criteria definition, which determined the quality of the data collected by requiring cases to contain enough substantive data for data analysis.

Cases were made complete and available for editing in two ways. The first and most common way of completing a case was through a CATI interview (N=3,519). However, in a few instances, cases were made complete through the use of worksheets and break-off questionnaires (N=271).

**7.2.2 The On-line Editing Process.** All edits and changes are made in the Shadow CATI system, so that the actual main data are not altered. This also allows the main database to serve as a reference when making changes in the Shadow data. By comparing the main and edited data, any revisions made through the editing process can be identified for each case. This comparison is especially useful for identifying added or deleted institutions, changes in organizational form, and verbatims which have been back-coded into the existing code frame.

Prior to the start of editing, all editing staff participated in one of the telephone interviewer trainings. All editing staff had prior editing and data cleaning experience. Editing occurred on a flow basis as cases were completed in the telephone center. Complete main cases were loaded into the Shadow CATI for editing in five separate batches starting in December 1999 and ending in June 2000.

By using the Shadow CATI, editors reviewed each case on-line and viewed the data provided in the CATI interview in context along with the question text. Furthermore, when an editor altered the original data the CATI system automatically processed and updated the calculated variables and flags for each case. The on-line editing process included correcting spelling errors, back-coding verbatims into the existing code frame, adding or deleting services and institutions not captured in the main CATI case, and reviewing comments and implementing edits based on those comments.

As part of its standard quality control protocol, supervisors reviewed decisions made by the editors. Especially during the first round of editing, some editing decisions were reversed or otherwise corrected. Supervisory review also detected the need for some change in protocol. For example, editors were instructed not to recontact institutions to verify information.

Even though the Shadow CATI was used to implement edits, problems sometimes occurred. When data had to be moved from one section to another, such as when an institution source needed to be added or dropped, editors did not always move the data correctly. In the future, this problem could be solved by providing editors with additional training or practice cases using a second editor for reviewing changes.

Making editing decisions on the basis of CATI comments was one of the more

difficult steps. Because respondents often provide critical information out of context, as it occurs to them, interviewers inserted important comments throughout the case. Comments were often recorded in one section of the questionnaire even when it was relevant to another section. To solve this problem, editors worked with a hard-copy Comment Sheet while they edited. For each batch of cases loaded into the editing system, a hard-copy Comment Sheet that lists all comments for the case was produced. For cases that had no comments, a page showing only the case identifier was produced.

During the editing process, editors reviewed the Comment Sheet identified any other hard-copy records that might pertain to the case. With all associated hard-copy materials in hand, an editor would edit the case directly into the Shadow CATI. Any editing notes about the case were recorded on the Comment Sheet.

All hard-copy materials necessary for editing were kept in one office. These materials are listed below:

1. *Comment sheets* - For every case, a comment sheet listed all comments made in the main CATI and the question number at which the comment was made.
2. *Extra Services Job Aid form* - This form had been completed by the interviewer as needed to indicate that an institution or service should be added or deleted. The interviewer could do this on-line during the interview because the change was mentioned by the respondent too late in the interview. These added or deleted institutions or services are reflected in the edited data only. A copy of this form can be found in Appendix V.
3. *FRB monitoring comments* - These comments are a listing of the observations or requested updates made by the FRB during the monitoring of the telephone interviews.

At the end of each case in the Shadow CATI is a screen that asks the editor to indicate the editing status of the case. There are two editing outcomes: (1) editing complete and no editing changes were made to the case (editing flag=2), or (2) editing complete and editing changes were made to the case (editing flag=3).

Where the editing process was complete a case was exported and delivered cumulatively to the FRB on a flow basis, along with the corresponding main data



files. NORC reviewed these edited delivery files in a process described in the Data Cleaning subsection.

Production statistics for the editing process were monitored. NORC measured the number of cases edited per week, cumulative cases edited, and minutes per edited case. Edited cases were included in weekly production reports. Other statistics were reported in weekly staff meetings.

### **7.3 DATA CLEANING**

The data cleaning stage consists of a systematic examination that removes data entry errors and identifies inconsistent or inappropriate responses either made or missed by the editing staff.

The CATI system automatically performed most of the data cleaning through programmed checks for valid responses and ranges as the interview proceeded. These quality constraints drastically limited the potential keystroke errors by the interviewers or editors. Additionally, the structure of the CATI system ensures that, for the most part, skips are followed correctly and no questions are left “unanswered”. However, while the need for most of the typical post-data collection cleaning was eliminated, misspellings and errors still remained. Therefore, the NORC project staff systematically reviewed the edited data to identify any problems or inconsistency, concurrently with the FRB review process. The project staff used the following materials and reports to identify problems in the data:

- ▶ Frequencies of edited cases
- ▶ Comparisons of the main and edited data
- ▶ Printouts of the main case data
- ▶ Databases of verbatim responses from the main data
- ▶ Databases of the comments from the main data
- ▶ Requests for changes made by the FRB
- ▶ Requests for retrieval made by the FRB

All data cleaning measures were made to the edited cases within the Shadow CATI system.

NORC generated a three different comparisons between the main and the edited cases, each with a different review focus:

1. *Total Case Comparison Report* - This report compared all questionnaire section

variables. Each case was reviewed to identify any instances where editing or program changes had created a new path in the Shadow CATI instrument that had not been taken during the live interview. The objective was to identify the variables that had (a) recently been pathed-to and (b) not been answered in the main. Once identified, NORC's goal was to verify that they were coded as "NOT ASKED" or .N. This coding was added to CATI after the data were collected to differentiate valid from invalid data, but the ex-post patch was only partially successful. In the future, this coding should be built into the CATI programming logic.

Editing changes made early in the questionnaire would cause the instrument to path through questions not asked during the live interview. Initially, editors were filling these unasked questions with the "don't know" response, which implied that the question had in fact not been asked. Because this option was not acceptable to the FRB, NORC enabled the "not asked" reserve code key.

2. *Changes in Services and Institution Proc Compare* - This report compared the variables found in Sections E, F, G, MRL, H, and Institution Flags. Every case on this report was reviewed in conjunction with the verbatims, comments, and main case printouts to either verify a change or change a variable back to reflect the original main data.
3. *Changes in Dollar Amounts Proc Compare* - This report compared the variables found in Sections P, R, S, and U. Every case on this report was reviewed to be certain that data reflected in these sections were what had been reported in the CATI interview and not what was documented on the worksheet.

Through the review process, the FRB and NORC identified 304 cases that required retrieval. For these cases survey staff generated a retrieval sheet showing the case's background and contact information, the questions to be retrieved, and the other questions affected by retrieval. Interviewers attempted to contact the firm again, and when successful, recorded the results of their efforts on the retrieval sheets. NORC project staff inserted this information into the edited data.

The FRB identified approximately 100 cases that required specific changes to the edited data. Many of these changes were identified through the FRB's review of the comment file. The primary changes were the addition or deletion of institutions, the addition or deletion of services, changes in the number of owners, and changes in firm type. The FRB forwarded these recommended changes to NORC's project staff, who implemented them on a flow basis.

## **7.4 DATA CODING**

NORC was responsible for two types of verbatim coding: race recoding and industry coding.

**7.4.1 Race Coding.** The race questions were only asked during the main interview for a subset of the cases. This subset included cases where a proxy completed the screener, the owner who complete a screener is no longer with the firm, or there is different ownership of the firm. The race questions were administered for 1,433 cases in the main interview. All but 41 of these cases provided race answers that fit into the existing code frame. After reviewing the race verbatim responses for these 41 cases, NORC was able to back-code all but one case into the existing code frame. The remaining case was listed as “Dominican Republic”.

**7.4.2 Industry Coding.** There were 1,212 instances where the respondent did not agree with the SIC description provided by Dun and Bradstreet (DNB) and provided a new description of their firm’s industry classification. For each of these cases, NORC reviewed the new SIC description provided by the respondent and compared it to the original DNB SIC description.

If the DNB SIC description was a better fit than the respondent’s new SIC description, NORC back- coded the question at B1 to a value of 1 (YES) in the edited data, changing the verbatim text field to blank. If NORC determined the respondent’s new SIC description warranted assigning a different SIC code, NORC left the edited data as is. In a separate data file, NORC provided the FRB with the new 4-digit SIC code and the SIC description associated with that code.

NORC’s industry coding review resulted in 584 back codes to the D&B industry code and 628 newly assigned SIC codes.

## **7.5 DATA RECODING PROCESS**

After exporting the main data directly out of Survey Craft, the main and the edited data required some recoding steps prior to delivery. NORC developed a series of systematic recoding processes that were applied to the data.

At the general level, an initial recode of all stored values representing "Refused", "Don't Know", "Exception", and "Not Asked" were recoded to "R", "D", "X", and "N" in the raw main data after it had been exported from the CATI. This process was applied to all appropriate variables in each of the survey data deliverables.

More specific variable-level recoding was also performed for ZIP codes, telephone numbers, and sex of owner for sole proprietorships.

- ▶ Out-of-range values for ZIP5, ZIP4, and PHONE were recoded to "D" ("Don't Know"). Despite training to the contrary, interviewers frequently used the values of 00000, 99999, or 11111 in the ZIP5 fields when the respondent did not know the true value. The recoding program systematically changed these invalid data to a "Don't Know" response.
- ▶ The values for sex of the owner for Sole Proprietorships and single owners were inverted. Originally, the coding specification indicated that females be coded 1 and males be coded 2, to flag the minority trait of female ownership. However, the CATI program coded these variable in reverse. These erroneous codes were switched back to the FRB's original specification through the recoding program.
- ▶ Verbatim text strings in the raw data were replaced with a flag to indicate whether a verbatim was set for that variable for that specific respondent. If so, the replacement variable was set to 1. If not, the same general rule specified above for "R", "D", and "X" was used. Note that the actual verbatim text was delivered in a separate file uniquely identified by the case identifier and verbatim variable combination.

In pre-data collection phases, NORC set up range checks at every verbatim question, such that the only responses accepted that were less than or equal to two characters were DK, RF, and EX. Because the "not asked" option had been enabled part way through the editing process at FRB request, this option had not been included in the extensive original programming. Consequently, there are places where the .N response is not accepted.

- ▶ Interviewer names were collected by the survey software. These were encoded uniquely to remove personal information from the delivered data.
- ▶ Two respondent level recodes were applied to both the main and the edited data. The variables affected were:

BREAKQ - Break-Off question: BREAKQ was set to the name of the variable at which a break-off occurred. Three hundred cases were recoded by case identifier. This variable was set in both the main and edited data.

OUTCOME - The outcome disposition variable was set similarly, that is, by case identifier. All stage 2 respondents (11,059) received an assignment for OUTCOME. Again, this variable was set for both the main and edited data.

## **7.6 GEO-CODING**

The FRB analyzed the SSBF data in many ways. One particular downstream analysis was geographic triangulation. Triangulation consists of analyzing the geographic positions of the respondent firms to their financial institutions. To facilitate this analysis, NORC and two subcontracted vendors geo-coded files of both the institution and firm addresses in order to achieve the most accurate physical locations possible. If the physical and mailing address differed these businesses were automatically earmarked for further review by the geo-coding vendors.

The addresses for those 3,300 firms that had the same information for their physical and mailing addresses were run through the Smart Mailer program to check for address plausibility. Of those 3,300 addresses, 729 addresses were flagged as having address inconsistencies. The 729 firms with inconsistent addresses were added to the 489 firms that had differing information for their physical and mailing addresses, for a total of 1,218 firm addresses; all were sent to the address geo-coding vendor for further processing.

During the second week in August 2000, when editing had been completed, NORC extracted 7,878 financial institution addresses from the edited data. This file included the loop number of the financial institution. Occasionally, the respondent had not provided the city, state and ZIP code for a financial institution; sometimes only a telephone number for the institution had been given. The 125 cases where a telephone number had been provided instead of an address were sent to the Telematch vendor for “reverse match” processing, whereby a telephone number is used to generate a city, state, and ZIP code for the financial institution. Note that reverse matching is not possible when only a toll-free number had been provided.

The remaining 7,753 financial institution addresses from the edited data were combined with the 1,218 updated firm addresses extracted from the main data for a total of 9,096 addresses sent to the Acxiom vendor. Acxiom of Phoenix, Arizona, handled the addresses where the city, state, and/or ZIP code were inconsistent. On the basis of their performance on the 1000-case pre-test, NORC determined that their Addressability software generated a new city when state and ZIP code were provided; a new city and state when only ZIP code was provided; and a new state when city and ZIP were provided. It also double-checked addresses that had all three data elements present.

However, the Addressability software could not generate data when only a state was given; a list of ZIP codes when only a city was given; a list of ZIP codes when only city and state were given; or prioritize data fields and generate new data when city, state, and ZIP code all were provided but were inconsistent.

Consequently, NORC and FRB designed a plan to complement the Acxiom processing to obtain the most accurate and complete data. The first step was for NORC to generate lists of ZIP codes via the U.S. Postal Service Database ([http://www.framed.usps.com/ncsc/lookups/lookup\\_ctystzip.html](http://www.framed.usps.com/ncsc/lookups/lookup_ctystzip.html)) when only city or city/state data were given and the vendor could not provide enough data. The second step was for NORC to add a variable to the file of geo-coded addresses upon return from the vendors. The variable has 10 values:

- 0 = no correction was made by the vendor (all items match)
- 1 = new city (city inconsistent, state and ZIP consistent)
- 2 = new state (state inconsistent, city and ZIP consistent)
- 3 = new ZIP (ZIP inconsistent, city and state consistent) No lists of ZIPs were inserted by the vendors. Those types of cases were left for more in-depth processing by NORC.
- 4 = new city and state (city and state incorrect, ZIP is correct)
- 5 = new city and ZIP (city and ZIP are incorrect, state is correct)
- 6 = new state and ZIP (state and ZIP are incorrect, city is correct)
- 7 = all 3 elements are inconsistent
- 8 = all new address elements (for telephone # cases only)
- 9 = new ZIP coded added via NORC research on U.S. Postal Service database

After all of the processing was complete, NORC delivered the geo-coding files to the FRB in a separate package containing the original exported data files that went to each of the two vendors, the output data files returned from the two vendors, and the processed NORC files (new variable added, U.S. POSTAL SERVICE ZIP processing). As requested by the COTR, institution loop position identifiers were included for each address.

## **7.7 DATA DELIVERIES**

Three major data files comprised the main data and edited data deliveries.

1. *Recorded Data*: This file represents the data captured in the CATI or Shadow CATI; the unique identifier in this file is the case identifier. For each record in the data set, NORC collapsed the variables associated with institution use, stripped most of the text fields, added a flag in place of the open-end response,

and added an X in all variables where an exception code was indicated.

Variables from the sections related to the use of financial institutions (sections E, F, G, and H) were collapsed to match the hard-copy questionnaire, as specified by the FRB. NORC modified exported SurveyCraft files to collapse the multiple institution iteration loops. Eight loops were collapsed from 21 to 3 iterations, and one loop was collapsed from 20 to 8 iterations.

2. *Comments File (includes comments, margin notes and exceptions):* Comments were stripped from the main data set for separate delivery. The unique identifier in this file is the SUID/Variable Name pair. Each comment was delivered in an ASCII bar delimited format. For each comment NORC delivered SUID, DLVNAME (variable name), SASLABEL (variable label), VALUE (variable value), and COMMENT1, COMMENT2, and COMMENT3 (in that order). Because there is no limit on comment size during data capture, large comments were parsed into three fields. Variable names were consistent with those in the main data set for the collapsed sections.
3. *Verbatim File (from open-end and other-specify responses):* Open-end and other-specify responses were stripped from the main data set and delivered separately. The unique identifier in this file was the SUID/VARIABLE NAME pair. Each verbatim was accompanied by the SUID, variable name, and variable label. In the main data set, for those open-end and other-specify responses, a two-digit value replaces the text response (“D” for don’t know, “R” for refused, or “1” to indicate a response exists). For each open-end and other-specify response NORC delivered SUID, Variable Name, SASLABEL (variable label) and VERBATIM (text field), in that order.

NORC established a password-protected web site for the exclusive use of the FRB. NORC posted data delivery files on this web site and then notified the FRB that the files were ready to be downloaded. The required deliveries usually consisted of many large files. Therefore, a compressed archive was created for each separate delivery set using the compression utility PKZIP, or WinZip. This allowed much quicker download times and added an additional security measure by password-encrypting the data. A list of the deliveries and their delivery dates follows in Table 7.1 below:

**Table 7.1 List of Files Delivered and Delivery Dates**

File Name	Main Recoded & Edit Data Description	Date	Case Counts
SSBFMAIN.ZIP	Main Data Files (Passing Cases)	Nov 03 1999	848
MAIN2.ZIP	Main Data Files (Passing Cases)	Nov 24 1999	1074
MAIN3.ZIP	Comments & Verbatims associated with 11/24/99 data	Nov 30 1999	associated w/1074
MAIN4.ZIP	Main Data Files (Failed Cases on the Cusp)	Dec 22 1999	16
MAIN5.ZIP	Edited Main Data Files	Feb 8 2000	1679
MAIN7.ZIP	Main Data Files (Failed Cases on the Cusp)	Feb 9 2000	43
MAIN8.ZIP	Main Data Files w/Comments & Verbatim Data (Passing Cases)	Apr 28 2000	3312
COMMENT1.ZIP	Corrected Comments File associated with 4/28/00 data	May 02 2000	associated w/3312
BUBBLE1.ZIP	Main Data Files (Failed Cases on the Cusp)	May 02 2000	81
EDITDATA.ZIP	Edited Data Files w/Comments & Verbatim Data	May 05 2000	2227
COMP0628.ZIP	Main Data Files w/Comments & Verbatim Data (Passing Cases)	Jun 28 2000	3498
CUSP0628.ZIP	Main Data Files (Failed Cases on the Cusp)	Jun 28 2000	138
FAIL0628.ZIP	Main Data Files (Failed Cases )	Jun 28 2000	153
EDIT0817.ZIP	Edited Data Files	Aug 17 2000	3789
SSBF0901.ZIP	Main Data Files w/Comments & Verbatim Data (Passing Cases)	Sep 01 2000	3561
FAIL0901.ZIP	Main Data Files w/Comments & Verbatim Data (Failing Cases)	Sep 01 2000	1017
INEL0901.ZIP	Main Data Files w/Comments & Verbatim Data (Ineligible Cases)	Sep 01 2000	89
EDT0928C.ZIP	Edited Data Files w/Comments & Verbatim Data (Passing Cases)	Sep 28 2000	3561
EDT0928F.ZIP	Edited Data Files w/Comments & Verbatim Data (Failing Cases)	Sep 28 2000	46
EDT0928I.ZIP	Edited Data Files w/Comments & Verbatim Data (Ineligible Cases)	Sep 28 2000	13
EDT1116C.ZIP	Final Data File for the Passing Cases	Nov 16 2000	3561
EDT1116F.ZIP	Final Data Files for Failing Cases	Nov 16 2000	46
EDT1116I.ZIP	Final Data Files for Ineligible Cases	Nov 16 2000	13
GEOCODE.ZIP	Final Files of geographical coding and address updates for institutions and firms	Nov 28 2000	9221 (addresses)
EDTMAIN2.CBK	Codebook	Nov 28 2000	NA
DICT1227.ZIP	Data Dictionary	Dec 27 2000	NA



## VIII. SAMPLING AND WEIGHTING METHODOLOGY

### 8.1 Introduction

The SSBF used a two-stage sampling design. The design steps prior to data collection were (1) determining objectives; (2) establishing a frame; and (3) drawing the screener sample. First, a large sample of firms was selected from the frame. This first stage sample was screened for minority ownership, size, and eligibility. The first stage was necessary primarily to identify minority ownership status for stratification and subsampling in the second stage. Then, during data collection, sample units were screened for eligibility, and (4) eligible cases were subsampled. Following data collection, (5) weights for analysis were calculated. Each of these steps for the 1998 SSBF is discussed below.

### 8.2 Objectives

The objectives were to have 6,000 completed interviews with at least a 70 percent response rate. Sample sizes by size class should have been sufficient to yield a 95 percent confidence interval of plus or minus .05 for an estimate of a proportion. Similarly, the sample sizes for minority groups should have been sufficient to yield 95 percent confidence intervals of plus or minus .05 for estimates of differences in proportions between firms owned by members of minority groups and firms owned by non-minorities.

Assuming equal weights within each subgroup, a sample of approximately 400 cases per analysis group is required to meet these requirements. The subgroups that were oversampled are African-American, Hispanic-, and Asian-owned firms, as well as firms in the 20-49, 50-99, and 100-499 employment size strata. For urban and rural areas, as well as the nine Census divisions, the sample was not allocated proportionately.

Table 8.1 shows the assumptions used to derive the size of the initial sample for screening. NORC attempted to achieve a 70 percent response rate, but planned contingencies for a lower response rate. Therefore, for sample planning purposes, we assumed a conservative 61 percent response rate for all strata except the Asian-owned firms. For non-minority strata, we assumed a conservative 31 percent confirmed ineligibility rate overall, where ineligibility may be determined in screening or at the interview stage. Consistent with the 1993 SSBF experience, we assumed a lower 25 percent ineligibility rate for the minority-owned firms. Finally, we assumed that the purchased Dun & Bradstreet (D&B) file contains minority-owned firms in the percentages shown in Table 8.1, based on weighted results from the 1993 SSBF. Oversampling by firm employment size in addition to minority status did not further

increase the initial sample size. Using these assumptions, we needed an initial sample size for screening of about 38,100 firms, as shown in Table 8.1.

**Table 8.1 Derivation of Initial Sample Size**

Stratum	Complete	Response Rate	Sub-sample	Sub-sampling Rate	Eligibles and Unknown	Inelig Rate	Initial Size	Percent of Population
Hispanic	400	.61	656	.53	1,229	.25	1,638	4.3
Asian	400	.40	1,000	1.00	1,000	.25	1,333	3.5
African-American	400	.61	656	.79	829	.25	1,105	2.9
Other, 1-19 emp	3,600	.61	5,902	.34	17,358	.31	25,156	89.3
Other, 20-49 emp	400	.61	656	.34	1,929	.31	2,795	
Other, 50-99 emp	400	.61	656	.34	1,929	.31	2,795	
Other, 100-499 emp	400	.61	656	.34	1,929	.31	2,795	
<b>Total</b>	<b>6,000</b>		<b>10,180</b>		<b>26,530</b>		<b>38,095</b>	

To plan for contingencies, we actually requested slightly more firms initially than shown in Table 8.1. We also requested a small number of firms with 500 or more employees, so that smaller firms inappropriately classified as 500+ would have a chance of being selected. We divided the sample into replicates for managing the screening workload. We ultimately received 39,240 firms from D&B.

As the project progressed, the objectives and plan were revised to conform to contemporary response experience. Three major changes were implemented. First, while Table 8.1 was designed on the assumption that cases not screened would nevertheless be eligible for selection into stage 2, cost considerations caused us to eliminate the unscreened cases after stage 1. To better understand and adjust for the implications of this change, we subsampled 5% of screener nonrespondents for an intense follow-up. This follow-up survey is described later in this chapter. Second, financial incentives were offered to some of the later stage 2 replicates in the hope that such action

would improve response rates. Third, because response rates were lower than expected in spite of incentives and an expanded data collection period, the original targets were eventually modified as shown in Table 8.2. Most of this chapter, however, is written as if it applies to the original sample targets.

**Table 8.2 Revised Sample Targets**

Stratum	Completes
Hispanic	300
Asian	300
African-American	300
Other, 1-19 employees	1,450
Other, 20-49 employees	300
Other, 50-99 employees	300
Other, 100-499 employees	300
<b>Total</b>	<b>3,250</b>

### 8.3 The Frame

Due to the dynamic nature of the business world, no frame can accurately capture the complete set of firms at any one point in time. All potential frames will have some coverage and timeliness issues. Some frames may also have issues with accessibility or usability. Both the 1988 and 1993 rounds of the SSBF used the Dun's Market Identifiers™ (DMI) file produced by D&B. The DMI file is the best commercially available frame that we know of. For these reasons, and to help maintain consistency with historical precedent, we used the May 1999 DMI file for the 1998 SSBF frame.

The DMI file for the United States is based on D&B's credit rating services and business telephone listings. The file includes the following items from the firm: telephone number, address, name of owner or chief executive, classification as a headquarters/branch/ division/single location or subsidiary/non-subsidiary, SIC code, sales volume, and total number of employees, among other things. D&B also provided minority status of the owner, when such information was provided by the business. The minority status is a relatively new item offered by D&B. Although we did not use D&B's minority status for sampling purposes, we did evaluate it by comparing it with the minority information we obtained through screening. Through related services, D&B also provided credit scores for the businesses selected for interviewing.

In preparing the file for use as a frame, all branch, division, and subsidiary records were removed. Also, firms with headquarters located outside of the United States (50 states and the District of Columbia) were removed. What remained were firms at the enterprise level; those that are ultimate parent and single-location companies.

Businesses with SIC codes that were out of scope were removed. Additional out-of-scope businesses may exist in other SIC codes, but these firms were identified through screening to avoid elimination of in-scope businesses. The SIC codes excluded prior to selection of the screening sample are shown in Table 8.3.

**Table 8.3 Out of Scope SIC Codes**

SIC	Industries
0000-0999	Agriculture, fishing, forestry
4311	US Postal Service
6000-6399	Non/depository institutions, security/commodity brokers, insurance carriers
6700-6799	Holding and other investment offices
8600-8699	Membership organizations
9000-9721	Public administration
821103	Public elementary/secondary schools (This is the only 6-digit SIC in the list.)

Given this reduced file, D&B removed duplicate records as identified by records having the same D&B number identifier. In general, D&B assigned a unique D&B number for each business in the DMI file through routine maintenance. If a business is sold or otherwise changed, the D&B number is retained only if the revised business is legally the same entity as the original business; otherwise a new D&B number is assigned. If a duplication is found, the first record meeting the above criteria is retained in the frame. Removing duplicates in this manner could introduce a slight bias based on the sort order of the file, but we expect any such bias to be trivial. Removing duplicates reduced the frame by .5%.

The remaining firms constituted the frame. The DMI file contained all variables needed for stratification for the first stage, described in the next section.

#### **8.4 Basic Sample Design**

As described above, we used a two-stage sample design for the 1998 SSBF. In the first stage, before the screening sample was selected, the DMI file was refined to eliminate certain SIC codes, branches, divisions, subsidiaries, and foreign-owned businesses that

were headquartered outside of the US. The refined DMI file was divided into strata, and establishments within strata were sorted by SIC code. We then selected a large, stratified, systematic sample of firms for screening.

During screening, we determined eligibility on the basis of employment, for-profit status, non-government ownership, operation during December 1998, non-subsidary headquarters office, and SIC. Although some of the information for determining eligibility is available from the frame, the frame may contain erroneous data. Consequently, these data were verified during the screening. We also determined the race/ethnicity of the owner, which was needed for stratification at the second stage.

The second stage was a stratified random subsample that utilized screening results. A portion of the screeners had to be completed to determine the stage 2 sample allocations and selection probabilities. Ideally, all screeners would have been completed before stage 2 sampling, but time constraints required us to begin second stage sampling and interviewing before all the screeners were complete. This introduced some complications which will be discussed below in Section 8.6.

### **8.5 Initial Stratification and Sample Selection: Stage 1**

For the first stage of sample selection, we stratified the de-duplicated frame by the nine Census divisions, urban/rural status, and employment size. An firm was considered urban if it is located in an MSA, or it is considered rural if it is in a non-MSA county. The size classes were 1-19 employees, 20-49 employees, 50-99 employees, 100-499 employees, 500 or more employees, and unknown size. For the first stage of sampling, there were 91 strata.

The 500+ and unknown size strata were sampled to allow any eligible firms in those strata a chance of being selected, but for efficiency the sample allocations in the 500+ stratum were small. The sampling rate in the 500+ stratum was designed to yield weights comparable to the largest average weight for the other size strata. The sample firms selected from the 500+ size class were not counted among the core total. The firms with unknown employment sizes were sampled according to their proportion in the frame, with no specified minimum.

The 1-19 stratum, by far the largest, was undersampled, while the remaining size classes were oversampled to achieve at least 400 completed interviews. The derivation of minimum sample sizes for size classes 20-49, 50-99, and 100-499 is shown in Table 8.1. The minimum for these size classes was larger than their proportionate share of the sample. In allocating the sample, the minimum sample size was rounded up to 2,800 for

convenience. Given a total target sample size and these minimums, the target sample size for the 1-19 size class was obtained by subtraction.

Within each employment size class (with the exception of the 500+ stratum), firms were stratified by Urban/Rural crossed with Census Division, yielding 91 strata. Sample sizes were allocated proportionately within employment size classes. Given the target sample marginal totals and marginal totals in the refined DMI file, we used a SAS program to rake stratum sample counts for the core strata until the marginal objectives were met. The raking algorithm also took into account the minimum sample sizes needed for the size classes. The raking algorithm, also known as iterative proportional fitting, converged very quickly.

After removing ineligible classes of firms, duplicates from the file as described above, D&B produced frequencies by stratum from the frame. The marginal totals were derived by summing the stratum totals. The marginal totals are shown in Table 8.4. Most marginal totals were based on the core frame, excluding the 500+ employment size class.

**Table 8.4 Marginal Totals From the Sampling Frame**

Marginal Category	Frame Count
Urban (in an MSA) - core	5,998,100
Rural (not in any MSA) - core	1,516,536
New England Division - core	441,913
East North Central Division - core	1,100,018
West North Central Division - core	1,094,422
Mid-Atlantic Division - core	544,249
South Atlantic Division - core	1,317,493
East South Central Division - core	393,764
West South Central Division - core	819,140
Mountain Division - core	483,613
Pacific Division - core	1,320,024
Unknown employment - core	25,235
1-19 employees - core	6,982,599
20-49 employees - core	337,117
50-99 employees - core	102,771
100-499 employees - core	66,914
500+ employees	13,783
Total frame firms	7,528,419

**Table 8.5 Ideal Sample Sizes for Core Sample of 38,900**

Region		Size Class					Division Totals
		unknown	1-19	20-49	50-99	100-499	
New England	urban	6	1,426	131	131	131	$\frac{441,913}{7,514,636} \cdot 38,900$
	rural	2	360	33	33	33	<b>2,288</b>
Mid-Atlantic	urban	15	3,548	327	327	327	$\frac{1,100,018}{7,514,636} \cdot 38,900$
	rural	4	897	83	83	83	<b>5,694</b>
East North Central	urban	15	3,530	325	325	325	$\frac{1,094,422}{7,514,636} \cdot 38,900$
	rural	4	893	82	82	82	<b>5,665</b>
West North Central	urban	8	1,756	162	162	162	$\frac{544,249}{7,514,636} \cdot 38,900$
	rural	2	444	41	41	41	<b>2,817</b>
South Atlantic	urban	18	4,250	392	392	392	$\frac{1,317,493}{7,514,636} \cdot 38,900$
	rural	5	1,075	99	99	99	<b>6,820</b>
East South Central	urban	5	1,270	117	117	117	$\frac{393,764}{7,514,636} \cdot 38,900$
	rural	1	321	30	30	30	<b>2,038</b>
West South Central	urban	11	2,642	244	244	244	$\frac{819,140}{7,514,636} \cdot 38,900$
	rural	3	668	62	62	62	<b>4,240</b>
Mountain	urban	7	1,560	144	144	144	$\frac{483,613}{7,514,636} \cdot 38,900$
	rural	2	394	36	36	36	<b>2,503</b>
Pacific	urban	18	4,258	393	393	393	$\frac{1,320,024}{7,514,636} \cdot 38,900$
	rural	5	1,077	99	99	99	<b>6,833</b>
Size Class Total		$\frac{25,235}{7,514,636} \cdot 38,900$	38,900 - 131 - 3(2800)	$\frac{\max(2800, 337,117)}{7,514,636} \cdot 38,900$	$\frac{\max(2800, 102,771)}{7,514,636} \cdot 38,900$	$\frac{\max(2800, 66,914)}{7,514,636} \cdot 38,900$	
		<b>131</b>	<b>30,369</b>	<b>2,800</b>	<b>2,800</b>	<b>2,800</b>	
urban total			$(5,998,100)/(7,514,636) \cdot (38,900)$				<b>31,050</b>
rural total			$(1,516,536)/(7,514,636) \cdot (38,900)$				<b>7,850</b>
Core Grand Total						<b>38,900</b>	

The strata totals and required minimum sizes were used to determine the sample allocation. Table 8.5 shows the (rounded) ideal sample sizes for stage 1, as determined by the raking algorithm, for a total core sample of 38,900 firms. The figure 38,900 was selected as slightly more than the 38,100 firms necessary in Table 1 in order to build in a small margin for errors in the assumptions. With the 500+ stratum expected to contribute no more than 100 firms to the sample, 38,900 core sample firms should have yielded close to 39,000 total sample firms.

Firms classified as having 500+ employees were included in the sample so that eligible firms mis-classified in this size class would have a chance of being selected. The sample from this group was small, though, because of the inefficiencies of finding eligible firms within this group. Among the size classes with fewer than 500 employees, the 1-19 size class had the lowest selection rate overall. The selection rate for this class was applied to the 500+ size class to obtain a desired sample size of 60.

$$\frac{6,982,599}{7,514,636} 13,783 = 60$$

The strata totals were used along with the ideal strata target sample sizes to determine the sampling rates. The rates, interpreted as selecting one in  $n$  businesses, are integer skip intervals for D&B to perform systematic selection. Table 8.6 gives the frame stratum totals and demonstrates the method for determining sample specifications for D&B.

**Table 8.6 Determination of Sampling Rates by Stratum for Stage 1 Screening Sample**

Urban/ Rural	Division	Total Employees Size Class	DMI Totals	Ideal Sample	Ideal Rate	Truncated Rate	Expected Sample
urban	New England	unknown	1,314	6.13	214.356	214	7
urban	New England	1-19	317,422	1425.51	222.673	222	1,430
urban	New England	20-49	16,296	131.43	123.990	123	133
urban	New England	50-99	4,999	131.43	38.035	38	132
urban	New England	100-499	3,486	131.43	26.524	26	135
urban	M Atlantic	unknown	4,823	15.26	316.055	316	16
urban	M Atlantic	1-19	938,620	3548.41	264.518	264	3,556
urban	M Atlantic	20-49	46,630	327.16	142.530	142	329
urban	M Atlantic	50-99	14,000	327.16	42.793	42	334
urban	M Atlantic	100-499	9,627	327.16	29.426	29	332
urban	E N Central	unknown	3,134	15.19	206.320	206	16
urban	E N Central	1-19	776,131	3530.36	219.845	219	3,544
urban	E N Central	20-49	46,952	325.49	144.250	144	327
urban	E N Central	50-99	15,241	325.49	46.825	46	332
urban	E N Central	100-499	10,270	325.49	31.552	31	332
urban	W N Central	unknown	731	7.55	96.821	96	8
urban	W N Central	1-19	273,734	1755.62	155.919	155	1,767
urban	W N Central	20-49	16,318	161.87	100.809	100	164
urban	W N Central	50-99	5,513	161.87	34.058	34	163
urban	W N Central	100-499	3,690	161.87	22.796	22	168
urban	S Atlantic	unknown	3,503	18.28	191.630	191	19
urban	S Atlantic	1-19	1,008,403	4249.93	237.275	237	4,255



Urban/ Rural	Division	Total Employees Size Class	DMI Totals	Ideal Sample	Ideal Rate	Truncated Rate	Expected Sample
urban	S Atlantic	20-49	48,491	391.84	123.752	123	395
urban	S Atlantic	50-99	14,432	391.84	36.831	36	401
urban	S Atlantic	100-499	8,995	391.84	22.956	22	409
urban	E S Central	unknown	630	5.46	115.385	115	6
urban	E S Central	1-19	224,867	1270.19	177.034	177	1,271
urban	E S Central	20-49	12,194	117.11	104.124	104	118
urban	E S Central	50-99	3,595	117.11	30.698	30	120
urban	E S Central	100-499	2,481	117.11	21.185	21	119
urban	W S Central	unknown	1,430	11.37	125.770	125	12
urban	W S Central	1-19	602,111	2642.36	227.869	227	2,653
urban	W S Central	20-49	27,470	243.62	112.758	112	246
urban	W S Central	50-99	8,363	243.62	34.328	34	246
urban	W S Central	100-499	5,339	243.62	21.915	21	255
urban	Mountain	unknown	1,094	6.71	163.040	163	7
urban	Mountain	1-19	305,557	1560.02	195.867	195	1,567
urban	Mountain	20-49	14,933	143.83	103.824	103	145
urban	Mountain	50-99	4,633	143.83	32.212	32	145
urban	Mountain	100-499	2,711	143.83	18.849	18	151
urban	Pacific	unknown	4,322	18.32	235.917	235	19
urban	Pacific	1-19	1,109,124	4258.10	260.474	260	4,266
urban	Pacific	20-49	49,659	392.59	126.491	126	395
urban	Pacific	50-99	15,421	392.59	39.280	39	396
urban	Pacific	100-499	9,411	392.59	23.972	23	410
rural	New England	unknown	322	1.55	207.742	207	2
rural	New England	1-19	91,996	360.42	255.247	255	361
rural	New England	20-49	4,100	33.23	123.382	123	34
rural	New England	50-99	1,160	33.23	34.908	34	35
rural	New England	100-499	818	33.23	24.616	24	35
rural	M Atlantic	unknown	407	3.86	105.440	105	4
rural	M Atlantic	1-19	80,905	897.17	90.178	90	899
rural	M Atlantic	20-49	3,407	82.72	41.187	41	84
rural	M Atlantic	50-99	898	82.72	10.856	10	90
rural	M Atlantic	100-499	701	82.72	8.474	8	88
rural	E N Central	unknown	669	3.84	174.219	174	4
rural	E N Central	1-19	227,274	892.60	254.620	254	895
rural	E N Central	20-49	9,883	82.30	120.085	120	83
rural	E N Central	50-99	2,788	82.30	33.876	33	85
rural	E N Central	100-499	2,080	82.30	25.273	25	84
rural	W N Central	unknown	394	1.91	206.283	206	2
rural	W N Central	1-19	230,218	443.88	518.649	518	445
rural	W N Central	20-49	9,010	40.93	220.132	220	41
rural	W N Central	50-99	2,876	40.93	70.266	70	42
rural	W N Central	100-499	1,765	40.93	43.122	43	42
rural	S Atlantic	unknown	764	4.62	165.368	165	5
rural	S Atlantic	1-19	218,649	1074.54	203.481	203	1,078
rural	S Atlantic	20-49	9,782	99.07	98.738	98	100
rural	S Atlantic	50-99	2,676	99.07	27.011	27	100
rural	S Atlantic	100-499	1,798	99.07	18.149	18	100
rural	E S Central	unknown	390	1.38	282.609	282	2
rural	E S Central	1-19	140,286	321.15	436.824	436	322
rural	E S Central	20-49	6,224	29.61	210.199	210	30
rural	E S Central	50-99	1,849	29.61	62.445	62	30
rural	E S Central	100-499	1,248	29.61	42.148	42	30
rural	W S Central	unknown	349	2.87	121.603	121	3
rural	W S Central	1-19	165,399	668.08	247.574	247	670
rural	W S Central	20-49	5,843	61.60	94.854	94	63
rural	W S Central	50-99	1,757	61.60	28.523	28	63
rural	W S Central	100-499	1,079	61.60	17.516	17	64

Urban/ Rural	Division	Total Employees Size Class	DMI Totals	Ideal Sample	Ideal Rate	Truncated Rate	Expected Sample
rural	Mountain	unknown	429	1.70	252.353	252	2
rural	Mountain	1-19	146,088	394.43	370.378	370	395
rural	Mountain	20-49	5,883	36.37	161.754	161	37
rural	Mountain	50-99	1,494	36.37	41.078	41	37
rural	Mountain	100-499	791	36.37	21.749	21	38
rural	Pacific	unknown	530	4.63	114.471	114	5
rural	Pacific	1-19	125,815	1076.60	116.863	116	1,085
rural	Pacific	20-49	4,042	99.26	40.721	40	102
rural	Pacific	50-99	1,076	99.26	10.840	10	108
rural	Pacific	100-499	624	99.26	6.287	6	104
both	all	500+	13,783	59.95	229.908	229	61
<b>Total</b>				<b>38,960</b>			<b>39,240</b>

The ideal stratum sample sizes were derived from the assumption of a core sample of 38,900 firms, as in Table 8.5. After truncating the selection intervals to integers and adding the firms with 500+ employees, the actual number of firms purchased from D&B was 39,240.

Selection within strata was by systematic sampling, where each stratum was sorted by SIC code. Every  $n^{\text{th}}$  record was selected, beginning with the first, where  $n$  is the sampling interval. Ideally, the systematic sampling procedure would have had a random start in each stratum, but that capability was not built into D&B's sampling mechanism. Furthermore, their sampling procedure required an integer sampling interval. The  $n$  values are given in the truncated rate column in Table 8.6.

The sample was divided into replicates for screening, so that each replicate was a random subsample of the full screening sample. To control costs, we monitored progress during the data collection period so that we released and worked only enough replicates to achieve our target sample sizes. The released replicates were worked until every case was resolved one way or another, in order to preserve the representativeness of the sample. NORC assigned cases to replicates by assigning a "permanent random number" (PRN) to each firm, sorting the file by PRN, and dividing the sorted file into sequences of 100 PRNs. Thus the 39,240 firms were assigned to 393 replicates, 392 of size 100 and the last of size 40. (Actually, the first replicate had 99 firms and the last had 41 due to a slight glitch in the assignment algorithm.)

Firms that were selected for the pretests were not removed from the frame; 13 of the pretest firms were selected for the stage 1 screening sample. The known eligible cases among the 13 were to be subjected to subsampling at the second stage, even though we did not intend to collect interview data from them. The known ineligibles were treated as

ineligibles for the main study, and the pretest nonrespondents were treated as screener nonrespondents of unknown eligibility. For calculating weights, the eligible cases were treated as screener completes and interview nonrespondents. Note, however, that these 13 cases were not included in response rate calculations for evaluating the interviewers.

## **8.6 Screening and Subsampling: Stage 2**

The screener was designed to identify those firms that were eligible for the study. The screener also provided information on a firm's minority status and employment size. This information enabled us to stratify the first stage sample for second stage selection.

The sample as received from D&B should have contained minority-owned businesses in the proportions that they occur in the DMI file, after stratifying. The first stage sample size was determined to provide enough cases to meet the target figures most difficult to achieve, taking estimated population rates, eligibility rates, and completion rates into account. The groups with more easily achieved targets were subsampled. The stratification and subsampling occurred after screening was complete.

After screening, we removed known ineligible cases, owner refusals, and screener nonrespondents of unknown eligibility from further consideration. The known eligible cases were assigned to strata for second stage sampling. The strata were defined similarly to stage 1 strata, with the addition of stratification by racial/ethnic group. Where screener employment, state, and minority data were available, they were used for stratification, rather than the D&B data. D&B's urban/rural indicator was used for second stage stratification, as were employment size for cases that could not be fully screened, such as proxy screeners with unknown employment.

In the second stage, there were 72 strata of non-minority owned businesses, and six strata for the minority-owned businesses, for a total of 78 strata. The major strata are listed in Table 8.9. The major strata were Hispanic, non-Hispanic Asian (simply called Asian), non-Hispanic non-Asian black (African-American), non-minority or "other" 1-19 employees, other 20-49 employees, other 50-99 employees, and other 100-499 employees.

For sample selection, the four racial/ethnic groups were mutually exclusive. This means that a firm screened as, e.g., both Hispanic and Asian was put into only one of the two specified racial/ethnic groups. Practically speaking, there was very little overlap among racial/ethnic categories. The "other" group contained all observations for which minority status could not be identified in the screener.

A small number of firms were owned by persons claiming membership in more than one minority group. Some of these firms had multiple owners of different racial/ethnic groups. Others were owned by a person of mixed heritage. For simplicity, we established the following hierarchical rules for stage 2 sampling stratification:

1. If ownership is more than 50% Hispanic, stratify the firm as Hispanic-owned.
2. Otherwise, if ownership is more than 50% Asian (including Hawaiian and Pacific Islander), stratify the firm as Asian-owned.
3. Otherwise, if ownership is more than 50% African-American, stratify the firm as African-American.
4. Otherwise, stratify the firm as “other.”

In addition to size class, the “other” group was fully stratified by census division and urbanicity. The minority groups may not have had sufficient numbers to support this level of stratification detail. In the prior round, each minority group was stratified only by urban/rural status. We used the same design in the 1998 SSBF.

Cases for which the employment could not be verified were assigned to a size class based on D&B’s employment figure for the firm. If the employment was also unknown to D&B, the case was assigned to the 1-19 size class, effectively collapsing what remained of the unknown size class with the 1-19 size class.

Table 8.1 shows the minority subsample sizes and rates expected given an initial core sample size of approximately 38,100 and assuming the response rates shown. These rates and sizes varied slightly once the final sample sizes in the replicates used was known. Furthermore, the subsampling rates varied within racial/ethnic groups due to additional stratification. The requirements for proportional urban/rural and census division representation, and the requirements for size class representation, could be achieved without additional increases to the sample size.

As with the first stage sampling, the second stage sample was allocated to strata using a raking program to help ensure that all target sample sizes were met. Four types of information were needed for the raking algorithm: (1) marginal totals, (2) the expected eligibility rate among stage 2 cases, (3) the maximum available cases by strata, and (4) target sample sizes for major strata. Each of these is described in turn. The allocation steps are illustrated for an artificial example in the appendix.

**(1) Marginal Totals.** Our original stage 1 marginal totals were based on percentages in the DMI file. Screening revealed a proportion of the cases to be ineligible. The marginal

totals should have been based on estimates of the eligible universe, and those estimates came from the screener responses. The more screeners that were complete, the better the estimates of the eligible universe. Unfortunately, only the first 200 of the stage 1 replicates were finalized at the time the stage 2 sampling began. Therefore, the marginal totals for stage 2 raking were estimated from weighted screener responses of the first 200 stage 1 replicates.

All cases in the first 200 replicates had to have been worked to the same extent that all screeners were ultimately worked, according to set rules for what constituted a finalized case. See Chapter 6 on Data Collection for the rules. The cases from the first 200 stage 1 replicates were classified into the marginal categories as shown in Table 8.7, based on their screener responses. We adjusted the stage 1 base weights (Section 8.11) to reflect only the first 200 replicates and their eligibility and response, so that the completed replicates were a properly weighted stratified random sample of the eligible population. Using these weights, we estimated the eligible universe for each marginal grouping, forcing each dimension to sum to the same total. The estimated marginal totals (rounded to integers) and proportions for stage 2 allocation are shown in Table 8.7.

**Table 8.7 Estimated Marginal Totals After Screening the First 200 Replicates**

Marginal Category	Estimated Total	Percentages
Hispanic	266,596	5.5
Asian	221,015	4.6
African-American	160,008	3.3
Other 1-19 and unknown	3,808,164	79.0
Other 20-49	242,347	5.0
Other 50-99	68,763	1.4
Other 100-499	55,968	1.2
Urban	3,794,751	78.7
Rural	1,028,110	21.3
New England Division	262,717	5.4
East North Central Division	651,723	13.5
West North Central Division	713,642	14.8
Mid-Atlantic Division	369,976	7.7
South Atlantic Division	867,870	18.0
East South Central Division	294,779	6.1
West South Central Division	496,058	10.3
Mountain Division	318,290	6.6
Pacific Division	847,805	17.6
<b>Total eligible firms</b>	<b>4,822,861</b>	<b>100.0</b>

**(2) Expected Stage 2 Eligibility Rate.** The stage 2 ineligibility rates depended upon the estimated ineligibility rates in stage 1 screening, which in turn were based on confirmed ineligibility rates and the estimated proportion of screener nonrespondents that were ineligible. Recall that in Table 8.1, the ineligibles included confirmed ineligibles from screening as well as those cases whose eligibility could not be confirmed during screening (firms we were unable to contact).

Even though cases of unknown eligibility were not subsampled for stage 2, some allowance was made for ineligibles in stage 2 sample allocation. A conservative estimate for the stage 2 ineligible rate for a stratum was expected to be:

$$\text{stage 2 ineligibility rate} = \max(0.02, (\text{total ineligibility rate} - \text{stage 1 ineligibility rate})).$$

The conservative stage 2 eligibility rates, or 1 minus the ineligibility rates, are shown in column 4 of Table 8.9.

**(3) Maximum Available Cases.** The next step prior to stage 2 sample allocation was determining the maximum number of cases available for each of the 78 stage 2 strata, after screening all the stage 1 replicates. The unweighted proportions in the first 200 stage 1 replicates were applied to the total number of cases in all stage 1 replicates to determine the expected maximums, as shown in Table 8.8.

**Table 8.8 Expected Stage 2 Stratum Sizes After Screening is Complete**

Major Stratum Race/Ethnicity and Size Class	Substratum		Expected Counts from All Replicates
	Census Division	Urban/Rural	
Hispanic		urban	859
		rural	126
Asian (non-Hispanic)		urban	726
		rural	82
African-American (non-Hispanic, non-Asian)		urban	536
		rural	49
Other 1-19 employees	New England	urban	585
		rural	194
	Mid-Atlantic	urban	1,401
		rural	514
	East North Central	urban	1,615
		rural	516

Major Stratum	Substratum		Expected Counts from All Replicates	
Race/Ethnicity and Size Class	Census Division	Urban/Rural		
	West North Central	urban	883	
		rural	249	
	South Atlantic	urban	1,764	
		rural	585	
	East South Central	urban	626	
		rural	222	
	West South Central	urban	1,036	
		rural	277	
	Mountain	urban	685	
		rural	206	
	Pacific	urban	1,585	
		rural	538	
	Other 20-49 employees	New England	urban	78
			rural	27
		Mid-Atlantic	urban	157
			rural	43
East North Central		urban	204	
		rural	43	
West North Central		urban	126	
		rural	20	
South Atlantic		urban	233	
		rural	69	
East South Central		urban	69	
		rural	14	
West South Central		urban	126	
		rural	24	
Mountain		urban	102	
		rural	24	
Pacific		urban	186	
		rural	55	
Other 50-99 employees	New England	urban	61	
		rural	18	
	Mid-Atlantic	urban	98	

Major Stratum	Substratum		Expected Counts from All Replicates
Race/Ethnicity and Size Class	Census Division	Urban/Rural	
		rural	33
		urban	135
	East North Central	rural	33
		urban	63
	West North Central	rural	16
		urban	145
	South Atlantic	rural	27
		urban	37
	East South Central	rural	16
		urban	100
	West South Central	rural	18
		urban	43
	Mountain	rural	14
		urban	153
	Pacific	rural	51
		urban	57
Other 100-499 employees	New England	rural	14
		urban	108
	Mid-Atlantic	rural	39
		urban	171
	East North Central	rural	33
		urban	82
	West North Central	rural	22
		urban	159
	South Atlantic	rural	39
		urban	45
	East South Central	rural	10
		urban	98
	West South Central	rural	22
		urban	75
	Mountain	rural	16
		urban	208
	Pacific	rural	39
		urban	



*(4) Target Sample Sizes.* The final step prior to stage 2 sample allocation was determining the target sample sizes by major strata. Comparing Table 8.8 with Table 8.1, shows that we were unlikely to have enough African-American and Asian respondents, even if we had screened all stage 1 replicates. Stage 2 sample allocations were constrained by the expected maximums available after screening was complete.

As shown in Table 8.9, we first estimated the ideal number of stage 2 cases needed, given the response rates and expected stage 2 eligibility rates shown. If the ideal sample size exceeded the maximum available for a major stratum, the difference was allocated to the “other” strata. In an attempt to keep the allocation among size classes as close to the universe proportions as possible, while maintaining at least 400 completes in each size class, the excess cases were allocated to the other 1-19 stratum. Given the stage 2 ineligibility rate, the maximums available, and the response rate assumption as in Table 8.1, the target counts for stage 2 sampling were derived as shown in Table 8.9.

**Table 8.9 Derivation of Stage 2 Sample Sizes**

Stratum	Ideal Completes	Response Rate	Assumed Stage 2 Elig Rate	Ideal Stage 2 Sample Count	Constrained Stage 2 Sample Count	Expected Completes
Hispanic	400	.61	.98	669	669	400
Asian	400	.40	.98	1,020	808	317
Af-Amer	400	.61	.98	669	585	350
Other, 1-19 emp	3,600	.61	.97	6,084	6,310	3,734
Other, 20-49 emp	400	.61	.955	687	687	400
Other, 50-99 emp	400	.61	.98	669	669	400
Other, 100-499 emp	400	.61	.98	669	669	400
<b>Total</b>	<b>6,000</b>			<b>10,468</b>	<b>10,397</b>	<b>6,000</b>

For stage 2 sample allocation, we rounded the stage 2 total sample size up to 11,000. The raking algorithm took into account the total desired sample size of 11,000, the marginal totals and proportions in Table 8.7, the target sample sizes in Table 8.9, and the expected maximum numbers of cases available by stratum in Table 8.8. The final allocations are shown in Table 8.10.

**Table 8.10 Stage 2 Sample Allocation**

Major Stratum	Substratum		Allocation
Race/Ethnicity and Size Class	Census Division	Urban/Rural	
Hispanic		urban	542.864
		rural	126.136
Asian (non-Hispanic)		urban	725.738
		rural	82.262
African-American (non-Hispanic, non-Asian)		urban	535.855
		rural	49.145
Other 1-19 employees	New England	urban	288.576
		rural	87.966
	Mid-Atlantic	urban	715.871
		rural	218.218
	East North Central	urban	783.885
		rural	238.951
	West North Central	urban	406.392
		rural	123.880
	South Atlantic	urban	953.738
		rural	290.727
	East South Central	urban	323.794
		rural	98.702
	West South Central	urban	544.884
		rural	166.097
	Mountain	urban	349.619
		rural	106.574
	Pacific	urban	931.253
		rural	283.873

Major Stratum	Substratum		Allocation	
	Census Division	Urban/Rural		
Other 20-49 employees	New England	urban	28.678	
		rural	8.742	
	Mid-Atlantic	urban	71.142	
		rural	21.686	
	East North Central	urban	77.901	
		rural	23.746	
	West North Central	urban	40.386	
		rural	12.311	
	South Atlantic	urban	94.781	
		rural	28.892	
	East South Central	urban	32.178	
		rural	9.809	
	West South Central	urban	54.149	
		rural	16.506	
	Mountain	urban	34.744	
		rural	10.591	
	Pacific	urban	92.546	
		rural	28.211	
	Other 50-99 employees	New England	urban	27.965
			rural	8.525
Mid-Atlantic		urban	69.374	
		rural	21.147	
East North Central		urban	75.965	
		rural	23.156	
West North Central		urban	39.383	
		rural	12.005	
South Atlantic		urban	92.425	
		rural	27.249	
East South Central		urban	31.378	
		rural	9.565	

Major Stratum	Substratum		Allocation
Race/Ethnicity and Size Class	Census Division	Urban/Rural	
	West South Central	urban	52.804
		rural	16.096
	Mountain	urban	33.881
		rural	10.328
	Pacific	urban	90.246
		rural	27.510
Other 100-499 employees	New England	urban	27.927
		rural	8.513
	Mid-Atlantic	urban	69.278
		rural	21.118
	East North Central	urban	75.860
		rural	23.124
	West North Central	urban	39.328
		rural	11.988
	South Atlantic	urban	92.297
		rural	28.135
	East South Central	urban	31.335
		rural	9.552
	West South Central	urban	52.731
		rural	16.074
	Mountain	urban	33.834
		rural	10.314
	Pacific	urban	90.121
		rural	27.472

We defined stage 2 replicates by major stratum instead of across all strata. The advantage of this approach was that we could release more or fewer replicates of different types to meet our target numbers if our estimated marginal totals proved to be faulty. For example, if the response rate for Hispanics was lower than expected, we could release more Hispanic replicates to help meet our targets. From a theoretical perspective, samples were drawn independently within strata, so replicates could be defined independently within strata, as well. Within the non-minority group, replicates

were defined by size class. Assigning replicates by racial/ethnic group and size class resulted in seven sets of replicates.

The replicates did not have to be the same size in each of the major strata. Our replicates had approximately 50 cases, except for the non-minority 1-19 stratum, which had replicates of size 100.

Within a replicate, substrata were represented in the same proportion as in the target sample counts. For example, an Asian replicate consisted of both urban and rural Asian-owned businesses, with approximately 90% urban and 10% rural, according to the proportions in Table 8.11. If our ratio of urban to rural Asian businesses was incorrect, we could not correct the raw proportions in the sample.

Using some random ordering, all eligible cases in the first 200 stage 1 replicates (batch 1) were sorted within stage 2 strata. If, for example, the sample allocation specified 45 urban Asian cases and 5 rural Asian cases for a replicate of size 50, then we selected the first 45 urban Asian cases and the first 5 rural Asian cases in the specified random order. (The target sample sizes per replicate were not necessarily integers, so the actual replicates were not exactly the same size.) The selected cases constituted a random sample of size 45 from the urban Asian stratum and size 5 from the rural Asian stratum. The next 45 urban Asian cases and 5 rural Asian cases in random order constituted the next Asian stratum. This process continued until all of the eligible Asian cases in the first 200 stage 1 replicates were assigned to stage 2 replicates. Then the stage 2 Asian replicates were examined, and only those with the required number of both urban and rural cases were released. The incomplete replicates were completed with eligible cases from the remaining stage 1 replicates (batch 2) when they were finalized. The same process was used for the other major strata.

One benefit of this approach was that the first stage 2 replicates could be released for interviewing before screening was completed in all the stage 1 replicates, saving time and costs. When the remaining stage 1 replicates were finalized, and their eligible cases were assigned to stage 2 replicates, the cases had to be sorted randomly by stage 1 replicate *within batches*, where the first 200 replicates constitute the first batch. If need be, the assignment to replicates could have taken place in more than two batches as long as the cases were sorted by batch and stage 1 replicate first.

As an aside, the sort order within stage 1 replicates was intended to be the same permanent random number that was used to assign cases to stage 1 replicates, although any random order would have sufficed. Sometime after assignments to stage 1 replicates

were successfully completed, the permanent random number variable was corrupted. When the software attempted to sort by the permanent random number for batch 1 of stage 2 replicates, the sort did not work, and the default sort order was case identification number (ID). The situation was detected after the first batch was released, but before the second batch was assigned to stage 2 replicates. A new random number was assigned to the “leftover” cases from batch 1 not already released and to all cases in batch 2. If all batch 1 replicates were released and worked, the sort order within replicates would not matter. Thus this situation did not affect the minority strata and the 50-99 and 100-499 non-minority strata, as all replicates in these major strata were released.

Within a major stratum, the only time default sort order was relevant was if the last stage 2 replicate released contained parts of a batch 1 stage 1 replicate for a stratum, and the remainder of that stage 1 replicate was not released. This happened for only four substrata in the non-minority 1-19 stratum, and affected at most only six cases. Even though the number of affected cases is small, the default sort order is a problem only if it is not suitably random in relation to the variables of interest. The case ID used in the default sort order is slightly correlated with age of firm, but does not appear to be correlated with other key variables. Having a small number of cases affected by a sort order that only slightly deviates from random is not likely to pose any problems to the survey results.

Initially we released fewer replicates than we ultimately expected to need. We monitored the sample during data collection and released additional replicates by racial/ethnic group and size class as needed. We released all minority replicates, even though the last ones did not have the expected urban/rural representation. The final sample consisted of only those firms selected within the released stage 2 replicates.

Subsampling actually occurred when the decision was made not to release some stage 2 replicates. The probabilities of selection were determined at that point according to the numbers of cases in the released and worked replicates. Ultimately we released all minority replicates, 60 replicates for non-minority firms of size 1-19, 20 replicates for non-minority 20-49, 17 replicates for non-minority 50-99, and 16 replicates for non-minority 100-499. Table 8.11 shows the number of worked cases in the released replicates.

### **8.7 Sample Design Aspects of Interviewing**

Table 8.11 shows the final status of the cases worked in stage 2 by major stratum, where the major strata were defined using information obtained in the screener.

**Table 8.11 Final Status of Stage 2 By Major Stratum**

Major Stratum Using Screener Information	Number of Cases Released and Worked	Ineligible Cases	Completed and Partial Cases	Final Accepted Cases
Hispanic	1,020	26	280	273
Asian	806	22	245	232
African-Amer.	618	23	250	241
Other 1-19	5,992	149	2,131	2,000
Other 20-49	990	27	341	312
Other 50-99	841	29	272	259
Other 100-499	792	37	258	244

The questionnaire for stage 2 interviews contained questions that repeated some of the screener questions for some respondents. A great deal of effort was made to conduct the screener interview with an owner. However, in some cases this proved infeasible. For those cases where the screener was conducted with a proxy (in lieu of an owner), the screener questions were repeated in the stage 2 main interview. The duplication was important to confirm eligibility for those cases. Because we did not select for stage 2 cases that had not completed the screener, we expected the ineligibility rates for stage 2 to be small.

By repeating the screener questions, some eligible cases turned out to be classified into the wrong sampling strata. Tables 8.12-14 summarize the accuracy of the stratification variables among cases that completed the main interview. If the screener questions were not repeated for a particular case, we assumed in these tables that the screener information was correct. We were unable to update urban/rural status without updated county information since urban/rural status is determined by the MSA status of the counties. Organization type is shown in Table 8.15, even though it was not a stratification variable.

**Table 8.12 Pre-Interview Race/Ethnicity Compared with Updated Information**

Screener Info	Hispanic	Asian	African-American	Other/Unknown	Total
Hispanic	250	3	1	19	273
Asian	0	220	1	11	232
African-American	1	0	236	4	241
Other	10	6	3	2,796	2,815
<b>Total</b>	<b>261</b>	<b>229</b>	<b>241</b>	<b>2,830</b>	<b>3,561</b>

**Table 8.13 Pre-Interview Size Class Compared with Updated Information**

Screener Info	1-19	20-49	50-99	100-499	Total
Unknown	2	1	0	0	3
1-19	2,621	10	5	3	2,639
20-49	17	333	16	2	368
50-99	2	13	253	17	285
100-499	6	5	12	243	266
<b>Total</b>	<b>2,648</b>	<b>362</b>	<b>286</b>	<b>265</b>	<b>3,561</b>

**Table 8.14 Pre-Interview Census Division Compared with Updated Information**

Screener Info	NE	MA	ENC	WNC	SA	ESC	WSC	MTN	PAC	Total
New England	154									154
Mid-Atlantic		439				1				440
E. N. Central			485		2					487
W. N. Central				285						285
South Atlantic		1			638				1	640
E. S. Central					1	201				202
W. S. Central	1						381			382
Mountain							1	238	1	240
Pacific									731	731
<b>Total</b>	<b>155</b>	<b>440</b>	<b>485</b>	<b>285</b>	<b>641</b>	<b>202</b>	<b>382</b>	<b>238</b>	<b>733</b>	<b>3,561</b>



**Table 8.15 Pre-Interview Organization Type Compared with Updated Information**

Screener Info	Sole Prop	Partnership	S-Corp	C-Corp	Total
Unknown	4	3	11	15	33
Sole Prop	1,326	31	56	51	1,464
Partnership	37	159	32	48	276
S-Corp	31	10	721	71	833
C-Corp	30	24	199	702	955
<b>Total</b>	<b>1,428</b>	<b>227</b>	<b>1,019</b>	<b>887</b>	<b>3,561</b>

**Table 8.16 Final Accepted Cases By Updated Classification**

Updated Classification	Final Accepted Cases
Hispanic	261
Asian	229
African-American	241
Other 1-19	2,018
Other 20-49	312
Other 50-99	258
Other 100-499	242

Table 8.16 summarizes the final, accepted cases by analytical classification, where the classification correspond to major sampling strata but are defined by updated information.

We considered conducting interviews after the business' 1998 income taxes had been filed on the assumption that the owners could provide better figures if their taxes were complete. This approach would have complicated the management of replicates. Screener responses in the early stage 1 replicates indicated that 70% of the businesses had already filed their taxes for 1998. The next largest group had no idea whether taxes had already been filed or when they would be filed. Based on this information, and the fact that the screener took much longer than anticipated, we decided that postponing interviewing until taxes were filed for the small percentage not already filed was not worth the difficulties it would have caused in managing the staggered workload in the context of random replicates.

## 8.8 Overview of Weighting Adjustments and Calculations

We calculated a sample weight for every firm in the sample. The weights were then used in weighted analyses of the data. The primary purpose of the weights is to eliminate bias due to unequal probabilities of selection. The second purpose is to correct the frame counts for firms included in the list frame that were not eligible for the study. The third purpose is to account for the effects of nonresponse. Informally, a sampling weight approximates the number of firms in the target population that a sampled firm represents.

The final weights were calculated in seven steps:

1. Calculation of base weights for stage 1. A base weight for a respondent is the reciprocal of the probability of selection.
2. Estimation of the eligible proportion among cases for whom eligibility could not be determined through screening. A five percent subsample of the unscreened cases was worked extensively beyond the normal screening period to determine the eligibility composition of the group.<sup>2</sup>
3. Adjustment for ineligible cases.
4. Adjustment for nonresponse at the screening stage.
5. Calculation of base weights for stage 2 of the two-stage sample design, based on the number of replicates actually used in stage 2.
6. Adjustment of stage 2 weights for eligibility.
7. Adjustment of weights for stage 2 nonresponse.

The weight calculations described in the following sections are illustrated in Figure 8 in the appendix.

Post-stratification was not required for this study, and it was not recommended because reliable control totals were not available. Vendors of business frames do not design their services to provide reliable estimates of total urban businesses owned by Asian-Americans, for example. Neither does the government routinely and publicly provide the categorical totals of interest to this study.

## 8.9 Stage 1 Base Weights

The first stage base weight or design weight for an firm is the reciprocal of the

---

<sup>2</sup>Although one criterion for eligibility was “in business during December, 1998,” approximately 32% of the original sample could not be contacted or did not complete a screening interview during the screening period. It was believed that a large proportion of the “unable to contact” screening firms was likely to be ineligible. Consequently, a sample of these firms was interviewed to obtain an estimate of the ineligibility rate for firms that were not contacted or refused to participate in the screener. See Section 8.11.

probability of selection for screening. The probability of selection is the product of the initial probability of selection within the stratum and the proportion of cases in replicates used. Since all stage 1 replicates were screened, the base weight is simply the initial sampling probability used by D&B.

Let  $P_{1i}$  be the probability of selection of firm  $i$ . The initial probability of selection is the stratum sample size divided by the frame size for the stratum, where these sizes are given in Table 8.6. In many applications of systematic sampling, the probability is simply the reciprocal of the sampling rate. If D&B had been able to use non-integer sampling intervals and a random start, then the probability would have been the reciprocal of the ideal rate in Table 8.6. Given the initial probability of selection, the screening base weight in the sample file is defined by:

$$W_{1i} = \frac{1}{P_{1i}}.$$

Recall that interim weights were calculated for the first 200 replicates prior to stage two sample selection. Since not all stage 1 replicates were screened at the time, the base weights were adjusted for the replicates used. Let  $m$  be the number of cases in the first 200 replicates. Then the interim base weights were adjusted as shown below:

$$\begin{aligned} W_{2i}^* &= W_{1i} \frac{39,240}{m}, \text{ in replicates } 1 - 200 \\ &= 0, \text{ in replicates } 201 - 240. \end{aligned}$$

No such adjustment was necessary for the final base weights because all stage 1 replicates had been screened when the final base weights were computed.

### **8.10 Screening Outcomes, Eligibility, and Nonresponse**

Screening had three main outcomes: eligible, ineligible, and unknown eligibility. The cases of unknown eligibility can also be called screener nonrespondents, or unscreened cases. A case was determined to be eligible if (1) the owner's responses satisfied all eligibility criteria, (2) a proxy's responses satisfied all eligibility criteria, or (3) a proxy did not know or refused to answer an eligibility question, but the screener was otherwise complete. A case was ineligible if (1) either the owner or a proxy gave an answer that failed the eligibility criteria, or (2) the firm was confirmed to be out-of-business.

The final screener status of each case was coded differently for different purposes. For example, the codes available in the CATI software were not as detailed as FRB requested, but were far more detailed than needed for weighting. The screener disposition codes, category and subcodes used by the CATI software, eligibility flags,

and ultimate classification for weighting are shown in Table 8.17.

**Table 8.17 Screener Dispositions and Eligibility Status**

Disp Code	CATI category /subcode	Description	Eligibility Flag	Weighting Classification
00	N.A.	Overlap with pretest; not released	use pretest eligibility	
11	33/90	Final noncontact; unconfirmed phone number	10	unknown eligibility*
12	33/91	Final locating problem	10	unknown eligibility*
21	33/92	Final unavailable during field period	11	unknown eligibility
22	33/95	Final incapacitated respondent	11	unknown eligibility
23	33/97	Final noncontact; phone number confirmed	11	unknown eligibility
24	33/98	Final language barrier	11	unknown eligibility
31	33/89	Final refusal	5	unknown eligibility
32	33/99	Final hostile refusal	5	unknown eligibility
64	19/0	Complete; DK eligible per proxy	4	assume eligible
66	19/0	Complete; RF eligible per proxy	6	assume eligible
67	19/0	Complete; eligible per owner	7	eligible
68	19/0	Complete; eligible per proxy	8	eligible
71	19/0	Complete; ineligible per owner	1	ineligible
72	19/0	Complete; ineligible per proxy	2	ineligible
73	19/0	Complete; DK eligible per owner	3	unknown eligibility
75	19/0	Complete; RF eligible per owner	5	unknown eligibility
79	19/0	Complete; ineligible tax filing for partnership	9	ineligible
81	7/81	Intro ineligible; not in business in December 1998	1	ineligible
82	7/82	Intro ineligible; not the headquarters	1	ineligible
83	7/83	Intro ineligible; subsidiary	1	ineligible
84	7/84	Intro ineligible; not-for-profit	1	ineligible
85	7/85	Intro ineligible; government-owned	1	ineligible
87	7/87	Intro ineligible; 500 or more employees	1	ineligible
88	7/88	Intro ineligible; ineligible business type	1	ineligible

\*For interim weights after the first 200 stage 1 replicates were completed, cases with these disposition codes were treated as ineligible.

About 32% of the total sample (12,422 cases) could not be contacted or did not complete a screener interview. Some portion of the unscreened cases, those for whom eligibility was unknown, were indeed ineligible, although we did not know which ones. As described in Section 8.11, we adjusted the weights of all cases of unknown eligibility so that they summed to an estimate of the base weights for the proportion of cases expected to be eligible. First we had to estimate the proportion of unscreened cases that would have been eligible and ineligible. One approach was to apply the proportions from those cases that were successfully screened. This approach was used in calculating the interim weights for stage 2 sample selection. However, it was likely that the eligibility rates for the unscreened cases was lower than for the screened cases because many of those we could not screen may have gone out of business.

We selected a 5% subsample of the unscreened cases for an intensive follow-up to learn more about the eligibility of this group. The screener nonrespondents were assigned to one of five subsampling strata. A 5% subsample was selected randomly within each stratum using a new random number generated for this purpose. The subsample in each stratum tended to have the same proportions of cases by original stratification variables as the entire set of unscreened cases. The strata and their follow-up sample counts are shown in Table 8.18.

**Table 8.18 Subsampling Strata for the 5% Follow-Up Sample**

Stratum	Screener Disposition	Eligibility Flag*	Sample Size	Number Contacted	Confirmed Eligibles
1. Unavailable	21, 22	11	67	49	25
2. Noncontact, confirmed number	23	11	39	21	12
3. Language Barrier	24	11	3	2	1
4. Refusal/DK	31, 32, 73, 75	3, 5	307	285	154
5. Noncontact, unconfirmed number	11	10	34	11	6
6. Locating problem	12	10	171	61	15
<b>Total</b>			<b>621</b>	<b>429</b>	<b>213</b>

\*For this purpose we used screener eligibility flags that had been corrected after editing. A small number of cases received new flag values, primarily because one interviewer had created artificial data. The affected cases were recontacted and the flags adjusted.

NORC interviewers followed specific protocols for attempting to screen these cases, and a case was classified as final unknown eligibility only when the full protocol had been completed. The protocol is described in Chapter 6, Data Collection.

Of the 621 cases selected for this screener nonresponse follow-up, we confirmed eligibility or ineligibility for 283 cases. The results are summarized in Table 8.19.

**Table 8.19 Eligibility Results from 5% Follow-Up**

Stratum	Ineligible	Don't Know	Refused	Eligible	Noncontact, Unconfirmed Phone	Locating Problem
1. Unavailable/ Incapacitated	7	0	17	25	18	0
2. Noncontact, confirmed number	6	0	3	12	18	0
3. Language barrier	1	0	0	1	1	0
4. Refusal/DK	16	4	111	154	20	2
5. Noncontact, unconfirmed number	5	0	0	6	23	0
6. Locating problem	35	0	11	15	110	0

Some assumptions were required for those cases among the 5% for whom eligibility status was never determined. For those cases in the unconfirmed number and locating problem strata, if they were still classified as noncontact unconfirmed number or locating problem after all our efforts, we assumed that the case was not in business and was ineligible. For all other cases in this sample, we had at least some confirmation that the telephone number was correct and that the firm might have been in business. Therefore, we assumed that the nonrespondents to the 5% follow-up were eligible in the same proportion as those for whom eligibility status was determined in the follow-up. For the 11 cases from the locating problem stratum that became refusals, we assumed that they were eligible in the same proportion as the confirmed cases in these strata, excluding the assumed ineligible.

Let  $E_k$  be the set of eligible cases for adjustment cell  $k$ , and let  $I_k$  be the set of ineligible cases in the cell. The proportion of eligibles among the 5% follow-up sample in cell  $k$  is

$$\phi_k = \frac{\sum_{j \in E_k} W_{1j} \gamma_j}{\sum_{j \in E_k \cup I_k} W_{1j} \gamma_j},$$

where  $\gamma_j$  is the subsampling probability, approximately .05 but varying slightly by stratum. For purposes of this calculation, we chose to ignore the nonresponse within the 5% follow-up group.

For most of the screener nonrespondents for whom eligibility was unknown, a group composed of nonrespondents to the 5 % follow-up experiment and those not selected for the 5 % follow-up experiment, the ratio  $\phi_k$  was the weight adjustment factor as defined in the above equation, and as applied in Section 8.11. For the noncontact unconfirmed number cases and locating problem cases,  $\phi_k$  was calculated in two steps. First,  $\phi_k$  was estimated using the 61 cases from the follow-up for whom eligibility or ineligibility had been confirmed. This ratio was applied to the weights of the 11 cases from the locating problem stratum that were refusals in the follow-up. With these 11 weights adjusted, and with the weights of the remaining 133 noncontact cases from these strata set to zero, we recalculated  $\phi_k$  for application to noncontact and locating problem cases not selected for the follow-up. This revised adjustment factor assumed that, if we had been able to work all the screener nonrespondents as hard as we did the 5% follow-up, the same proportions of eligibles, ineligibles, refusals, and remaining noncontacts would have been observed.

The adjustment cells and eligibility adjustment factors are shown in Table 8.20.

### **8.11 Adjusting Stage 1 Weights for Eligibility**

In this step, weights of confirmed ineligible cases were set to zero. The ineligible cases included

- ▶ Branches, divisions, and subsidiaries of a business
- ▶ Businesses that had closed
- ▶ Businesses that had not yet opened by the end of 1998
- ▶ Not-for-profit businesses
- ▶ Businesses with ineligible SIC codes
- ▶ Government-owned businesses
- ▶ Firms with 500 or more employees.

Although the sampling frame eliminated most such firms, some ineligible firms were identified through screening in stage 1.

Eligibility was based on the screener dispositions available at the time of stage 2

sampling, as defined in Table 8.17.<sup>3</sup> After the first 200 replicates were subsampled for stage 2, a few disposition codes were corrected. For stage 1 weighting, however, we retained the disposition codes used in subsampling the first 200 stage 1 replicates for stage 2. For the remaining stage 1 replicates, the corrected disposition codes were applied in subsampling and in the weight calculations. For the 13 cases selected both for the main study and one of the pretests, eligibility for weighting was determined using the pretest dispositions.

We also adjusted the weights of all cases of unknown eligibility so that they summed to an estimate of the base weights for the proportion of cases expected to be eligible. The eligibility proportions were determined from the 5% screener nonresponse follow-up as described in Section 8.10. We did not predict the eligibility status of each unscreened case individually. We adjusted their weights in aggregate.

In adjusting weights, the sample firms were first categorized into rather large adjustment cells. Cell definitions are usually based on variables that are available for all cases and that are likely to be correlated with eligibility or response rates. For stage 1, the eligibility adjustment cells were defined by a combination of the screener disposition code (5% follow-up strata) and number of employees. Table 8.20 shows the adjustment cells.

Adjustment cells cannot be too small, or the estimated eligibility rates will be unstable. Based on disposition code stratification, the only cell likely to be too small was the language barrier group. If cells were further divided by original sampling stratification variables, other cells may have been too small. Adjustment cells were collapsed if a cell had fewer than 20 of the 5% nonresponse follow-up cases for whom eligibility was determined. If collapsing of cells was necessary, cells were merged with a cell that had most of the same values of the stratification variables and similar eligibility rates. Decisions on collapsing of adjustment cells were made after review of the data.

Let  $E_1$  be the set of confirmed and assumed eligible cases for a stage 1 stratum or adjustment cell  $k$ ; let  $I_1$  be the set of confirmed and assumed ineligible cases in the cell, and let  $U_1$  denote the set of cases of unknown eligibility in the cell. These classifications were based on the screener disposition codes as of the time of stage 2 sampling. Let  $\phi_k$  denote the proportion of eligible cases among the 5% nonresponse follow-up

---

<sup>3</sup>As noted, in a few cases eligibility flags had to be corrected during editing. If the corrected flag was available at the time of sampling it was used. If cases that had not yet been selected were selected for stage 2, identified because the eligibility status in the main interview differed from that at the time of sampling, the correction was made during the stage 2 weight adjustments.



corresponding to adjustment cell k as defined in Section 8.10. Then the eligibility-adjusted weights were defined by:

$$\begin{aligned} W_{3i} &= W_{1i} \quad , i \in E_1, \\ &= W_{1i} \phi_k, \quad i \in U_1, \\ &= 0 \quad , i \in I_1. \end{aligned}$$

The adjustment cells and estimated eligibility rates actually used are shown in Table 8.20.

**Table 8.20 Screener Eligibility Adjustment Cells**

Cell	5% Follow-Up Stratum	Cases in Cell, Known or Unknown Eligibility	Cases of Known Eligibility Status from 5% Follow-Up	Initial Eligibility Rate	Final Eligibility Rate $\phi_k$
Unavailable / Incapacitated, 0-19 Employees	1	1093	23		0.91
Refused / DK, 0-19 Employees	4	5025	129		0.92
Refused / DK, 20+ Employees	4	1123	41		0.89
Noncontact, Unconfirmed Number, and Locating Probs.	5, 6	4093	61	0.37	0.12
All Other	2, 3	1088	29		0.68

For the interim weights used to estimate the eligible universe prior to stage 2 sample selection, the ratios  $\phi_k$  were not available, so the proportion of eligible screener nonrespondents was assumed to be the same as the proportion of eligible screener respondents. The interim eligibility-adjusted weights were defined by:

$$\begin{aligned}
W^*_{3i} &= W^*_{2i} && , i \in E_1, \\
&= W^*_{2i} \frac{\sum_{E_1} W^*_{2j}}{\sum_{E_1 \cup I_1} W^*_{2j}}, && i \in U_1, \\
&= 0 && , i \in I_1.
\end{aligned}$$

Recall that  $W^*_{2i}$  was the interim base weight adjusted for the replicates completed at the time of the first batch of stage 2 sampling. See the end of Section 8.9 for more details. Once all replicates were screened,  $W^*_{2i}$  was no longer necessary. Thus the interim weight  $W^*_{3i}$  was not the same as the final eligibility-adjusted weight  $W_{3i}$ .

### 8.12 Adjusting Stage 1 Weights for Nonresponse

When there is significant nonresponse, the weights should be adjusted to account for differences between intended and actual sample sizes and to adjust for any bias in the sample that occurs as a result of varying response rates. If all nonresponse occurs randomly, then the nonresponse adjustment is simply a way of accounting for a reduced sample size. If nonresponse occurs nonrandomly, then the respondents are a biased sample of the target population. If the adjustment cells are correlated with varying response patterns, adjusting the weights for nonresponse reduces the effects of potential nonresponse bias.

Cases of unknown eligibility, including owner refusals, were essentially screener nonrespondents, and they were not subjected to sampling for stage 2. Therefore, their weights were set to zero and the weights of the confirmed and assumed eligible cases (eligibility flag = 4, 6, 7, or 8) were adjusted for nonresponse. Recall that the weights of the screener nonrespondents were already reduced to represent only the eligible cases among the nonrespondents.

The adjustment cells could be defined by the original stratification variables (as received from D&B) because they were available for all cases. Defining cells by sampling strata or other criteria related to response patterns relaxes the random nonresponse assumption somewhat, allowing response rates to vary by the characteristics of the adjustment cells.

Experience in the prior round indicated that response rates may vary by state or by smaller size classes than those used for sampling in the 1998 SSBF. The size classes for the nonresponse adjustment of the stage 1 weights are 0-2 employees, 3-4, 5-9, 10-19, 20-49, 50-99, 100-499, and 500 or more employees. Thus we potentially had as many

as 808 nonresponse adjustment cells [(50 states x 2 urban/rural classifications x 8 size classes) + 8 size classes for the District of Columbia]. We obtained information on state and employee when we purchased the sample from D&B.

As mentioned above, adjustment cells can not be too small because response rates among very small groups can vary dramatically, increasing the variability in the weights.

Adjustment cells were collapsed if either of the following conditions were met:

- a. A cell has fewer than 20 cases
- b. A cell has an adjustment factor greater than 2.0, where the adjustment factor for a cell was defined by:

$$\text{adj factor} = \frac{\sum_{E_1 \cup U_1} W_{3j}}{\sum_{E_1} W_{3j}}, \quad i \in E_1,$$

$$= 0 \quad , \quad i \in (I_1 \cup U_1).$$

If collapsing of cells was necessary, cells were merged with other cells that had most of the same values of the stratification variables and similar response rates. Decisions on collapsing of adjustment cells were made after review of the data. Collapsing was done first by state within census division, where the states were sorted in order of adjustment factor. If further collapsing was necessary, a division was collapsed with a neighboring state or division. Collapsing was not needed beyond this level except for firms with 500+ employees. Because so few of these firms were actually eligible, this cell was extremely small. Urban and rural were combined for this size class, which still did not produce a large enough cell. Finally, the 500+ cases were collapsed with the 100-499 cell that had the largest adjustment factor because the 500+ cell also had a large adjustment factor. Adjustment factors were recalculated for collapsed cells.

Within a collapsed cell, the nonresponse-adjusted weights were defined by:

$$W_{4i} = W_{3i} * \text{adj factor}.$$

The response rate implied by the calculation of stage 1 weights, the reciprocal of the adjustment factor, was the weighted counterpart to response rate definition 3 or 4 in AAPOR (1998), where some portion of the unknown eligibility cases were considered

eligible.<sup>4</sup> The adjustment cells and response rates actually used are shown in Table 8.21.

**Table 8.21 Cells for Stage 1 Nonresponse Adjustments Based on Frame Variables**

Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
1	0-2	Rural	5	DE, MD	29	1.059
2	0-2	Rural	6	AL, MS	51	1.103
3	0-2	Rural	4	MO, ND, SD	64	1.121
4	0-2	Rural	1	ME, RI, VT	60	1.132
5	0-2	Rural	5	WV	39	1.153
6	0-2	Rural	3	IN	60	1.170
7	0-2	Rural	8	ID, NM, NV	51	1.208
8	0-2	Urban	5	DC, SC, WV	119	1.232

<sup>4</sup>The response rate definitions in AAPOR (1998):

- I = Complete interview
- P = Partial interview
- R = Refusal and break-off
- NC = Non-contact
- O = Other
- UH = Unknown if household/occupied HU
- UO = Unknown, other

e = Estimated proportion of cases of unknown eligibility that were eligible

$$RR1 = I / [(I + P) + R + NC + O] + (UH + UO)]$$

Response Rate 1 (RR1), or the minimum response rate, is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-off plus non-contacts plus others) plus all cases of unknown eligibility (unknown if housing unit, plus unknown, other).

$$RR2 = (I + P) / [(I + P) + R + NC + O] + (UH + UO)]$$

Response Rate 2 (RR2) counts partial interviews as respondents.

$$RR3 = I / [(I + P) + R + NC + O] + e(UH + UO)]$$

Response Rate 3 (RR3) estimates what proportion of cases of unknown eligibility were actually eligible. In estimating e, one must be guided by the best available scientific information on what share eligibles make up among the unknown cases and one must not select a proportion in order to boost the response rate. The basis for the estimate must be explicitly stated and detailed. It may consist of separate estimates (Estimate 1, Estimate 2) for the sub-components of unknowns and/or a range of estimators based of differing procedures. In each case, the basis of all estimates must be indicated.

$$RR4 = (I + P) / [(I + P) + R + NC + O] + e(UH + UO)]$$

Response Rate 4 (RR4) allocates cases of unknown eligibility as in RR3, but also includes partial interviews as respondents as in RR2.

$$RR5 = I / [(I + P) + R + NC + O]$$

$$RR6 = (I + P) / [(I + P) + R + NC + O]$$

Response Rate 5 (RR5) is either a special case of RR3 in that it assumes that e=0 (i.e. that there are no eligible cases among the cases of unknown eligibility) or the rare case in which there are no cases of unknown eligibility. Response Rate 6 (RR6) makes that same assumption and also includes partial interviews as respondents. RR5 and RR6 are only appropriate when it is valid to assume that none of the unknown cases were eligibles or when there were no unknown cases. RR6 represents the maximum response rate.

Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
9	0-2	Rural	9	AK	40	1.243
10	0-2	Rural	7	AR	42	1.262
11	0-2	Rural	9	WA	149	1.264
12	0-2	Rural	8	MT	22	1.265
13	0-2	Rural	3	IL	68	1.267
14	0-2	Urban	9	AK, HI	36	1.268
15	0-2	Urban	4	MN, SD	223	1.272
16	0-2	Urban	6	AL	108	1.273
17	0-2	Rural	6	TN	26	1.276
18	0-2	Rural	4	IA, NE	61	1.278
19	0-2	Rural	9	HI	33	1.281
20	0-2	Rural	2	NY	172	1.284
21	0-2	Rural	5	FL	50	1.296
22	0-2	Rural	7	TX	139	1.303
23	0-2	Urban	8	NV	42	1.306
24	0-2	Rural	2	NJ, PA	203	1.314
25	0-2	Rural	3	WI	81	1.321
26	0-2	Urban	6	KY	71	1.324
27	0-2	Rural	9	OR	134	1.330
28	0-2	Rural	5	GA	92	1.337
29	0-2	Rural	9	CA	113	1.347
30	0-2	Urban	4	IA	63	1.353
31	0-2	Urban	1	NH, VT	43	1.359
32	0-2	Urban	7	LA	103	1.362
33	0-2	Rural	5	SC	39	1.365
34	0-2	Urban	3	WI	114	1.367
35	0-2	Rural	4	MN	39	1.371
36	0-2	Urban	8	AZ	156	1.373
37	0-2	Urban	5	NC	189	1.375
38	0-2	Rural	3	OH	59	1.375
39	0-2	Urban	9	WA	291	1.393
40	0-2	Urban	7	OK	84	1.393
41	0-2	Urban	9	OR	112	1.394
42	0-2	Rural	7	LA	28	1.394
43	0-2	Urban	3	MI	282	1.405
44	0-2	Rural	6	KY	34	1.406
45	0-2	Rural	7	OK	50	1.408
46	0-2	Urban	5	FL	608	1.413
47	0-2	Urban	8	MT, NM	81	1.413
48	0-2	Rural	8	AZ, UT, WY	47	1.417
49	0-2	Urban	4	MO	206	1.422
50	0-2	Urban	7	TX	644	1.424
51	0-2	Rural	1	NH	33	1.431
52	0-2	Urban	3	IN	103	1.433
53	0-2	Urban	3	IL	318	1.437
54	0-2	Urban	6	MS	40	1.450
55	0-2	Rural	5	NC	94	1.453
56	0-2	Urban	5	MD	163	1.467
57	0-2	Rural	4	KS	21	1.471
58	0-2	Urban	6	TN	175	1.478
59	0-2	Urban	9	CA	1078	1.482
60	0-2	Urban	8	CO, WY	182	1.494
61	0-2	Urban	3	OH	314	1.502
62	0-2	Urban	5	GA	194	1.506
63	0-2	Urban	2	PA	333	1.518
64	0-2	Urban	1	CT	140	1.519
65	0-2	Urban	4	KS	73	1.526
66	0-2	Rural	3	MI	77	1.526
67	0-2	Urban	1	ME	25	1.545
68	0-2	Urban	2	NJ	311	1.545

Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
69	0-2	Rural	5	VA	55	1.547
70	0-2	Urban	4	NE	42	1.559
71	0-2	Urban	8	ID	26	1.570
72	0-2	Urban	4	ND	21	1.587
73	0-2	Urban	2	NY	568	1.607
74	0-2	Urban	8	UT	52	1.608
75	0-2	Urban	5	DE, VA	188	1.625
76	0-2	Urban	7	AR	49	1.633
77	0-2	Rural	8	CO	24	1.638
78	0-2	Urban	1	MA	224	1.690
79	0-2	Urban	1	RI	39	1.742
80	0-2	Rural	1	CT, MA	50	1.824
81	3-4	Rural	6	MS	36	1.140
82	3-4	Rural	4	KS, MO, ND, NE, SD	65	1.159
83	3-4	Rural	5	VA	44	1.175
84	3-4	Urban	5	DC	20	1.195
85	3-4	Rural	8	MT, NM, UT, WY	54	1.218
86	3-4	Urban	4	ND, SD	27	1.218
87	3-4	Rural	5	SC	30	1.220
88	3-4	Rural	3	WI	51	1.231
89	3-4	Rural	5	WV	23	1.277
90	3-4	Rural	6	AL, KY	36	1.277
91	3-4	Rural	1	ME, NH, RI, VT	55	1.280
92	3-4	Rural	4	IA	22	1.306
93	3-4	Urban	9	AK, HI	38	1.307
94	3-4	Rural	5	NC	78	1.336
95	3-4	Rural	9	CA	82	1.342
96	3-4	Rural	4	MN	22	1.355
97	3-4	Rural	3	IL	48	1.358
98	3-4	Rural	7	TX	91	1.372
99	3-4	Urban	5	NC	161	1.380
100	3-4	Urban	6	AL	119	1.384
101	3-4	Urban	8	MT, NM	56	1.393
102	3-4	Rural	5	GA, MD	91	1.397
103	3-4	Urban	1	NH, VT	37	1.398
104	3-4	Rural	3	MI	55	1.406
105	3-4	Urban	7	AR	32	1.421
106	3-4	Urban	9	OR	91	1.422
107	3-4	Urban	4	IA	46	1.423
108	3-4	Rural	9	OR	76	1.424
109	3-4	Rural	7	OK	28	1.426
110	3-4	Rural	8	AZ, ID, NV	37	1.439
111	3-4	Urban	5	SC	81	1.449
112	3-4	Rural	2	NJ, PA	142	1.451
113	3-4	Urban	3	WI	84	1.457
114	3-4	Urban	4	KS	63	1.461
115	3-4	Urban	7	LA	98	1.461
116	3-4	Rural	9	HI	24	1.468
117	3-4	Urban	4	MO	164	1.477
118	3-4	Rural	2	NY	104	1.479
119	3-4	Urban	6	KY	69	1.490
120	3-4	Urban	3	MI	284	1.492
121	3-4	Urban	7	OK	63	1.498
122	3-4	Rural	9	AK	25	1.504
123	3-4	Urban	8	UT	55	1.507
124	3-4	Urban	5	DE, WV	40	1.508
125	3-4	Urban	9	WA	192	1.513
126	3-4	Urban	3	OH	275	1.518
127	3-4	Urban	4	NE	33	1.529
128	3-4	Urban	6	TN	185	1.531

Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
129	3-4	Urban	6	MS	37	1.535
130	3-4	Urban	4	MN	163	1.543
131	3-4	Urban	5	FL	564	1.554
132	3-4	Rural	9	WA	91	1.561
133	3-4	Rural	6	TN	21	1.566
134	3-4	Urban	7	TX	672	1.580
135	3-4	Urban	5	GA	180	1.594
136	3-4	Urban	5	MD	169	1.620
137	3-4	Urban	8	ID, NV	78	1.628
138	3-4	Urban	9	CA	1138	1.630
139	3-4	Rural	3	IN	34	1.635
140	3-4	Urban	5	VA	159	1.640
141	3-4	Urban	3	IN	127	1.653
142	3-4	Urban	3	IL	335	1.667
143	3-4	Urban	2	NJ	309	1.671
144	3-4	Rural	5	DE, FL	49	1.674
145	3-4	Urban	2	PA	309	1.686
146	3-4	Rural	8	CO	25	1.693
147	3-4	Rural	3	OH	45	1.708
148	3-4	Urban	1	CT	156	1.709
149	3-4	Urban	8	AZ	132	1.735
150	3-4	Rural	7	AR, LA	57	1.759
151	3-4	Urban	1	MA, ME, RI	297	1.773
152	3-4	Urban	2	NY	596	1.783
153	3-4	Urban	8	CO, WY	178	1.818
154	3-4	Rural	1	CT, MA	43	1.946
155	5-9	Rural	9	CA, HI	54	1.084
156	5-9	Rural	3	IL, IN	31	1.089
157	5-9	Rural	5	DE, FL, WV	25	1.099
158	5-9	Rural	8	AZ, CO, ID, MT, NM, NV, UT, WY	52	1.113
159	5-9	Urban	4	ND, NE	28	1.120
160	5-9	Rural	6	AL, KY, MS, TN	54	1.133
161	5-9	Rural	3	MI	27	1.161
162	5-9	Rural	9	WA	32	1.162
163	5-9	Urban	5	SC, WV	39	1.170
164	5-9	Urban	8	ID, MT, NM	24	1.172
165	5-9	Urban	6	TN	71	1.185
166	5-9	Rural	2	NJ, NY	56	1.186
167	5-9	Urban	5	NC	70	1.222
168	5-9	Rural	7	AR, LA, OK, TX	72	1.228
169	5-9	Rural	5	SC	20	1.231
170	5-9	Urban	4	IA	26	1.236
171	5-9	Urban	4	MO	64	1.250
172	5-9	Rural	4	IA, KS, MN, MO, ND, NE, SD	49	1.252
173	5-9	Urban	1	ME, NH, RI	37	1.252
174	5-9	Urban	3	IL	171	1.260
175	5-9	Rural	2	PA	56	1.269
176	5-9	Urban	7	TX	241	1.269
177	5-9	Urban	8	UT	29	1.275
178	5-9	Urban	6	AL, MS	78	1.282
179	5-9	Urban	4	MN, SD	62	1.282
180	5-9	Urban	8	AZ	59	1.282
181	5-9	Rural	5	GA, VA	49	1.291
182	5-9	Urban	3	OH	137	1.299
183	5-9	Urban	5	FL	203	1.300
184	5-9	Rural	3	WI	20	1.308
185	5-9	Urban	7	AR, LA	65	1.308
186	5-9	Urban	9	AK, CA, HI	380	1.312
187	5-9	Urban	5	VA	73	1.331
188	5-9	Urban	9	OR	32	1.351

Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
189	5-9	Urban	3	IN	58	1.355
190	5-9	Urban	1	MA	81	1.361
191	5-9	Rural	5	MD, NC	39	1.362
192	5-9	Urban	8	CO, NV, WY	71	1.363
193	5-9	Urban	5	DC, DE, GA	97	1.372
194	5-9	Urban	2	NY	223	1.378
195	5-9	Rural	1	CT, MA, ME, NH, RI, VT	39	1.380
196	5-9	Urban	2	NJ	126	1.391
197	5-9	Urban	6	KY	33	1.392
198	5-9	Rural	9	AK, OR	36	1.393
199	5-9	Urban	1	CT, VT	52	1.406
200	5-9	Urban	7	OK	43	1.439
201	5-9	Urban	4	KS	30	1.450
202	5-9	Rural	3	OH	33	1.450
203	5-9	Urban	2	PA	113	1.477
204	5-9	Urban	3	WI	55	1.479
205	5-9	Urban	5	MD	52	1.487
206	5-9	Urban	3	MI	111	1.615
207	5-9	Urban	9	WA	60	1.847
208	10-19	Urban	9	AK, HI, OR	34	1.058
209	10-19	Rural	5	DE, FL, GA, MD, SC, WV	34	1.063
210	10-19	Urban	3	IN	35	1.088
211	10-19	Urban	5	DC, DE, SC, WV	26	1.094
212	10-19	Rural	6	AL, KY, MS, TN	20	1.103
213	10-19	Urban	6	KY, MS	29	1.111
214	10-19	Urban	5	NC	34	1.131
215	10-19	Rural	7	AR, OK, TX	24	1.132
216	10-19	Rural	9	AK, CA, HI, OR, WA	46	1.140
217	10-19	Rural	8	AZ, CO, ID, MT, NV, UT, WY	24	1.144
218	10-19	Urban	8	AZ, MT, NM, NV	64	1.151
219	10-19	Rural	3	IL, IN, MI, OH, WI	56	1.154
220	10-19	Rural	4	IA, KS, MO, NE	24	1.158
221	10-19	Urban	4	IA, MO, NE, SD	65	1.170
222	10-19	Urban	3	WI	28	1.182
223	10-19	Rural	1	CT, MA, ME, NH, VT	28	1.193
224	10-19	Rural	5	NC, VA	31	1.205
225	10-19	Urban	5	MD	32	1.217
226	10-19	Urban	7	AR, LA, OK	42	1.263
227	10-19	Urban	1	MA, ME, NH, RI, VT	76	1.267
228	10-19	Urban	8	CO	34	1.268
229	10-19	Urban	6	AL	33	1.279
230	10-19	Urban	9	CA	208	1.280
231	10-19	Urban	5	GA	40	1.284
232	10-19	Urban	3	OH	62	1.299
233	10-19	Urban	2	NJ	61	1.322
234	10-19	Urban	7	TX	111	1.362
235	10-19	Rural	2	NJ, NY, PA	43	1.378
236	10-19	Urban	9	WA	33	1.401
237	10-19	Urban	5	FL	95	1.404
238	10-19	Urban	4	KS, MN, ND	56	1.409
239	10-19	Urban	5	VA	34	1.417
240	10-19	Urban	3	IL	70	1.435
241	10-19	Urban	6	TN	29	1.470
242	10-19	Urban	8	ID, UT	26	1.476
243	10-19	Urban	2	NY	93	1.478
244	10-19	Urban	3	MI	55	1.482
245	10-19	Urban	1	CT	26	1.500
246	10-19	Urban	2	PA	54	1.546
247	20-49	Urban	1	ME, NH, RI, VT	26	1.083
248	20-49	Urban	4	MO, ND, NE, SD	57	1.085



Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
249	20-49	Urban	9	AK, HI, OR, WA	58	1.134
250	20-49	Urban	8	ID, MT, NV, UT, WY	40	1.159
251	20-49	Rural	8	AZ, CO, ID, MT, NM, UT, WY	31	1.163
252	20-49	Urban	5	DC, SC, VA, WV	61	1.170
253	20-49	Rural	9	HI, WA	20	1.174
254	20-49	Urban	4	IA, KS	32	1.179
255	20-49	Urban	7	AR, LA	39	1.195
256	20-49	Urban	3	IN	30	1.196
257	20-49	Urban	3	OH	77	1.203
258	20-49	Rural	2	NY, PA	51	1.204
259	20-49	Urban	3	WI	28	1.207
260	20-49	Rural	7	AR, LA, OK, TX	42	1.232
261	20-49	Rural	3	IL, IN, MI, OH, WI	64	1.234
262	20-49	Urban	1	CT	26	1.247
263	20-49	Rural	1	CT, MA, ME, NH, RI, VT	25	1.249
264	20-49	Urban	3	IL	70	1.262
265	20-49	Urban	7	OK, TX	151	1.265
266	20-49	Rural	5, 6	AL, DE, FL, GA, KY, MD, MS, NC, SC, TN, VA, WV	85	1.270
267	20-49	Urban	3	MI	53	1.271
268	20-49	Urban	5	MD	37	1.286
269	20-49	Urban	8	CO	38	1.286
270	20-49	Rural	4	IA, KS, MN, MO, NE, SD	27	1.293
271	20-49	Urban	6	KY, MS, TN	71	1.299
272	20-49	Urban	9	CA	258	1.312
273	20-49	Urban	8	AZ, NM	51	1.316
274	20-49	Urban	5	FL	115	1.321
275	20-49	Urban	1	MA	57	1.350
276	20-49	Urban	2	PA	71	1.357
277	20-49	Urban	4	MN	36	1.379
278	20-49	Urban	2	NJ	70	1.383
279	20-49	Urban	5	DE, GA	48	1.419
280	20-49	Rural	9	AK, CA, OR	47	1.431
281	20-49	Urban	6	AL	25	1.446
282	20-49	Urban	2	NY	114	1.486
283	20-49	Urban	5	NC	39	1.567
284	50-99	Urban	4	MN, ND, NE, SD	52	1.025
285	50-99	Urban	3	IN	30	1.078
286	50-99	Rural	9	AK, CA, HI, OR, WA	69	1.149
287	50-99	Urban	5	DC, DE, MD, WV	57	1.159
288	50-99	Rural	2	NY	26	1.162
289	50-99	Urban	4	MO	32	1.169
290	50-99	Urban	8	MT, NM, NV, UT, WY	38	1.176
291	50-99	Urban	6	AL, KY, MS	48	1.182
292	50-99	Urban	3	OH	58	1.196
293	50-99	Urban	9	AK, HI, OR, WA	69	1.221
294	50-99	Rural	3	IL, IN, MI, OH, WI	53	1.245
295	50-99	Rural	7	AR, LA, OK, TX	34	1.262
296	50-99	Urban	4	IA, KS	30	1.270
297	50-99	Rural	4	IA, KS, MN, MO, ND, NE, SD	28	1.276
298	50-99	Urban	5	NC	38	1.290
299	50-99	Urban	3	MI	46	1.291
300	50-99	Urban	3	IL	72	1.300
301	50-99	Rural	5, 6	AL, DE, FL, GA, KY, MD, MS, NC, SC, TN, VA, WV	79	1.303
302	50-99	Urban	7	OK, TX	147	1.304
303	50-99	Urban	3	WI	34	1.312
304	50-99	Urban	8	CO, ID	33	1.331
305	50-99	Urban	1	MA, ME, RI, VT	63	1.340

Cell	Employment	Urban / Rural	Census Division	States (with sample cases)	Sample Size	Adjustment Factor
306	50-99	Urban	6	TN	41	1.343
307	50-99	Urban	2	PA	79	1.357
308	50-99	Urban	5	FL	100	1.361
309	50-99	Rural	1	CT, MA, ME, NH, RI, VT	28	1.366
310	50-99	Urban	9	CA	229	1.369
311	50-99	Rural	8	AZ, CO, ID, MT, NM, NV, UT, WY	24	1.372
312	50-99	Urban	5	VA	37	1.374
313	50-99	Rural	2	PA	34	1.376
314	50-99	Urban	2	NY	102	1.384
315	50-99	Urban	8	AZ	31	1.429
316	50-99	Urban	5	SC	21	1.429
317	50-99	Urban	5	GA	35	1.435
318	50-99	Urban	2	NJ	63	1.468
319	50-99	Urban	7	AR, LA	32	1.495
320	50-99	Urban	1	CT, NH	36	1.582
321	100-499	Urban	5	DE, GA	28	1.116
322	100-499	Urban	3	IL	73	1.122
323	100-499	Urban	6	AL, KY, MS	48	1.123
324	100-499	Urban	3	WI	25	1.135
325	100-499	Urban	8	CO, ID, NM, NV, UT, WY	64	1.155
326	100-499	Urban	5	VA	38	1.162
327	100-499	Urban	3	IN	30	1.173
328	100-499	Urban	4	MN, NE, SD	55	1.174
329	100-499	Rural	9	AK, CA, HI, OR, WA	44	1.176
330	100-499	Urban	9	AK, WA	41	1.189
331	100-499	Rural	7, 8	AR, CO, ID, LA, MT, NM, OK, TX, UT, WY	38	1.208
332	100-499	Rural	5, 6	AL, DE, FL, GA, KY, MD, MS, NC, SC, TN, VA, WV	71	1.212
333	100-499	Rural	1, 2	CT, MA, ME, NH, NY, RI, VT	36	1.213
334	100-499	Urban	1	MA, ME, NH, RI, VT	53	1.229
335	100-499	Rural	3, 4	IA, IL, IN, KS, MI, MN, MO, NE, OH, WI	57	1.233
336	100-499	Urban	9	OR	22	1.237
337	100-499	Urban	5	MD, SC	56	1.237
338	100-499	Urban	3	OH	44	1.239
339	100-499	Urban	9	CA, HI	231	1.258
340	100-499	Rural	2	PA	21	1.276
341	100-499	Urban	4	IA, KS, MO, ND	72	1.287
342	100-499	Urban	2	NJ	50	1.297
343	100-499	Urban	3	MI	46	1.304
344	100-499	Urban	7	TX	124	1.308
345	100-499	Urban	5	DC, NC	44	1.343
346	100-499	Urban	7	AR, LA, OK, TX	39	1.349
347	100-499	Urban	2	PA	65	1.378
348	100-499	Urban	5	FL, WV	89	1.433
349	100-499	Urban	6	TN	29	1.446
350	100-499	Urban	8	AZ	39	1.607
351	100-499	Urban	2	NY	83	1.620
352	100-499, 500+	Rural and Urban	all	CA, CT, IA, IL, MD, MN, NJ, NY, OH, PA, TX, VA	36	1.655

A similar nonresponse adjustment was applied to the interim weights prior to selecting the stage 2 sample. Figure 2 in the Appendix demonstrates the calculation of interim weights.

### 8.13 Stage 2 Base Weights

At this point, cases were assigned to stage 2 strata on the basis of updated information from the screener. To the extent that the stage 1 stratification was correct and consistent with the screener data, the stage 2 stratification maintained equal weights within stage 2 substrata. Since all stage 2 cases were assigned to stage 2 replicates, the initial stage 2 base weight was the final adjusted weight for stage 1,  $W_{4i}$ . The base weights had to be adjusted for replicates actually used, however. Let  $N_{rep}$  be the total number of cases selected in stage 2 for firm  $i$ 's major stratum, and let  $n_{rep}$  be the number of cases in the major stratum replicates actually used in stage 2. The ratio  $n_{rep}/N_{rep}$  is the probability the case was selected for the subsample in stage 2. If all replicates in a major stratum were used, then the ratio was 1, and no adjustment was necessary. Otherwise, the adjusted base weights are defined by:

$$W_{5i} = W_{4i} \frac{N_{rep}}{n_{rep}}, \quad \text{in replicates used,}$$

$$= 0, \quad \text{in replicates not used.}$$

The values of  $N_{rep}$  and  $n_{rep}$  and the corresponding ratio adjustments are shown in Table 8.22.

**Table 8.22 Stage 2 Base Weight Adjustments for Replicates Used**

Major Stratum	$n_{rep}$ (cases used)	$N_{rep}$ (cases available)	Adjustment Ratio
Hispanic	1020	1020	1.000
Asian	806	806	1.000
African-American	618	618	1.000
Other, 1-19 emp	5992	13,406	2.237
Other, 20-49 emp	990	1,603	1.619
Other, 50-99 emp	841	1,117	1.328
Other, 100-499 emp	792	1,222	1.543

### 8.14 Adjusting Stage 2 Weights for Eligibility

The disposition codes for the main interview cases were slightly different from those used in the screener. Table 8.23 gives the outcome or disposition codes for the main interview. Some cases selected for stage 2 turned out to be ineligible for the study. This happened, for example, when a proxy gave inaccurate information during screening. Table 8.24 summarizes the number of cases by original screener eligibility classification and final eligibility.

**Table 8.23 Outcome Dispositions for the Main Interview**

Outcome Code	Description	Classification	Number of Cases
67	CATI Interview	Complete	3,426
69	Breakoff	Complete	135
77	CATI Interview Failed	Nonrespondent	80
79	Breakoff Failed	Nonrespondent	136
81	Not in business Dec 98	Ineligible	127
82	Not the main office	Ineligible	59
83	Subsidiary	Ineligible	36
84	Not-for-profit	Ineligible	11
86	Government-owned	Ineligible	16
87	500+ employees	Ineligible	47
89	Ineligible business type	Ineligible	17
91	Pretest case not worked	Nonrespondent	6
92	Deceased respondent	Nonrespondent	13
94	Unavailable	Nonrespondent	218
95	Refused	Nonrespondent	5,060
96	Language barrier	Nonrespondent	19
97	Field period ended, likely refused	Nonrespondent	1,234
98	Field period ended, unconfirmed	Nonrespondent	249
99	Unlocatable	Nonrespondent	170

**Table 8.24 Screener Eligibility Status for Ineligible Cases in Stage 2**

Eligibility Category	Original Screener Eligibility Flag		
	4 - DK eligibility per proxy	7 - Eligible per owner	8 - Eligible per proxy
81 - not in business Dec 98	1	98	28
82 - not the main office	1	28	30
83 - subsidiary		13	23
84 - Not-for-profit		5	6
86 - Government-owned		2	14
87 - 500+ employees	2	24	21
89 - Ineligible business type		12*	5

\* Two of these cases had their screener eligibility flags corrected to 10 or 11, unknown eligibility based on supervisory review of cases. See Section 11.

We adjusted the weights for revised eligibility information obtained during the main interview. At this point we assumed that any survey nonrespondents were eligible. Thus the only adjustment needed for eligibility was to set the weights of ineligible cases to zero. Let  $E_2$  and  $I_2$  denote the eligible and ineligible cases after stage 2 interviewing. Then,

$$\begin{aligned} W_{6i} &= 0, & i \in I_2, \\ &= W_{5i}, & i \in E_2. \end{aligned}$$

### 8.15 Adjusting Stage 2 Weights for Nonresponse

Nonrespondents at stage 2 included eligible firms in which the owner refused to participate, had a language barrier, or was unavailable during the data collection period. To distinguish completed cases from partially completed nonrespondents, NORC applied the following criteria approved by the FRB. A completed case was one with 75% item response overall, a 75% item response for Sections P, R and S (the income statement and balance sheet), a 75% item response for Section U (credit history), and a 75% item response for the entrance questions in Sections E, F and MRL (sources of financial services and most recent loans). For cases that came very close to meeting these criteria (so-called bubble cases), the FRB applied slightly looser criteria of a 75% (plus or minus one percent) item response on the balance sheet and substantial item response for the entrance questions in Sections E, F and G. For these bubble cases, failure to meet the criterion for section U was not fatal. Seventy six bubble cases met these FRB criteria and were included in the final sample.

To adjust for nonresponse, we first assigned all stage 2 eligible cases to nonresponse adjustment cells. The cells were intended to group together cases that were similar in propensity to respond and similar in response patterns in order to minimize nonresponse bias. Many variables could be used to define the adjustment cells, but the variables had to be available for all cases, whether or not they responded. One of the issues in defining cells was whether to update the information with corrections obtained in the main interview. It could be argued that mis-stratified cases tended to be more like cases in the strata they should have been in. However, the data collection protocols were based on sampling strata, not corrected strata, so mis-stratified cases received the same treatment as other cases in their sampling strata. More importantly, the updated screener variables were available for only a subset of the stage 2 cases. Thus the nonresponse adjustment cells were based on variables as of the time of stage 2 sampling. The remaining issue was which variables to use.

In the prior round, response rates varied by minority status, urban/rural status, geography, and size class. That is, the stage 2 sampling strata made reasonable cells for nonresponse

adjustments. In the 1998 SSBF, we also considered using stage 2 stratification variables in defining adjustment cells.

Grigorian, et al. (2000) showed that credit scores were strongly correlated with propensity to respond. The credit score ranges that D&B provided were designed to separate firms into distinct classes of credit worthiness. The credit score ranges are shown in Table 8.25 The D&B credit score range was also considered in defining adjustment cells. Another variable highly correlated with propensity to respond was the organization type, the legal structure of the firm. For defining cells, we collapsed similar organization types together as shown in Table 8.26.

**Table 8.25 Credit Score Ranges**

Credit Score Range	Credit Scores	Credit Risk Level
1	91-100	Low
2	71-90	Moderate
3	31-70	Average
4	11-30	Significant
5	1-10	High
0	0	In Bankruptcy or Not at Same Location, Collapsed with 5

**Table 8.26 Organization Type**

Consolidate Type	Detailed Type	Collapse Precedence (if needed)
1 Sole Proprietor	1 Sole Proprietor 9 LLC, tax filed as sole proprietor	Collapse with Partnership
2 Partnership	2 Partnership 3 LLP, tax filed as partnership 7 LLC, tax filed as partnership	Collapse with Sole Proprietor
5 S-Corporation	5 S-Corporation	Collapse with C-corporations
6 C-Corporation	6 C-Corporation 4 LLP, tax filed as corporation 8 LLC, tax filed as corporation	Collapse with S-corporations
0 Don't Know	-2 Don't Know D Don't Know*	Collapse with Sole Proprietor

\*It was requested that NORC's reserve code of -2 for "Don't Know" be converted to "D" prior to delivery. Because data came from different sources in the processing system, it is likely that not all sources got converted. Hence, the two different values for "Don't Know."

Another variable we considered was industry division. The industry divisions were based on the SIC codes provided by D&B, and categorized the same way that the State of Illinois does for employment information. The industry divisions are shown in Table 8.27.

**Table 8.27 Industry Divisions**

Division	SIC Range	Collapsed Precedence (if needed)
Mining	1000-1499	Collapsed with Construction
Construction	1500-1999	Collapsed with Mining
Durable Manufacturing	2400-2599, 3200-3999	Collapsed with Nondurable Manufacturing
Nondurable Manufacturing	2000-2399, 2600-3199	Collapsed with Durable Manufacturing
Transportation/ Communication/Utilities	4000-4999	do not collapse
Wholesale Trade	5000-5199	Collapsed with Retail Trade
Retail Trade	5200-5999	Collapsed with Wholesale Trade
Finance/Insurance/Real Estate	6000-6999	do not collapse
Nonprofessional Services	7000-7999	Collapsed with Professional Services
Professional Services	8000-8999	Collapsed with Nonprofessional Services

Late in the study, some entire replicates were offered financial incentives for completing the questionnaire. Thus the presence or absence of a monetary incentive prior to the interview attempt was another reasonable variable to use in defining adjustment cells. (We did not make special allowances for the small number of cases offered an incentive after refusal as part of an experiment or through discretionary funds. Cases offered an incentive prior to the interview as part of an experiment were included with other cases offered the pre-incentive.)

We found that response rates varied for different reasons in different race/ethnicity groups. The African-American group was sensitive to the incentives, and their propensity to respond increased with their credit scores. The Asian group response rates varied more by organization type and the incentive. The Hispanic group showed very little response variability, although the variables most likely to be significant were industry and organization type. The non-minority group was divided into cells by size, organization type, urban/rural status, industry division, and to a lesser extent by credit score range and census division.

Adjustment cells were collapsed if they contained fewer than 20 observations, or if the

adjustment factor was too large. While our usual criterion for “too large” is 2.0, the lower response rates for this questionnaire necessitated a larger threshold. Thus cells were collapsed if the adjustment factor was greater than 4.0. The Hispanic cells were collapsed by organization type, then by industry. If Asian cells needed to be collapsed, they were collapsed first by incentive, then by organization type. If African-American cells needed to be collapsed, they were collapsed first by organization type, then incentive, then credit score range. Non-minority cells were collapsed first by census division, then credit score range, then by industry division, urban/rural, and organization type as needed. The non-minority cells did not have to be collapsed by size. The final adjustment cells and the adjustment factors are shown in Table 8.28.

**Table 8.28 Nonresponse Adjustment Cells for Stage 2 Weights <sup>5</sup>**

Hispanic Cells				
Cell	Cell Size	Adjust Factor	Industry	Org Type
10001	522	3.953	Mining, Construction, Manufacturing, TCU, Trade	All
10002	228	3.853	Non-profit Services	Non-Corps, S-Corporation
10003	42	3.534	Non-profit Services	C-Corporation
10004	102	3.232	Professional Services	Sole Proprietorship
10005	21	2.377	Professional Services	S-Corporation
10006	33	3.583	Professional Services	C-Corporation
10007	46	2.389	FIRE	All
Asian Cells				
Cell	Cell Size	Adjust Factor	Org Type	Incentive
20001	463	3.873	Non-Corp	Both
20002	321	3.243	Corporations	Both

<sup>5</sup> Abbreviations used in the table:

constr = construction

mod = moderate

mfg = manufacturing

avg = average

nondur = nondurable

sig = significant

dur = durable

NE = New England

TCU = transportation, communication, utilities

MA = Mid-Atlantic

svcs = services

ENC = East North Central

nonprof = nonprofessional

WNC = West North Central

prof = professional

SA = South Atlantic

FIRE = finance, insurance, real estate

ESC = East South Central

corp = corporations

WSC = West South Central

prop = proprietorship

MTN = Mountain

partner = partnership

PAC = Pacific

DK = don't know



African-American Cells					
Cell	Cell Size	Adjust Factor	Credit Score Range	Incentive	Org Type
30001	58	1.880	Low, Mod	Both	All
30002	37	2.535	Moderate	Both	All
30003	161	2.520	Average	No	Non-Corp
30004	23	2.153	Average	No	S-Corp
30005	49	1.740	Average	Both	All
30006	123	3.105	Significant	No	Non-Corp
30007	71	2.473	Significant	Both	All
30008	44	3.007	High	No	Non-Corp
30009	29	2.572	High	Both	All

Non-Minority Cells								
Cell	Cell Size	Adjust Factor	Size	Org Type	Urban/Rural	Industry	Credit Score Range	Census Division
40001	180	3.887	1-19	DK, Sole Prop	Both	Svcs, FIRE	All	All
40002	28	3.362	1-19	Sole Prop	Rural	Mining, Const	Mod, Avg, Sig, High	All but Wsc, Mtn
40003	58	2.488	1-19	Sole Prop	Rural	Mfg	All	All
40004	35	2.760	1-19	Sole Prop	Rural	Wholesale	All	All
40005	67	2.511	1-19	Sole Prop	Rural	Retail	Low, Mod	All
40006	113	3.860	1-19	Sole Prop	Rural	Retail	Average	All
40007	58	2.747	1-19	Sole Prop	Rural	Retail	Sig, High	All
40008	63	2.384	1-19	Sole Prop	Rural	Nonprof Svcs	Low, Mod	All
40009	111	3.188	1-19	Sole Prop	Rural	Nonprof Svcs	Average	All
40010	52	3.976	1-19	Sole Prop	Rural	Nonprof Svcs	Sig, High	All
40011	28	2.559	1-19	Sole Prop	Rural	Prof Svcs	Low, Mod	All
40012	69	1.992	1-19	Sole Prop	Rural	Prof Svcs	Avg., Sig, High	All
40013	38	2.056	1-19	Sole Prop	Rural	FIRE	Mod, Avg, Sig	All but Esc
40014	86	3.498	1-19	Sole Prop	Urban	Constr	Low, Mod	All
40015	200	3.340	1-19	Sole Prop	Urban	Mining, Const	Mod, Avg, Sig, High	All
40016	59	3.186	1-19	Sole Prop	Urban	Durable Mfg	All	All
40017	46	2.137	1-19	Sole Prop	Urban	Nondur Mfg	All	All
40018	46	3.897	1-19	Sole Prop	Urban	TCU	Mod, Avg, Sig, High	All
40019	34	2.625	1-19	Sole Prop	Urban	Wholesale	Low, Mod	All
40020	43	3.587	1-19	Sole Prop	Urban	Wholesale	Average	All
40021	28	3.985	1-19	Sole Prop	Urban	Wholesale	Sig, High	All but Mtn
40022	96	3.549	1-19	Sole Prop	Urban	Retail	Low, Mod	All
40023	97	3.505	1-19	Sole Prop	Urban	Retail	Average	NE, MA, ENC, WNC, SA, ESC
40024	35	2.598	1-19	Sole Prop	Urban	Retail	Average	Wsc, Mtn
40025	35	2.323	1-19	Sole Prop	Urban	Retail	Average	Pacific
40026	86	3.207	1-19	Sole Prop	Urban	Retail	Significant	All
40027	26	2.970	1-19	Sole Prop	Urban	Retail	High	All
40028	93	2.654	1-19	Sole Prop	Urban	Nonprof Svcs	Low, Mod	All
40029	22	2.211	1-19	Sole Prop	Urban	Nonprof Svcs	Moderate	Pacific
40030	131	3.342	1-19	Sole Prop	Urban	Nonprof Svcs	Average	NE, MA, ENC, WNC
40031	46	3.516	1-19	Sole Prop	Urban	Nonprof Svcs	Average	South Atlantic
40032	28	1.978	1-19	Sole Prop	Urban	Nonprof Svcs	Average	East S Central
40033	38	2.568	1-19	Sole Prop	Urban	Nonprof Svcs	Average	Wsc, Mtn
40034	70	2.272	1-19	Sole Prop	Urban	Nonprof Svcs	Average	Pacific

Non-Minority Cells								
Cell	Cell Size	Adjust Factor	Size	Org Type	Urban/Rural	Industry	Credit Score Range	Census Division
40035	98	3.030	1-19	Sole Prop	Urban	Nonprof Svcs	Significant	All
40036	37	3.067	1-19	Sole Prop	Urban	Nonprof Svcs	High	All but NE
40037	116	2.536	1-19	Sole Prop	Urban	Prof Svcs	Low, Mod	All but Pac
40038	25	2.826	1-19	Sole Prop	Urban	Prof Svcs	Moderate	Pacific
40039	48	2.599	1-19	Sole Prop	Urban	Prof Svcs	Average	NE, MA
40040	137	2.949	1-19	Sole Prop	Urban	Prof Svcs	Average	ENC, WNC, SA, ESC, WSC, MTN
40041	51	3.369	1-19	Sole Prop	Urban	Prof Svcs	Average	Pacific
40042	81	2.932	1-19	Sole Prop	Urban	Prof Svcs	Significant	All but Pac
40043	28	3.447	1-19	Sole Prop	Urban	Prof Svcs	Significant	Pacific
40044	20	2.322	1-19	Sole Prop	Urban	Prof Svcs	High	All but NE
40045	41	2.934	1-19	Sole Prop	Urban	FIRE	Low, Mod	All but ENC
40046	62	2.775	1-19	Sole Prop	Urban	FIRE	Average	All
40047	26	1.869	1-19	Sole Prop	Urban	FIRE	Sig, High	All but SA
40048	26	3.498	1-19	Partner	Rural	Mining, Const, Mfg, TCU	All	All
40049	35	2.824	1-19	Partner	Rural	Trade	All	All
40050	47	3.334	1-19	Partner	Rural	Svcs, FIRE	All	All
40051	78	3.167	1-19	Partner	Urban	Mining, Const, Mfg, TCU	All	All
40052	28	3.227	1-19	Partner	Urban	Wholesale	Low, Mod, Avg, Sig	All
40053	40	2.430	1-19	Partner	Urban	Retail	Low, Mod, Avg	All
40054	28	2.870	1-19	Partner	Urban	Retail	Sig, High	All
40055	102	3.897	1-19	Partner	Urban	Services	All	All
40056	88	3.322	1-19	Partner	Urban	Prof Svcs, FIRE	All	All
40057	61	2.506	1-19	S-corp	Rural	Mining, Const, Mfg, TCU	All	All
40058	58	2.187	1-19	S-corp	Rural	Trade	All	All
40059	56	3.328	1-19	S-corp	Rural	Services	All	All
40060	134	3.396	1-19	S-corp	Urban	Mining, Const	All	All
40061	42	2.615	1-19	S-corp	Urban	Durable Mfg	All	All
40062	32	2.591	1-19	S-corp	Urban	Nondur Mfg	All	All but Wsc
40063	39	2.867	1-19	S-corp	Urban	Tcu	All	All but NE
40064	36	2.322	1-19	S-corp	Urban	Wholesale	Low, Mod	All
40065	56	3.482	1-19	S-corp	Urban	Wholesale	Avg., Sig, High	All
40066	51	2.681	1-19	S-corp	Urban	Retail	Low, Mod	All but NE
40067	53	2.842	1-19	S-corp	Urban	Retail	Average	All
40068	52	2.800	1-19	S-corp	Urban	Retail	Sig, High	All
40069	77	2.903	1-19	S-corp	Urban	Nonprof Svcs	Low, Mod	All
40070	113	2.509	1-19	S-corp	Urban	Nonprof Svcs	Avg, Sig	All
40071	20	2.301	1-19	S-corp	Urban	Nonprof Svcs	High	All but NE
40072	35	1.834	1-19	S-corp	Urban	Prof Svcs	Low, Mod	All
40073	53	2.430	1-19	S-corp	Urban	Prof Svcs	Average	All
40074	32	2.681	1-19	S-corp	Urban	Prof Svcs	Sig, High	MA, ENC, WNC, SA, MTN, PAC
40075	66	2.676	1-19	S-corp	Urban	FIRE	All	All
40076	58	1.861	1-19	C-corp	Rural	Mining, Const, Mfg, TCU	All	All
40077	43	3.472	1-19	C-corp	Rural	Trade	All	All
40078	52	2.170	1-19	C-corp	Rural	Retail	Avg., Sig, High	All

Non-Minority Cells								
Cell	Cell Size	Adjust Factor	Size	Org Type	Urban/Rural	Industry	Credit Score Range	Census Division
40079	38	2.654	1-19	C-corp	Rural	Nonprof Svcs	Low, Mod, Avg, Sig	All but Wnc
40080	35	2.103	1-19	C-corp	Rural	Prof Svcs	All	All
40081	20	1.923	1-19	C-corp	Rural	FIRE	All	All but Esc, Mtn
40082	412	3.995	1-19	C-corp	Urban	Mining, Const, Mfg, TCU, Wholesale	All	All
40083	158	3.472	1-19	C-corp	Urban	Retail	Low, Mod, Avg, Sig	All
40084	23	2.617	1-19	C-corp	Urban	Retail	High	All
40085	70	3.234	1-19	C-corp	Urban	Nonprof Svcs	Low, Mod	All
40086	67	2.701	1-19	C-corp	Urban	Nonprof Svcs	Average	All
40087	55	3.397	1-19	C-corp	Urban	Nonprof Svcs	Sig, High	All
40088	68	2.286	1-19	C-corp	Urban	Prof Svcs	Low, Mod	All
40089	54	2.340	1-19	C-corp	Urban	Prof Svcs	Average	All
40090	52	2.522	1-19	C-corp	Urban	Prof Svcs	Sig, High	All
40091	37	3.107	1-19	C-corp	Urban	FIRE	Low, Mod	All
40092	48	3.413	1-19	C-corp	Urban	FIRE	Avg., Sig, High	All
40093	287	3.779	20-49	All but C-corp	Both	All	All	All
40094	99	3.520	20-49	S-corp	Urban	Const, Mfg, TCU, Wholesale	All	All
40095	49	3.730	20-49	S-corp	Urban	Retail	All	All
40096	58	2.307	20-49	S-corp	Urban	Svcs, FIRE	All	All
40097	43	2.871	20-49	C-corp	Rural	Mining, Mfg, TCU	All	All
40098	70	2.432	20-49	C-corp	Rural	Trade, Svcs, FIRE	All	All
40099	162	3.609	20-49	C-corp	Urban	Mining, Const, Mfg, TCU, Wholesale	All	All
40100	24	1.722	20-49	C-corp	Urban	Retail	Low, Mod	All but Esc
40101	25	3.005	20-49	C-corp	Urban	Retail	Average	All
40102	36	2.741	20-49	C-corp	Urban	Retail	Sig, High	All
40103	51	2.765	20-49	C-corp	Urban	Nonprof Svcs	All	All
40104	59	2.672	20-49	C-corp	Urban	Prof Svcs, FIRE	Low, Mod, Avg, Sig	All
40105	84	3.965	50-99	DK, Sole Prop	Both	All but Mining	All	All
40106	67	3.679	50-99	Partner	Both	All but Mining	All	All
40107	23	1.565	50-99	S-corp	Rural	Const, Mfg, TCU	All	All but NE
40108	33	3.083	50-99	S-corp	Rural	Trade, Svcs, FIRE	Low, Mod, Avg	All but Esc
40109	28	3.660	50-99	S-corp	Urban	Construction	All	All
40110	56	2.449	50-99	S-corp	Urban	Mfg., TCU	All	All
40111	27	2.346	50-99	S-corp	Urban	Wholesale	All	All but Wnc
40112	58	2.616	50-99	S-corp	Urban	Retail	All	All
40113	55	2.677	50-99	S-corp	Urban	Svcs, FIRE	All	All
40114	92	2.170	50-99	C-corp	Rural	All	All	All
40115	42	3.924	50-99	C-corp	Urban	Construction	All	All but NE

Non-Minority Cells								
Cell	Cell Size	Adjust Factor	Size	Org Type	Urban/Rural	Industry	Credit Score Range	Census Division
40116	36	3.449	50-99	C-corp	Urban	Durable Mfg	All	All
40117	211	3.876	50-99	C-corp	Urban	Nondur. Mfg, TCU, Trade, Svcs, FIRE	All	All
40118	60	3.206	100-499	DK, Sole Prop	Both	All but Mining	All	All
40119	81	3.971	100-499	Partner	Both	All but Mining	All	All
40120	34	2.396	100-499	S-corp	Rural	Mining, Const, Mfg, TCU	All	All
40121	36	2.142	100-499	S-corp	Rural	Trade, Svcs, FIRE	Low, Mod, Avg, Sig	All
40122	26	2.237	100-499	S-corp	Urban	Construction	All	All but Esc
40123	150	3.800	100-499	S-corp	Urban	All but Mining & Const	All	All
40124	39	2.919	100-499	C-corp	Rural	Const, Mfg, TCU	All	All
40125	43	1.917	100-499	C-corp	Rural	Trade, Svcs, FIRE	All	All
40126	110	3.546	100-499	C-corp	Urban	Mining, Const, Mfg	All	All
40127	92	3.817	100-499	C-corp	Urban	TCU, Trade	All	All
40128	44	3.691	100-499	C-corp	Urban	Nonprof Svcs	All	All but Wsc
40129	40	3.806	100-499	C-corp	Urban	Prof Svcs, FIRE	Low, Mod, Avg, Sig	All but NE

Let  $R_2$  denote the set of stage 2 respondents in a cell, and let  $NR_2$  denote the stage 2 nonrespondents within the cell. The nonresponse-adjusted weights for the cell are calculated as:

$$W_{7i} = W_{6i} \frac{\sum_{R_2 \cup NR_2} W_{6j}}{\sum_{R_2} W_{6j}}, \quad i \in R_2,$$

$$= 0 \quad , \quad i \in NR_2.$$

Here  $W_{7i}$  is the stage 2 nonresponse-adjusted weight, while  $W_{6i}$  is the stage 2 eligibility-adjusted weight as defined in Section 8.14. Notice that the adjustment factor is the reciprocal of a stage 2 response rate within the cell, but not necessarily the same as the response rates shown in Chapter 7, which were unweighted. For stage 2, the response rate implied in the adjustment factor is the weighted counterpart of definition 1 or 2 of

AAPOR (1998), depending on how partial interviews are counted. Refer to the footnote in Section 8.12 for the AAPOR definitions.

The sum of the final weights is an estimate of the eligible universe. The sum of  $W_7$  is 5,291,246 firms.

### **8.16 Weight Smoothing**

Increased variability in the weights increases the design effect. Some variability among the weights is inevitable and is not necessarily considered a problem. If, however, a small number of firms appear to have a disproportionate influence on the estimates or a significant effect on the variances, then one might choose to smooth the outlier weights at the end of the weighting process. For example, minority-owned businesses identified in a non-minority stratum may have a weight far larger than firms in minority strata. Smoothing the weights may introduce a small bias, but the intent is to reduce the variance so that the mean squared error is reduced.

One way to smooth the weights is by trimming the outlier values to a predetermined constant or to the next largest non-outlier value in the same adjustment cell or stratum. Trimming reduces the estimates of totals and introduces a bias. Trimming is done when the reduction in variance more than offsets the bias, and when proportions are the analytical goal rather than totals.

Another possible approach to smoothing is to apply the weight removed from the outliers to other cases in similar subsets of the sample. In this way estimates of some totals can be preserved, in spite of bias. One way to accomplish this is through shrinkage estimators in which every weight is pulled toward a mean weight. A third approach, described in Small Business Survey Group (1999) for a design with extensive stratification errors, essentially evens the median weights among cases from different sampling strata; this approach is probably the most drastic.

Kish (1992) briefly discussed trimming and shrinking weight adjustments. One of the difficulties is that trimming and shrinking are often done for weights for specific analytical variables. This leads to multiple sets of weights, or weights that are good for one dimension and not so good for another dimension. Before deciding on a method, we first analyzed the weights to determine whether weight smoothing was required.

Using updated information from the main interview regarding a firm's characteristics or its owner's demographics, we compared mean weights of subsets by their sampling strata. We reviewed the distributions of weights for subsets of the sample, evaluating the

severity of the outliers. We also checked for cases whose final weights each accounted for more than 1% of the total weight for major subsets of the sample.

There was a small number of cases whose weights could be considered outliers for their analytical subgroups. Table 8.29 gives the largest weight for each analytical subgroup, where a subgroup is defined by the most recent data available. The table shows which sampling stratum the case came from, the percentage of total subgroup weight that this case was responsible for, and the weight divided by the mean and largest weight for cases whose sampling stratum matched the analytical subgroup.

**Table 8.29 Review of Cases with Largest Weights**

Final Analytical Subgroup*	Largest Weight	Sampling Stratum	Percent of Final Subgroup* Total Weight	Ratio to Accurately Stratified Subgroup** Mean Weight	Ratio to Accurately Stratified Subgroup** Largest Weight	Accurately Stratified Subgroup* * Size
Hispanic	3879	Other	1.3%	3.42	2.07	250
Asian	2382	Other	<1%	2.25	1.16	220
African-Amer.	2511	Other	1.3%	3.28	1.63	236
Other	5932	Other	<1%	3.65	1.00	2796
1-19 employees	5932	1-19	<1%	3.21	1.00	2621
20-49 employees	3258	20-49	1.1%	4.16	1.00	333
50-99 employees	2148	50-99	2.6%	7.86	1.00	253
100-499 employ.	2597	100-499	4.2%	11.65	1.00	243
sole proprietors	5267	sole prop	<1%	2.86	1.00	1326
partnerships	5932	unknown	1.6%	3.55	1.20	159
S-corporations	5078	S-corp	<1%	4.30	1.00	721
C-corporations	5088	partner	<1%	4.26	1.28	702

\* The “final subgroup” or “final analytical subgroup” refers to updated classifications after data collection was completed. For example, if a screener proxy did not know the owner’s race and ethnicity, the owner was stratified as “other” for stage 2. If during the interview the owner said he was Hispanic, then his final analytical subgroup is Hispanic, regardless of the stage 2 stratum.

\*\* The term “accurately stratified subgroup” in these columns refers to only those cases whose sampling stratum was correct for the labeled subgroup. The accurately stratified subgroups are subsets of the final analytical subgroups. The Hispanic owner who was incorrectly stratified as “other” is not part of the accurately stratified Hispanic subgroup, nor is he part of the accurately stratified Other subgroup.

Typically NORC does not adjust the largest weights when they are as close to the mean weights as these were. The largest ratios relative to the means occurred for the larger

size classes, where the distribution of weights was highly skewed. Most of the differences in the weights were due to sampling proportions, not to response rates. Thus these differences were intentional, in a sense, and not at all unusual. Since the largest weights were not so extreme for SSBF, any benefits gained from trimming or shrinking probably are not worth the complexity. Thus we did not smooth the weights, and  $W_7$  is the final weight.

### **8.17 Quality Control**

In order to determine the sampling rates for each of the screening sample strata, we had to determine the universe counts from the DMI file. D&B provided counts on the refined DMI files, broken out into 91 sampling strata and 1,225 smaller subgroups. We also obtained subgroup totals from the D&B web site. Although the totals from the D&B web site did not match the totals provided by D&B due to differences in the way duplicates were handled, the web totals did provide a check of reasonableness for D&B's totals in carrying out the frame instructions. The totals are not an estimate of the eligible universe.

After allocating sample specifications by raking, we checked that the allocated totals summed to the correct total sample size, and that the allocations by subgroup maintained the desired percentages.

We checked the D&B sample file for ineligible cases, but found no firms that should have been excluded by D&B on the basis of SIC, geography, subsidiary status, etc. We also computed frequencies on the stage 1 sample file and compared them with the specifications to verify that the counts and percentages were as expected.

During and after screening, we compared the sample percentages with our initial assumptions and with the experience of prior rounds. If our estimated universe proportions were far different from what we initially expected, we might have considered revising the distribution of cases in stage 2 replicates. Although we ended up with a slightly different proportion of urban and rural minority-owned firms than expected, we ultimately selected all minority-owned firms from the screener into the stage 2 sample. Otherwise, the stage 2 sample proportions were based on the proportions found in the first 200 replicates of the screener.

Weighted analyses are appropriate for sample designs with unequal probabilities of selection. Proper calculation of the weights is essential. Throughout the weighting process, NORC conducted a number of simple quality checks to ensure that the weights were being calculated correctly. For example, the sum of the weights was calculated at

each step to check for consistency and reasonableness. The sum of the stage 1 base weights matched the frame totals provided by D&B. The sum of the nonresponse-adjusted weights exactly matched the sum of the eligibility-adjusted weights prior to nonresponse adjustment. Another simple test checked the calculations themselves. Adjustment factors were rederived by computing the ratios of adjusted to unadjusted weights. Plots and other checks of the distribution of weights at various stages were performed to identify outliers and other anomalies.

Many of the assumptions built into the sampling and weighting plan were verified after data collection. This methodology report contains tables of eligibility rates (Table 6.8), response rates (Chapter 6), stratification errors (Tables 8.12 - 8.15), and other measures that were used to evaluate the assumptions. Many of these tables are included in this chapter. The evaluations will be useful for planning the next round of SSBF.



## REFERENCES

The American Association for Public Opinion Research (1998). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for RDD Telephone Surveys and In-Person Household Surveys*. Ann Arbor, Michigan: AAPOR.

(<http://www.aapor.org/ethics/stddef.html>)

Grigorian, K., Haggerty, C., and Wolken, J. (2000). "SSBF: The 1998 Survey of Small Business Finances." Presented at the meetings of the American Association for Public Opinion Research, Portland, OR, May 2000.

Small Business Survey Group (1999). *Codebook for 1993 National Survey of Small Business Finances (NSSBF)*. Washington, DC: Board of Governors of the Federal Reserve System.

Kish, L. (1992). "Weighting for Unequal Pi." [i is a subscript] in *Journal of Official Statistics*, Vol. 8, No. 2, pp. 183-200. Statistics Sweden.

## SAMPLING APPENDIX – ILLUSTRATION OF SUBSAMPLING AND WEIGHTING

The second stage sampling procedures and final weight calculations are illustrated with an artificial example. Figure 1 lists the stage 1 sample units for this example. The units were selected from the strata with the probabilities shown. The base weights are the reciprocals of the probability of selection. The cases were assigned permanent random numbers, sorted by the random numbers, and assigned to stage 1 replicates based on the random sort order.

**Figure 1. Stage 1 Sample for Screening**

Case ID	Stratum	Probability	Base Weight	PRN	Replicate
10000220	1	0.2	5.00	0.022678	1
10000350	1	0.2	5.00	0.048658	1
10000020	1	0.2	5.00	0.052512	1
10000450	1	0.2	5.00	0.066893	1
10000370	2	0.4	2.50	0.075071	1
10000440	1	0.2	5.00	0.078250	1
10000490	2	0.4	2.50	0.081624	1
10000040	1	0.2	5.00	0.082994	1
10000800	2	0.4	2.50	0.087266	1
10000120	1	0.2	5.00	0.094377	1
10000100	2	0.4	2.50	0.110228	2
10000540	1	0.2	5.00	0.130231	2
10000290	2	0.4	2.50	0.146493	2
10000060	2	0.4	2.50	0.154750	2
10000750	2	0.4	2.50	0.159935	2
10000070	2	0.4	2.50	0.188004	2
10000880	2	0.4	2.50	0.189275	2
10000200	1	0.2	5.00	0.193399	2
10000500	2	0.4	2.50	0.229146	2
10000480	1	0.2	5.00	0.255317	2
10000080	2	0.4	2.50	0.260083	3
10000560	1	0.2	5.00	0.275559	3
10000980	2	0.4	2.50	0.279883	3
10000630	1	0.2	5.00	0.302375	3
10000210	1	0.2	5.00	0.302906	3
10000720	1	0.2	5.00	0.307770	3
10000580	1	0.2	5.00	0.311606	3
10000520	2	0.4	2.50	0.332392	3
10000660	2	0.4	2.50	0.344567	3
10000810	2	0.4	2.50	0.344809	3
10000280	1	0.2	5.00	0.345002	4
10000610	2	0.4	2.50	0.350249	4
10000530	2	0.4	2.50	0.361872	4
10000240	1	0.2	5.00	0.379936	4
10000740	1	0.2	5.00	0.382952	4
10000470	1	0.2	5.00	0.394660	4
10000690	1	0.2	5.00	0.398768	4
10000940	1	0.2	5.00	0.400343	4
10000250	1	0.2	5.00	0.407765	4
10000330	1	0.2	5.00	0.415168	4
10000550	1	0.2	5.00	0.429430	5
10000030	1	0.2	5.00	0.436292	5
10000190	2	0.4	2.50	0.442496	5
10000140	1	0.2	5.00	0.457137	5

Case ID	Stratum	Probability	Base Weight	PRN	Replicate
10000420	1	0.2	5.00	0.483855	5
10000460	1	0.2	5.00	0.495999	5
10000050	1	0.2	5.00	0.496444	5
10000640	1	0.2	5.00	0.504101	5
10000960	2	0.4	2.50	0.506076	5
10000820	2	0.4	2.50	0.507257	5
10000340	1	0.2	5.00	0.507506	6
10000870	1	0.2	5.00	0.508228	6
10000730	1	0.2	5.00	0.520323	6
10000860	1	0.2	5.00	0.520961	6
10000010	1	0.2	5.00	0.523471	6
10000850	1	0.2	5.00	0.545241	6
10000670	2	0.4	2.50	0.558844	6
10000170	2	0.4	2.50	0.565172	6
10000260	2	0.4	2.50	0.567990	6
10000390	1	0.2	5.00	0.586952	6
10000110	1	0.2	5.00	0.592509	7
10000900	2	0.4	2.50	0.595035	7
10000770	1	0.2	5.00	0.614345	7
10000840	1	0.2	5.00	0.615531	7
10000700	1	0.2	5.00	0.634353	7
10000150	1	0.2	5.00	0.664652	7
10000780	1	0.2	5.00	0.666333	7
10000270	1	0.2	5.00	0.672346	7
10000410	2	0.4	2.50	0.678215	7
10000680	2	0.4	2.50	0.686267	7
10000380	2	0.4	2.50	0.692264	8
10000970	2	0.4	2.50	0.718608	8
10000620	1	0.2	5.00	0.726025	8
10000760	2	0.4	2.50	0.732056	8
10000180	2	0.4	2.50	0.747837	8
10000710	1	0.2	5.00	0.751013	8
10000300	2	0.4	2.50	0.755140	8
10000310	2	0.4	2.50	0.773551	8
10000650	2	0.4	2.50	0.774850	8
10000160	2	0.4	2.50	0.778283	8
10000950	2	0.4	2.50	0.781280	9
10000230	1	0.2	5.00	0.789512	9
10000590	1	0.2	5.00	0.795185	9
10000600	2	0.4	2.50	0.808327	9
10001000	1	0.2	5.00	0.817494	9
10000930	1	0.2	5.00	0.819736	9
10000360	2	0.4	2.50	0.820462	9
10000790	2	0.4	2.50	0.824809	9
10000990	1	0.2	5.00	0.843679	9
10000090	2	0.4	2.50	0.858017	9
10000910	1	0.2	5.00	0.869463	10
10000130	1	0.2	5.00	0.876541	10
10000320	2	0.4	2.50	0.881638	10
10000430	1	0.2	5.00	0.885336	10
10000400	2	0.4	2.50	0.885812	10
10000890	2	0.4	2.50	0.901276	10
10000920	1	0.2	5.00	0.916377	10
10000510	2	0.4	2.50	0.947879	10
10000830	2	0.4	2.50	0.963325	10
10000570	1	0.2	5.00	0.973918	10

Suppose that replicates 2 and 3 are screened first and constitute the first batch. On the basis of these replicates, we estimate the marginal totals and determine the desired sample allocation for

stage 2. Figure 2 illustrates the calculation of interim weights for the two replicates and the calculation of marginal totals for stage 2 strata. The stage 1 strata serve as the major strata for stage 2, but the strata are further divided into substrata A, B, C, and D on the basis of screener information. Notice that one case in the example was sampled in stratum 2, but was reassigned to stratum 1 as a result of screening. The case retains its original stratum for these interim weight calculations.

The base weights are adjusted for the fact that only two replicates are available. There are 100 cases in the entire stage 1 sample, but only 20 in the two replicates, so the weights are adjusted upward by a factor of  $5=100/20$ . The weights of ineligible cases are set to zero. The weights of cases of unknown eligibility are adjusted downward by the proportion of confirmed eligible cases in the stratum. Then the nonrespondents' weights are set to zero, and the weights of the completed, eligible cases are adjusted upward to account for the weights of the nonrespondents.

The marginal totals for the stage 2 strata are the sums of the nonresponse-adjusted interim weights for the cases in those strata, using the corrected screener information for classifying cases to strata. The proportions are used to allocate sample for stage 2.

Figure 3 shows the information needed for stage 2 sample allocation. First, the estimated marginal totals are repeated. Second, the expected eligibility rate for stage 2 is shown, although the derivation of these figures is not illustrated. Third, the expected completion rates for stage 2 are given. The completion rates have been assumed from the beginning and are not derived from the current data. The ideal final sample sizes (completed cases) are given as project requirements. The stage 2 target sample sizes are obtained by dividing the ideal completed cases by the expected completion and eligibility rates. The target sample figures are rounded. The completed cases by stratum are the actual unweighted counts from the first two completed replicates. The expected maximum available from a stratum after all screening is complete is the actual count for the first two completed replicates adjusted up to the total number of cases in all screeners, or a factor of  $5=100/20$ .

**Figure 2. Interim Stage 1 Weight Adjustments for Stage 2 Sample Allocation**

Case ID	Stage 1 Stratum	PRN	Stage 1 Replicate	W1 Base Weight	Screener Disp	W2 Base Wts For 2 Reps	Adjust for Eligibility	W3 Wts Adj for Eligibility	Adjust for NR	W4 Wts Adj for NR	Stage 2 Stratum
10000480	1	0.255317	2	5.00	eligible	25.00	1.00	25.00	1.60	40.00	A
10000720	1	0.307770	3	5.00	eligible	25.00	1.00	25.00	1.60	40.00	B
10000560	1	0.275559	3	5.00	eligible	25.00	1.00	25.00	1.60	40.00	A
10000580	1	0.311606	3	5.00	eligible	25.00	1.00	25.00	1.60	40.00	B
10000540	1	0.130231	2	5.00	unknown	25.00	0.80	20.00	0.00	0.00	
10000630	1	0.302375	3	5.00	ineligible	25.00	0.00	0.00		0.00	
10000210	1	0.302906	3	5.00	unknown	25.00	0.80	20.00	0.00	0.00	
10000200	1	0.193399	2	5.00	unknown	25.00	0.80	20.00	0.00	0.00	
10000750	2	0.159935	2	2.50	eligible	12.50	1.00	12.50	1.09	13.64	C
10000100	2	0.110228	2	2.50	eligible	12.50	1.00	12.50	1.09	13.64	C
10000060	2	0.154750	2	2.50	eligible	12.50	1.00	12.50	1.09	13.64	D
10000810	2	0.344809	3	2.50	eligible	12.50	1.00	12.50	1.09	13.64	D
10000520	2	0.332392	3	2.50	eligible	12.50	1.00	12.50	1.09	13.64	C
10000080	2	0.260083	3	2.50	eligible	12.50	1.00	12.50	1.09	13.64	D
10000500	2	0.229146	2	2.50	eligible	12.50	1.00	12.50	1.09	13.64	A
10000880	2	0.189275	2	2.50	ineligible	12.50	0.00	0.00		0.00	
10000980	2	0.279883	3	2.50	ineligible	12.50	0.00	0.00		0.00	
10000070	2	0.188004	2	2.50	unknown	12.50	0.64	7.95	0.00	0.00	
10000660	2	0.344567	3	2.50	ineligible	12.50	0.00	0.00		0.00	
10000290	2	0.146493	2	2.50	ineligible	12.50	0.00	0.00		0.00	

**Figure 3. Setup for Stage 2 Sampling Based on First Completed Replicate Results**

Estimated Marginal Totals (sum of weights)	Major Stratum 1		Stratum A	93.64	53.93%
	Major Stratum 1		Stratum B	80.00	46.07%
	Major Stratum 2		Stratum C	40.91	50.00%
	Major Stratum 2		Stratum D	40.91	50.00%
Expected Eligibility Rate for Stage 2 (assumed)			Major Stratum 1	0.95	
Expected Eligibility Rate for Stage 2 (assumed)			Major Stratum 2	0.90	
Expected Completion Rate for Stage 2 (assumed)			Major Stratum 1	0.75	
Expected Completion Rate for Stage 2 (assumed)			Major Stratum 2	0.75	
Ideal Sample Completes By Major Stratum (given)			Major Stratum 1	10	
Ideal Sample Completes By Major Stratum (given)			Major Stratum 2	10	
					Rounded
Target Stage 2 Sample Per Major Stratum (ideal completes / exp comp rate / exp elig rate)			Major Stratum 1	14.04	14
Target Stage 2 Sample Per Major Stratum (ideal completes / exp comp rate / exp elig rate)			Major Stratum 2	14.81	15
Ideal Total Stage 2 Sample Size					29
Completed Eligible Screeners By Stage 2 Stratum (observed)			Stratum A	3	
Completed Eligible Screeners By Stage 2 Stratum (observed)			Stratum B	2	
Completed Eligible Screeners By Stage 2 Stratum (observed)			Stratum C	3	
Completed Eligible Screeners By Stage 2 Stratum (observed)			Stratum D	3	
Estimated Max Available Eligible Screeners By Stage 2 Stratum for All Stage 1 Reps (observed x 100 / 20)			Stratum A	15	
Estimated Max Available Eligible Screeners By Stage 2 Stratum for All Stage 1 Reps (observed x 100 / 20)			Stratum B	10	
Estimated Max Available Eligible Screeners By Stage 2 Stratum for All Stage 1 Reps (observed x 100 / 20)			Stratum C	15	
Estimated Max Available Eligible Screeners By Stage 2 Stratum for All Stage 1 Reps (observed x 100 / 20)			Stratum D	15	

**Figure 4. Results of Raking Steps for Stage 2 Sample Allocation**

	Stratum A	Stratum B	Major	Stratum C	Stratum D	Major
Raking Criterion	53.93%	46.07%	Stratum 1	50.00%	50.00%	Stratum 2
Starting values	5.00	5.00	10.00	5.00	5.00	10.00
Marginal proportions?	5.39	4.61	10.00	5.00	5.00	10.00
Exceed Max Available?	5.39	4.61	10.00	5.00	5.00	10.00
Major Strat Targets Met?	7.55	6.45	14.00	7.50	7.50	15.00
Marginal proportions?	7.55	6.45	14.00	7.50	7.50	15.00
Exceed Max Available?	7.55	6.45	14.00	7.50	7.50	15.00
Select	7.55	6.45	14.00	7.50	7.50	15.00

Figure 4 is an illustration of the raking process to achieve at least 14-15 cases for each of the major strata, with proportional allocation of substrata within major strata, and not exceeding the expected maximum numbers of cases available. The starting values are arbitrary. Raking for marginal proportions adjusts the allocations within each major stratum to the proportions shown at the top of the columns, maintaining the total for the major stratum. Raking for maximum available reduces the allocation for a substratum if the maximum is exceeded. Raking for the major stratum targets raises the allocation for all substrata within a major stratum to reach the target number of 14 or 15 cases. Raking continues iteratively until the table converges. The last row in Figure 4 is the overall sample allocation for stage 2.

Figure 5 shows the allocation proportions scaled to replicate level. For operational reasons, the replicates will be of size four in this example. Notice that two of the replicate allocations are not integers.

**Figure 5. Allocation to Replicates**

	Replicate Size	Replicate Allocation
Major Stratum 1	4	
Stratum A		2.16
Stratum B		1.84
Major Stratum 2	4	
Stratum A		2
Stratum B		2

Figure 6 illustrates the assignment of cases to stage 2 replicates, in the order in which the stage 1 replicates are finalized. Actually, the order of finalization within a batch is not important, as long as the batch order is maintained. Within stage 1 replicates, eligible cases in this example are assigned to stage 2 replicates in order of the permanent random numbers assigned earlier. Any replicable random ordering will do. Any string of cases in random number order is a random subsample, so the use of PRNs for both first and second stage replicates preserves the fact that each replicate is a random subsample of eligible cases from the original sample.

The stage 2 replicates are not all the same size or exactly the same composition by substrata. A counter tracks the number of eligible cases in a substratum in order. When the counter exceeds a multiple of the replicate allocation for that stratum (Figure 5), the cases are assigned to the next replicate. For example, the replicate allocation for stratum A is 2.16. The first two cases in PRN order are assigned to replicate 1. For the third case, the counter exceeds 2.16, so the case is assigned to replicate 2. Replicate 2 for each major stratum is not complete using only the batch 1 cases and is not released for interviewing until batch 2 is assigned to replicates. Figure 7 summarizes the number of cases per stage 2 replicate for each of the substrata within the major strata

**Figure 6. Assigning Eligibles to Stage 2 Replicates as Stage 1 Replicates are Finalized**

Batch	Case ID	PRN	Stage 1 Replicate	Completion Order	Stage 2 Replicate
1	10000500	0.229146	2	1	1
1	10000480	0.255317	2	1	1
1	10000560	0.275559	3	2	2
2	10000020	0.052512	1	3	2
2	10000450	0.066893	1	3	3
2	10000470	0.394660	4	4	3
2	10000940	0.400343	4	4	4
2	10000340	0.507506	6	5	4
2	10000860	0.520961	6	5	5
2	10000390	0.586952	6	5	5
2	10000550	0.429430	5	6	6
2	10000050	0.496444	5	6	6
2	10000110	0.592509	7	7	7
2	10000840	0.615531	7	7	7
2	10000710	0.751013	8	8	7
2	10000230	0.789512	9	9	8
2	10000590	0.795185	9	9	8
2	10000910	0.869463	10	10	9
1	10000720	0.307770	3	2	1
1	10000580	0.311606	3	2	2
2	10000350	0.048658	1	3	2
2	10000280	0.345002	4	4	3
2	10000740	0.382952	4	4	3
2	10000250	0.407765	4	4	4
2	10000330	0.415168	4	4	4
2	10000730	0.520323	6	5	5
2	10000030	0.436292	5	6	5
2	10000460	0.495999	5	6	6
2	10000150	0.664652	7	7	6
2	10000780	0.666333	7	7	7
2	10000620	0.726025	8	8	8
2	10001000	0.817494	9	9	8
2	10000430	0.885336	10	10	9
1	10000100	0.110228	2	1	1
1	10000750	0.159935	2	1	1
1	10000520	0.332392	3	2	2
2	10000370	0.075071	1	3	2
2	10000530	0.361872	4	4	3
2	10000670	0.558844	6	5	3
2	10000410	0.678215	7	7	4
2	10000380	0.692264	8	8	4
2	10000160	0.778283	8	8	5
2	10000360	0.820462	9	9	5
2	10000510	0.947879	10	10	6
1	10000060	0.154750	2	1	1
1	10000080	0.260083	3	2	1
1	10000810	0.344809	3	2	2
2	10000800	0.087266	1	3	2
2	10000610	0.350249	4	4	3
2	10000190	0.442496	5	6	3
2	10000960	0.506076	5	6	4
2	10000900	0.595035	7	7	4
2	10000970	0.718608	8	8	5
2	10000650	0.774850	8	8	5
2	10000950	0.781280	9	9	6
2	10000600	0.808327	9	9	6
2	10000400	0.885812	10	10	7



**Figure 7. Number of Cases Per Replicate By Stratum**

Major Stratum 1			Major Stratum 2		
Rep	Stratum A	Stratum B	Rep	Stratum C	Stratum D
1	2	1	1	2	2
2	2	2	2	2	2
3	2	2	3	2	2
4	2	2	4	2	2
5	2	2	5	2	2
6	2	2	6	1	2
7	3	1	7		1
8	2	2			
9	1	1			

Figure 8 illustrates the calculations for the final weights. All stage 1 replicates were screened, so there is no adjustment to base weights for stage 1 replicates. The base weights are adjusted for eligibility and nonresponse at stage 1 using methods comparable to those used to calculate interim weights on batch 1. This time, however, the screener eligibility adjustments are based on eligibility proportions of a follow-up sample of nonrespondents rather than the eligibility proportions of the screener respondents.. The nonresponse-adjusted weights for stage 1 are the base weights for stage 2. Not all replicates were used for stage 2, so the weights are adjusted for the number of replicates before applying another eligibility and nonresponse adjustment. For stage 1 eligibility adjustments, the eligibility rates applied to the cases of unknown eligibility are shown in the eligible column, based on the confirmed cases. (We actually used the eligibility rate from the 5% nonresponse follow-up, but the method is the same.) For stage 2, all nonrespondents are considered eligible.

The sums of the weights are used for checking the work at each stage. The sum of the base weights is an estimate of the total frame size, prior to stage 1 sampling. The stage 1 weights adjusted for eligibility sum to an estimate of the eligible population. Nonresponse adjustments preserve that total. After subsampling, the replicate-adjusted weights sum to another estimate of the eligible population, and the sum should approximate the stage 1 nonresponse-adjusted weights. Eligibility adjustments at stage 2 will reduce the estimate of the total eligible population somewhat. Finally, the nonresponse adjustments should maintain the total.

**Figure 8. Calculation of Final Weights, Assuming All Stage 1 Replicates Are Used**

Case ID	Stage 1 Sampling Stratum	Rev Major Stratum	W1 Base Weight	Stage 1 Disp	W3 Wts Adj for Eligibility	W4 Wts Adj for NR	Stage 2 Stratum	Stage 2 Rep	Reps Used	W5 Subsamp Wts	Stage 2 Disp	W6 Wts Adj for Eligibility	W7 Wts Adj for NR
0000500	2	1	2.5	eligible	2.50	3.04	A	1	Y	5.28	complete	5.28	5.90
0000480	1	1	5	eligible	5.00	6.08	A	1	Y	10.56	complete	10.56	11.80
0000020	1	1	5	eligible	5.00	6.08	A	2	Y	10.56	complete	10.56	11.80
0000560	1	1	5	eligible	5.00	6.08	A	2	Y	10.56	complete	10.56	11.80
0000470	1	1	5	eligible	5.00	6.08	A	3	Y	10.56	complete	10.56	11.80
0000450	1	1	5	eligible	5.00	6.08	A	3	Y	10.56	complete	10.56	11.80
0000340	1	1	5	eligible	5.00	6.08	A	4	Y	10.56	complete	10.56	11.80
0000940	1	1	5	eligible	5.00	6.08	A	4	Y	10.56	complete	10.56	11.80
0000390	1	1	5	eligible	5.00	6.08	A	5	Y	10.56	complete	10.56	11.80
0000860	1	1	5	eligible	5.00	6.08	A	5	Y	10.56	NR	10.56	0.00
0000550	1	1	5	eligible	5.00	6.08	A	6		0.00	out	0.00	0.00
0000050	1	1	5	eligible	5.00	6.08	A	6		0.00	out	0.00	0.00
0000710	1	1	5	eligible	5.00	6.08	A	7		0.00	out	0.00	0.00
0000110	1	1	5	eligible	5.00	6.08	A	7		0.00	out	0.00	0.00
0000840	1	1	5	eligible	5.00	6.08	A	7		0.00	out	0.00	0.00
0000590	1	1	5	eligible	5.00	6.08	A	8		0.00	out	0.00	0.00
0000230	1	1	5	eligible	5.00	6.08	A	8		0.00	out	0.00	0.00
0000910	1	1	5	eligible	5.00	6.08	A	9		0.00	out	0.00	0.00
0000720	1	1	5	eligible	5.00	6.08	B	1	Y	10.56	complete	10.56	14.08
0000580	1	1	5	eligible	5.00	6.08	B	2	Y	10.56	ineligible	0.00	0.00
0000350	1	1	5	eligible	5.00	6.08	B	2	Y	10.56	complete	10.56	14.08
0000740	1	1	5	eligible	5.00	6.08	B	3	Y	10.56	complete	10.56	14.08
0000280	1	1	5	eligible	5.00	6.08	B	3	Y	10.56	NR	10.56	0.00
0000330	1	1	5	eligible	5.00	6.08	B	4	Y	10.56	complete	10.56	14.08
0000250	1	1	5	eligible	5.00	6.08	B	4	Y	10.56	complete	10.56	14.08
0000730	1	1	5	eligible	5.00	6.08	B	5	Y	10.56	NR	10.56	0.00
0000030	1	1	5	eligible	5.00	6.08	B	5	Y	10.56	complete	10.56	14.08
0000460	1	1	5	eligible	5.00	6.08	B	6		0.00	out	0.00	0.00
0000150	1	1	5	eligible	5.00	6.08	B	6		0.00	out	0.00	0.00
0000780	1	1	5	eligible	5.00	6.08	B	7		0.00	out	0.00	0.00
0000620	1	1	5	eligible	5.00	6.08	B	8		0.00	out	0.00	0.00
0001000	1	1	5	eligible	5.00	6.08	B	8		0.00	out	0.00	0.00
0000430	1	1	5	eligible	5.00	6.08	B	9		0.00	out	0.00	0.00
0000140	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
0000270	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
0000240	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
0000440	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00

Case ID	Stage 1 Sampling Stratum	Rev Major Stratum	W1 Base Weight	Stage 1 Disp	W3 Wts Adj for Eligibility	W4 Wts Adj for NR	Stage 2 Stratum	Stage 2 Rep	Reps Used	W5 Subsamp Wts	Stage 2 Disp	W6 Wts Adj for Eligibility	W7 Wts Adj for NR
I0000040	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000640	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000700	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000850	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000870	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000990	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000130	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000420	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000920	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000630	1	1	5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000570	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000010	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000120	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000200	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000210	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000220	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000540	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000690	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000770	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000930	1	1	5	unknown	3.51	0.00				0.00	out	0.00	0.00
I0000750	2	2	2.5	eligible	2.50	2.99	C	1	Y	3.12	complete	3.12	4.45
I0000100	2	2	2.5	eligible	2.50	2.99	C	1	Y	3.12	complete	3.12	4.45
I0000370	2	2	2.5	eligible	2.50	2.99	C	2	Y	3.12	ineligible	0.00	0.00
I0000520	2	2	2.5	eligible	2.50	2.99	C	2	Y	3.12	complete	3.12	4.45
I0000670	2	2	2.5	eligible	2.50	2.99	C	3	Y	3.12	complete	3.12	4.45
I0000530	2	2	2.5	eligible	2.50	2.99	C	3	Y	3.12	NR	3.12	0.00
I0000410	2	2	2.5	eligible	2.50	2.99	C	4	Y	3.12	NR	3.12	0.00
I0000380	2	2	2.5	eligible	2.50	2.99	C	4	Y	3.12	complete	3.12	4.45
I0000160	2	2	2.5	eligible	2.50	2.99	C	5	Y	3.12	complete	3.12	4.45
I0000360	2	2	2.5	eligible	2.50	2.99	C	5	Y	3.12	NR	3.12	0.00
I0000510	2	2	2.5	eligible	2.50	2.99	C	6	Y	3.12	complete	3.12	4.45
I0000080	2	2	2.5	eligible	2.50	2.99	D	1	Y	3.12	complete	3.12	3.74
I0000060	2	2	2.5	eligible	2.50	2.99	D	1	Y	3.12	complete	3.12	3.74
I0000800	2	2	2.5	eligible	2.50	2.99	D	2	Y	3.12	complete	3.12	3.74
I0000810	2	2	2.5	eligible	2.50	2.99	D	2	Y	3.12	complete	3.12	3.74
I0000190	2	2	2.5	eligible	2.50	2.99	D	3	Y	3.12	NR	3.12	0.00
I0000610	2	2	2.5	eligible	2.50	2.99	D	3	Y	3.12	complete	3.12	3.74
I0000960	2	2	2.5	eligible	2.50	2.99	D	4	Y	3.12	complete	3.12	3.74

Case ID	Stage 1 Sampling Stratum	Rev Major Stratum	W1 Base Weight	Stage 1 Disp	W3 Wts Adj for Eligibility	W4 Wts Adj for NR	Stage 2 Stratum	Stage 2 Rep	Reps Used	W5 Subsamp Wts	Stage 2 Disp	W6 Wts Adj for Eligibility	W7 Wts Adj for NR
I0000900	2	2	2.5	eligible	2.50	2.99	D	4	Y	3.12	NR	3.12	0.00
I0000970	2	2	2.5	eligible	2.50	2.99	D	5	Y	3.12	complete	3.12	3.74
I0000650	2	2	2.5	eligible	2.50	2.99	D	5	Y	3.12	complete	3.12	3.74
I0000950	2	2	2.5	eligible	2.50	2.99	D	6	Y	3.12	complete	3.12	3.74
I0000600	2	2	2.5	eligible	2.50	2.99	D	6	Y	3.12	complete	3.12	3.74
I0000400	2	2	2.5	eligible	2.50	2.99	D	7		0.00	out	0.00	0.00
I0000090	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000180	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000310	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000660	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000760	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000790	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000260	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000980	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000290	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000880	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000830	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000890	2	2	2.5	ineligible	0.00	0.00				0.00	out	0.00	0.00
I0000320	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
I0000070	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
I0000170	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
I0000300	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
I0000490	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
I0000680	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
I0000820	2	2	2.5	unknown	1.67	0.00				0.00	out	0.00	0.00
<b>Sum of Weights</b>			<b>390</b>		<b>269</b>	<b>269</b>				<b>267</b>		<b>253</b>	<b>253</b>