U.S. Department
of Labor

Employment
and Training
Administration

O∗net™

O*NET® Data Collection Program

Office of Management and Budget
Clearance Package Supporting
Statement and Data Collection
Instruments

July 11, 2002

O*net®

**U.S. Department
of Labor**

**Employment
and Training
Administration**

O*NET® Data
Collection Program

**Office of Management and Budget
Clearance Package Supporting Statement
and Data Collection Instruments**

July 11, 2002

Table of Contents

i

Table of Contents (continued)

List of Exhibits

List of Acronyms

AIR .... American Institutes for Research
BL.................. Business Liaison
BLS................ Bureau of Labor Statistics
CIS................. Career Information System
CMS .... Case Management System
CIP................. Classification of Instructional Programs
CFR .... Code of Federal Regulation
CSM .... Composite Size Measure
CTDOL.......... Connecticut Department of Labor
DOL............... Department of Labor
DOT............... *Dictionary of Occupational Titles*
ETA .... Employment and Training Administration
FAQ............... Frequently Asked Questions
GL.................. Gatekeeper Liaisons
I/O................. Industrial/Organizational
IT .... Information Technology
OES .... Occupational Employment Statistics
OMB.............. Office of Management and Budget
O*NET .... Occupational Information Network
OSCAR.......... Occupation and Skill Computer-Assisted Researcher

List of Acronyms (continued)

OU .... Occupational Unit
OVAE .... Office of Vocational and Adult Education
POC .... Point of Contact
PPS .... Probability Proportionate to Size
SSL .... Secure Socket Layer
SIC .... Standard Industrial Code
SOC .... Standard Occupational Classification
STAR .... Strategies and Tactics for Averting Refusals
SME .... Subject Matter Expert

A.1   Circumstances of Information Collection

This is a request for 3-year OMB clearance to continue with the *Occupational Information Network* (O*NET®) *Data Collection Program*. Data are currently being collected for approximately 201 occupations, and this request is to continue the collection of data for an additional 463 O*NET occupations over the next 3 years (Fiscal Years 2003–2005), subject to annual budget levels. Data for the remaining 310 O*NET occupations will be collected in FY2006 and FY2007. The O*NET *Data Collection Program* is an ongoing activity to populate and maintain a current database on the detailed characteristics of workers, occupations, and skills. The continued population of the O*NET database is important because the resulting updated O*NET database will be the most comprehensive standard source of occupational information in the U.S. The O*NET *Data Collection Program* is at the center of an extensive network of occupational/skill information used by a wide range of audiences, including individuals making career decisions, public agencies and schools making training investment decisions, educational institutions preparing a future workforce, and employers making staffing and training decisions. This program provides a common language and framework to meet the administrative needs of various federal programs, including workforce investment and training programs of the Departments of Labor and Education. The O*NET database and companion O*NET career exploration tools are being used by many private companies and public organizations to develop applications that use O*NET, tailored to meet their customer needs. We describe the uses of the O*NET Program in further detail in *Section A.2*. Also, information about the O*NET *Data Collection Program* can be found at the National O*NET Consortium's public website, www.onetcenter.org (the O*NET portal page that links to several O*NET-related websites), and the U.S. Department of Labor, Employment and Training Administration (ETA) website, www.doleta.gov/programs/onet.

The O*NET *Data Collection Program* employs a multiple method approach to updating the O*NET database. The primary method employs a two-stage sample design to survey establishments and workers within those establishments. In addition to this primary method, alternative methods include association member lists and subject matter experts (SMEs). These alternative methods have not yet been implemented. In all methods, the O*NET survey instruments are used. More detailed information on these multiple methods is presented in *Sections B.2.11* and *B.2.12*.

In 1999, a pretest was conducted to assess the impact of incentives and other methodological components on response rates. A report documenting the pretest activity and results is included as *Appendix A*.[1]

The remainder of this section describes the O*NET Program and reviews statutory and regulatory information.

A.1.1 What Is the O*NET Program?

The O*NET Program is a comprehensive system for collecting and disseminating information on occupational and worker requirements. The O*NET data replaces the Department of Labor's increasingly outdated *Dictionary of Occupational Titles* (DOT) and provides additional information not available in the DOT. The DOT is no longer updated or maintained by DOL.

As depicted in *Exhibit A-1*, the O*NET Program uses a data structure, the Content Model, to organize occupational information and provides a common language of standardized and defined occupational descriptors and measures for use by all audiences. The O*NET Content Model is the result of extensive research, and its development is fully documented (Peterson, Mumford, Borman, Jeanneret, & Fleishman, 1995; Peterson, Mumford, Borman, et al., 1997).

The Content Model comprises six domains:

1. *Worker Characteristics*: includes Abilities, Interests, and Work Styles;

2. *Worker Requirements*: includes Basic Skills, Cross-Functional Skills, General Knowledge, and Education;

3. *Experience Requirements*: includes Training, Experience, and Licensing Requirements;

4. *Occupation Requirements*: includes Generalized Work Activities, Work Context, and Organizational Context;

5. *Occupation-Specific Requirements*: includes Occupational Knowledges, Occupational Skills, Tasks, Machines, Tools, and Equipment; and

6. *Occupation Characteristics*: includes Labor Market Information, Occupational Outlook, and Wages, based on existing data sources.

---

[1] The appendices to the pretest report are not included in *Appendix A*. See RTI, 2000, for a complete version of the report.

Exhibit A-1. O*NET Content Model

Each domain of the Content Model employs a hierarchical structure used to group information. For example, the Worker Characteristics domain contains three types of information: *Abilities, Interests*, and *Work Styles*. The Abilities domain, in turn, contains four types of abilities: *Cognitive, Psychomotor, Physical*, and *Sensory*. Each of these types of abilities contains further levels of detail. For example, Psychomotor abilities include *Fine Manipulative, Control Movement,* and *Reaction Time and Speed* abilities. Finally, the Fine Manipulative abilities contain three specific descriptors: *Arm-Hand Steadiness, Manual Dexterity,* and *Finger Dexterity*.

The descriptors and rating scales for O*NET data were developed through extensive research, drawing primarily from job analysis in industrial/organizational (I/O) psychology (Peterson et al., 1995). The O*NET Program takes the best knowledge about both content and methodology from the last 60 years of research since the first DOT. The scales used for the O*NET ratings are *Importance*, *Level*, and *Frequency*. Each item (descriptor) in the O*NET questionnaires may use one or more scales. For example, the O*NET skill descriptor "Coordination" is rated on both a 5-point *Importance* scale and a 7-point *Level* scale. Refer to ***Appendix B*** for the complete set of O*NET questionnaires.

Exhibit A-2 summarizes the number of descriptors and scales included in the revised O*NET *Data Collection Program* questionnaires. Data are being collected using 239 descriptors that include 400 scales contained in the revised Content Model domains. Ratings for one of the

revised domain questionnaires, Abilities, are being developed using trained analysts. No data collection is planned for the Occupational Characteristics domain (see ***Section A.1.2*** for a discussion of the preferred data source). Information for Occupational Characteristics will be provided through links to the employment, wage, and long-term projections databases produced by the Bureau of Labor Statistics (BLS), the state employment security agencies, and other agencies.

Exhibit A-2. Summary of O*NET *Data Collection Program* Questionnaires

O*NET *Data Collection Program* Questionnaires	Number of Items and Scales			Data Source
	Number of Items	Number of Scales for Each Item	Total Number of Scales	
Skills	35	2	70	Incumbents[a]
Knowledge	33	2	66	Incumbents[a]
Work Styles[b]	16	1	16	Incumbents[a]
Education & Training[b]	5	1	5	Incumbents[a]
Generalized Work Activities	41	2	82	Incumbents[a]
Work Context	57	1	57	Incumbents[a]
Abilities	52	2	104	Analysts
Tasks[c]	varies	2	varies	Incumbents[a]
Total (not including tasks)	239		400	

[a]   Subject matter experts (SMEs) will rate the same questionnaires as incumbents for those occupations that will use this method.
[b]   The Knowledge Questionnaire packet also includes the Work Styles and Education & Training Questionnaires.
[c]   All incumbents will be asked to complete a task questionnaire in addition to the domain questionnaire they are asked to complete.

The first version of the O*NET database released to the public was O*NET 98. O*NET 98 moved from the DOT's more than 12,000 occupations to a more user-friendly 1,122 occupations, based on the BLS Occupational Employment Statistics (OES) codes. The O*NET 98 database contained 306 descriptors and 684 scales. A review of O*NET 98–specific scales and descriptors during the preparation for pretest data collection led to some consolidation or dropping of descriptors and scales to reduce burden and increase the employee response rate.[2] The O*NET 98 database was first replaced with the O*NET 3.1 database and has subsequently been replaced by the O*NET 4.0 database.

---

[2] See *Revision of O*NET Data Collection Instruments,* available at www.onetcenter.org (Hubbard et al., 2000).

The current database, O*NET 4.0, contains the same data used in O*NET 98; however, the occupations are restructured and coded to the 1998 Standard Occupational Classification (SOC). As a result of the restructuring and recoding, O*NET 4.0 has 974 occupations. O*NET 4.0 has a web-based accessing application called O*NET OnLine™, which is available to the public at www.onetcenter.org (the O*NET portal page) or directly at online.onetcenter.org. The O*NET 4.0 database has been restructured to incorporate improvements made to the O*NET data collection instruments and is the structure currently being offered to developers.

Data in the O*NET database include the mean ratings on each of the items (or descriptors) in the O*NET questionnaires. Ratings have been standardized to facilitate interpretation and comparison. Examples of specific data in the O*NET database include mean rating data on level and importance for various questionnaire items and text information on occupational definitions, descriptor definitions, scale anchors, and task descriptions.

The Foundation for the O*NET Program

The DOT, first published by DOL in 1939, provided the occupational classification and descriptions needed for operation of the public Employment Service. The collection of information for the DOT was done primarily through observation and interview techniques conducted by trained occupational analysts. These techniques were applied to a small number of job incumbents from what can best be characterized as a "convenience" sample. The DOT has not been maintained or updated since 1991; many of its observations were conducted during the 1970s.

In the late 1980s, the ETA of the Department of Labor began a review of the DOT program. The purpose was to address concerns about the cost and difficulty of maintaining the DOT using the existing methods and the need for additional and more current information. In 1990, as part of the DOT review, the Secretary of Labor appointed the Advisory Panel on the DOT. In response to its charge to advise on improvements to the DOT, the Panel called for development of a new database of occupational information, a concept that has evolved into the O*NET database. An important theme throughout the Panel's recommendations was the development of a common language about jobs, occupations, and skills. The Panel's complete report is presented in *Appendix C.*

A.1.2 The O*NET Data Collection Approach

The O*NET *Data Collection Program* is a critical step in the full updating of the O*NET database to reflect the most current occupational skills and attributes. In the research leading to the O*NET *Data Collection Program*, various sources and methods for collecting occupational

information were examined, including collection of data from job incumbents and supervisors, and development of ratings by SMEs and occupational analysts. Based on this work, the O*NET team has determined that the preferred source of data for most domains (Skills, Generalized Work Activities, Work Context, and Knowledge, Education and Training, and Work Styles) is job incumbents, while analysts are preferred for the Abilities domain, which tends to be more abstract. In addition, other SMEs, such as supervisors and trainers, may be used where access to job incumbents proves extremely difficult.

Previous studies that compare various sources of job analysis ratings suggest that incumbents "...seemed best able to provide information across all descriptor domains" (Peterson, Owens-Kurtz, Hoffman, Arabian, & Whetzel, 1990; Fleishman & Mumford, 1988). "[In addition], large samples of knowledgeable job incumbents are available, which should contribute to the reliability of the resulting descriptive system" (Peterson, Mumford, Levin, Green, & Waksberg, 1999). Furthermore, the world of work is constantly changing and technological advancements are occurring so rapidly that an efficient and effective way to remain current and accurate is to obtain the information directly from those performing the work.

Workers selected to participate in the O*NET *Data Collection Program* as part of a random sample are provided with questionnaires and asked to rate the requirements of their own jobs as defined by the questionnaire items. The responses are tabulated into statistics, such as mean ratings for each scale.

The Advisory Panel on the DOT also recommended using sampling techniques to ensure the representativeness and accuracy of the occupational data. By contrast, no systematic sampling methods were used in the development of the original DOT.

Collecting representative information from job incumbents presents many challenges; among them is determining the best method for identifying a representative sample of job incumbents in each occupation. In particular, the O*NET Program is concerned with identifying sampling approaches that minimize burden on employers and the public, achieve broad coverage of the workers in each occupation, ensure acceptable response rates, and are cost-effective.

Three types of sampling frames are available for identifying samples of workers in each occupation: lists of individual workers identified through professional and trade associations, licensing agencies, and unions; households; and employers. Each option was evaluated in light of the criteria mentioned above, and each offered advantages and disadvantages.

Identifying sampling frames of workers through professional and trade associations and unions is a good approach for some occupations. One advantage of this method is lower

response burden, as contacts with a sample of employers are replaced with contact with one or a few associations. This procedure also removes the cost of soliciting and maintaining employer cooperation but adds the cost of soliciting and maintaining association cooperation.

However, a major disadvantage of the Association List Sample approach is coverage. This method can be used only where association membership encompasses a significant share of employment in the occupation, which may be difficult to identify accurately. The membership of many associations consists of people in multiple occupations, retired individuals, and other interested parties. Further, few associations keep occupation information on their membership. Where coverage of employment in specific occupations by association membership is significant but not broad enough, the Association List and the General Employer Sample approaches to incumbent data collection may be used in a dual-frame sample design, as described in *Section B.1.6*.

The employer survey approach provides the advantage of lower response burden and cost when compared with a household survey, as well as good coverage for the large majority of occupations. Response burden and costs are lower for two primary reasons. First, there are more workers per employer than per household. Thus, the number of contacts required to identify workers is fewer. Second, employer contacts can be minimized by focusing on the employers most likely to employ workers in each occupation for which the sample is needed. Efficient sample design is possible because the distribution of employment in an occupation is usually a function of the industry of the employer. Employment by occupation by industry is measured by the federal-state OES program national estimates provided by BLS.

The employer survey provides good coverage of wage and salary employment as long as an acceptable employer sampling frame is available. Coverage of self-employment is more difficult, although the sampling frame planned for use in O*NET surveys includes many establishments operated by self-employed workers. Coverage of unpaid family workers in an employer sampling frame is minimal or zero, but the number of workers involved is quite small (135,000 nationwide in 1999 [BLS, January 2000]).

The employer/job incumbent survey approach was selected as the primary method to update the O*NET database and was pretested in 1999/2000. The O*NET Program is currently collecting the first round of data using this approach. Achieving high response rates with the employer survey approach can be problematic, as this approach requires cooperation at two levels—first by the employer and then by the sampled worker. However, based on the results of the pretest and the first round of data collection, this method has proven successful. Even though the response rates are acceptable, we will continue to take steps to enhance the response rates. Current and future efforts to improve response rates are discussed in *Section B.3*.

The General Employer Sample approach provides the best method for most occupations, and the Association List Sample should be used for a small number of occupations, either in a dual frame or alone, depending on the coverage provided by the Association List Sample. The dual-frame approach may be used to supplement the general employer sample approach where the latter may not be sufficient to obtain the necessary number of observations. With both approaches, continuous improvement in survey design and methods is planned to increase response rates.

A.1.3 Overview of the O*NET *Data Collection Program*

The O*NET *Data Collection Program* establishes the foundation for continuous data collection to update and maintain the O*NET database using collection methods designed to obtain high-quality and current data. The primary method to update the O*NET database is a General Employer Sample, that is, a survey of establishments and workers within those establishments. It employs a two-stage, establishment-based survey design, including a statistical sample of establishments expected to employ workers in the specific occupations being surveyed and a sample of workers in the occupations within the sampled establishments. These sampled workers are asked to complete the survey instruments.

In the General Employer Sample data are collected from sampled workers via four survey questionnaires: Skills, Knowledge (which includes Work Styles and Education and Training), Generalized Work Activities, and Work Context. These questionnaires are used with sampled workers identified using the two-stage sample design. Each sampled worker is randomly assigned only one of the four questionnaires. The workers are also asked to provide basic demographic information and to complete a brief task inventory for their specific occupation. In addition, they are offered the option of completing the questionnaire online. Data for a fifth domain, Abilities, are planned to be provided by trained analysts. ***Exhibit A-2*** in ***Section A.1.2*** lists the questionnaires, the number of items and scales, and data sources.

For selected occupations—such as ones with small employment size, ones in which employees work in remote locations, and ones for which no employment data exist for use in the sample design—the O*NET Program plans to employ two other methods, either in addition to, or in place of, the primary method. The first of these methods involves sampling from member rosters of professional associations that include a significant portion of the occupation's workers in their membership. The second method identifies and uses SMEs to complete the survey instruments. Each SME is asked to complete the above-mentioned four domain questionnaires, the demographic items, and the task inventory for the specific occupation being surveyed. The Association List Sample and SME method have not yet been implemented, as the focus during the first data collection wave has been to refine and improve the primary method, the General

Employer Sample, and to gather more data to identify occupations for which the secondary methods are more suitable. More details regarding the use of associations and SMEs are provided in *Sections B.2.11* and *B.2.12*.

The O*NET *Data Collection Program* has been designed to reduce the burden on establishments and employees in the following ways:

- Each incumbent responds to only one of the four questionnaires.

- The number and complexity of the questionnaires have been kept to a minimum. Each questionnaire requires only 25 to 30 minutes to complete.

- Informational materials sent to establishments and employees have been streamlined.

- Procedures for sampling employees within an establishment have been simplified and automated.

- Use of the web-based questionnaire will be encouraged by prominent identification of this option in survey materials.

A.1.4 Summary of the Data Collection Procedure

The O*NET *Data Collection Program*'s primary method of collecting data, the General Employer Sample, will involve the following broad steps, which are discussed in more detail in later sections.

1. **Sample Design and Selection**. The first stage of the sample, the sample of establishments, is based on the most current OES data. The OES survey data provide a distribution of employment in the occupation by Standard Industrial Classification (SIC) for the nation. The O*NET sample of establishments for an occupation is drawn roughly in proportion to each industry's share of employment in the occupation. At the later stages of selection, the sample of workers within the sampled establishments is selected randomly from among all workers in the establishment who work in the occupation, with the assistance of a point of contact (POC) in the establishment. Each sampled worker is assigned randomly one of four survey instruments. The sample design is discussed in detail in *Section B.1*.

2. **Solicitation and Data Collection**. The survey staff contacts the sampled establishments to solicit their participation and to identify the specific POC to assist in the survey. This individual, usually a human resources manager, works with the Business Liaison (BL) to determine if any of the POC's employees fit any of the occupations targeted for that establishment. The BL and POC use the occupation's description and relevant tasks to make those determinations. If the POC confirms that he or she has employees in the occupations of interest, the POC assists in drawing the sample of workers for the occupation or occupations. Survey materials

are mailed either to the POC for distribution or, if the POC prefers, directly to the sampled workers at their work address. Each sampled worker is asked to complete the survey during non-work time, using either the pencil/paper or the Internet version and, for pencil/paper users, to return the response in the postage-paid envelope directly to the survey processing center. Survey operations are conducted using an automated control system. The POC is provided with some nominal gifts as an incentive to participate in the employee sampling process and a cash incentive for participating in employee data collection. Monetary incentives are given to all sampled workers. ***Section B.2*** describes the survey procedures. ***Section A.3*** describes the automated tools to be used (control system and Internet). Incentives for respondents are described in ***Section A.9***.

3.   **Data Analysis and Publication**. The data are "cleaned" using standard edit procedures and through the identification and evaluation of anomalous cases. The data are summarized by computing the mean and standard error of the ratings for each descriptor, by occupation. Detailed analysis of nonresponse is then conducted. In addition, a variety of analyses are planned, focusing on reliability and validity of the ratings, demographic characteristics of the respondents, and the factor structure of the domains. The data will be released annually for public use through regularly scheduled updates to the database. Technical information, such as sampling error measures, will also be made available to users in paper and electronic reports. Data analysis and publication are discussed more fully in ***Section A.16.***

4.   **Schedule**. The major components of the O*NET *Data Collection Program* include sampling, data collection operations, and analysis. ***Section A.16*** provides the expected schedule. The schedule assumes OMB clearance is received on or before September 30, 2002. A change from that assumption would require a recalculation of the data collection schedule. Data collection is carried out using overlapping collection waves, with each wave designed to cover approximately 50 primary and 50 secondary O*NET occupations (see ***Section B.1.2)***. Projected waves of data collection represent a continuation of the current data collection activities. Each wave is divided into sub-waves to allow for more efficient sampling with less burden and more completed occupations in each wave.

A.1.5 Summary of the First Wave Experience to Date

Wave 1.1 data collection began in June 2001 and was completed in May 2002. In Wave 1.1, 12,274 establishments were selected for the sample. This first wave of data collection was designed to obtain responses from employees in 201 O*NET occupations of interest. The sample of establishments was selected with probability proportional to size where the size measure for an establishment is roughly proportional to its expected number of employees in the 201 occupations. This selection method gave establishments with a greater expected number of employees in these occupations a greater chance of being selected for Wave 1.1.

During the first phase of the data collection process, the 12,274 establishments were contacted by an O*NET interviewer (referred to as a Business Liaison, or BL) who explained the O*NET data collection effort and attempted to elicit participation from a suitable POC at the establishment. End of survey results suggested 15% of the establishments contacted were ineligible for the study because they were out of business, a frame duplicate, or otherwise out of scope for the survey for some other reason (e.g., industry change). Among the eligible establishments, 64% agreed to participate in the data collection effort.

After an establishment agreed to participate in the study, it was sent relevant information about the occupations of interest. Subsequent calls were made to the establishment to gather data on its total number of employees in up to 10 occupations of interest. These 10 occupations were randomly chosen for each establishment from the 201 target occupations, while giving greater probability to the occupations with the most employees within the establishment. The employment information obtained from the POC on the 10 occupations was used to select up to 15 employees within each establishment. In Wave 1.1, a total of 19,623 employees were selected and sent an incentive and a questionnaire (usually through the POC, to preserve confidentiality). At the end of data collection, 63% of the selected employees had responded. Additional information on the data collection methodology is provided in ***Section B.2***.

It is difficult to make response rate comparisons between the O*NET *Data Collection Program* and other establishment surveys due to some unique design characteristics. First, participation in the O*NET survey is completely voluntary rather than mandatory. Experiments conducted by the Bureau of the Census suggest that simply informing respondents that their response is mandatory adds about 20 percentage points to overall response (Worden and Hoy, 1992; Tulp, Hoy, Kusch, & Cole, 1991). Thus, since it is a voluntary survey, we would expect response rates to be about 20 percentage points lower than the average federally mandated survey.

Another unique and important feature of the O*NET design is the fact that the survey organization conducting the data collection does not have direct, personal contact with the ultimate respondent for the survey, the sampled employees. Rather, the requirement of respondent anonymity means that participation at the employee level relies exclusively on the interactions between the POC and the employee. In their review of establishment mail survey response rates, Paxson, Dillman, and Tarnai (1995) found that establishment surveys that featured anonymous mailings report response rates that were more than 30 percentage points lower than surveys having direct, personal contact with the respondents.

Another unique characteristic of the O*NET survey is that it requires participation at two stages of response—the POC level and the employee level—whereas the typical establishment

survey requires participation at only one level, the establishment level. Because very few surveys incorporate such a design, survey methods literature is essentially devoid of examples upon which to base a reasonable response rate expectation for the O*NET *Data Collection Program*. Therefore, the comparison of O*NET response rates with other establishment surveys is done separately for each stage of participation. First, we compare the O*NET establishment-level response rate with other mail establishment surveys having only one response stage at the establishment level. Then we compare the O*NET employee-level response rate with the response rate of other establishments' self-conducted employee surveys.

Paxson et al. (1995) analyzed the response rates for 46 surveys conducted by both government and nongovernment organizations. Among the surveys in their study, 26 were conducted by the Social and Economic Sciences Research Center (SESRC) at Washington State University and 20 were conducted by the U.S. Bureau of the Census. The SESRC, directed by Dr. Donald Dillman, is well known for its development of the Total Design Method (TDM) approach to mail surveys (Dillman, 1978; 2000) and its high response rates in implementing that methodology. Further, the 20 Census Bureau surveys include 12 well-established and ongoing, mandatory surveys. The average response rate for all 46 surveys is 63%, but if only voluntary surveys are considered, the average response rate drops to 55%. These results suggest that the O*NET establishment-level response rate of 64% exceeds expectations for this type of survey.

The second response stage of the O*NET survey can be compared to establishment surveys in which the ultimate sampling units are the employees of the establishment. Since federally sponsored surveys of employees within organizations are quire rare, the literature on their response rates is sparse. Most surveys of this type are employee satisfaction surveys. For example, one well-documented government-sponsored survey of employees is the Public Service Employee Survey, administered to more than 190,000 employees of the federal Public Service of Canada and conducted by Statistics Canada in 1999. A questionnaire was delivered to each employee by a government agent who personally requested that the employee complete the questionnaire and return it by mail. Multiple follow-ups of nonrespondents were made by e-mail and interoffice mail to maximize the response rate. No incentive was used; however, because all sample members were also employees of the organization conducting the survey and could fill out the survey on government time, the use of an incentive was thought to be unnecessary. The final overall response rate for the survey was 55%.

The Office of Personnel Management (OPM) developed the Organizational Assessment Survey (OAS) and has encouraged all federal agencies to survey their employees in order to evaluate organizational performance, benchmark best practices, and align performance with important and measurable outcomes. The experience OPM has had in implementing these surveys within numerous federal agencies provides some evidence of response rates for

employee surveys conducted by the U.S. federal government. The OAS design is very similar to the Canadian Public Service Employee Survey design. The surveys are self-administered and are conducted by each agency for its own employees. Furthermore, like the Public Service Employee Survey, the OAS request to participate is personalized and made directly to the employee by his or her employer. The features of the design offer a significant advantage over the O*NET survey design, as previously noted.

Although the results of the OAS surveys are not publicly available, an official at OPM was able to provide some general information regarding OAS response rates (C. Simons, personal communication, March 21, 2002). According to OPM, response rates for OAS surveys vary considerably by agency, from 30% to 80%. However, the average response across all agencies is approximately 57%. This result is similar to the Canadian experience. It is further supported by Roth and BeVier (1998), who conducted a meta-analysis of 173 surveys in the field of Human Resource Management and Organization Behavior (HRM/OB). They found that surveys implementing many of the response-rate-enhancing features of the O*NET survey had response rates in the range of 23% to 78%, with a median of 51%. These results suggest the O*NET employee response rate of 63% also exceeds expectations for mail surveys of employees within their organizations.

In summary, results from the survey methods literature and from other federal surveys in the U.S. and in Canada suggest that the current 64% establishment response rate and the 63% employee response rate exceed expectations for federal surveys that use a similar data collection approach. However, methods for further improving response rates will continue to be explored, and response rates are fully expected to continue to increase as the O*NET *Data Collection Program* progresses. **Section B.3** summarizes the enhancements that will be implemented in future waves to maximize response rates and further reduce the impact of potential nonresponse bias in estimates produced from the O*NET data collection effort.

A.1.6 Statutory and Regulatory Information

Section 309 of the *Workforce Investment Act* requires the Secretary of Labor to oversee the "development, maintenance, and continuous improvement of a nationwide employment statistics system," which shall include, among other components, "skill trends by occupation and industry." The states are to develop similar statewide employment statistics systems.

The Secretary of Labor's *Workforce Information System Plan for FY 2001-2005*, released in October 2001, includes as one of its priorities the collection of occupational skills data, stating: "During FY 2001, ETA initiated the data collection program for the Occupational Information Network—O*NET—to update the database and refresh it on a regular basis. ETA

also will continue research and development on O*NET measurement concepts and data collection methods" (U.S. Department of Labor [USDOL], October 2001). This citation in the plan demonstrates that the O*NET Program is the primary response vehicle for collecting skills information across all occupations. Updating the entire O*NET database is a critical component of the nationwide labor market information system to support employer, workforce, and education information needs.

In addition, the *Plan* establishes the improvement of occupational information products as one of its major goals, noting: "The quality and availability of information about occupations will be maintained and improved by enhanced occupational and career information products, short-term employment forecasts, and skills-oriented information using O*NET and other skills research and databases." Providing the most current information on the O*NET skills—the purpose of the proposed data collection effort—is key to supporting the needs of customers through the wide variety of public and private sector products that depend in part on the availability of O*NET information.

The *Workforce Investment Act* contains numerous references to identifying the skill requirements of jobs; for example, Section 154 requires local determinations of the "skills and education that are necessary to obtain the employment opportunities" in the local area, and Section 134 requires that "information on job skills necessary" be provided to obtain jobs listed for the local area.

The *Workforce Investment Act* also contains references to identifying and assessing the skills of individuals; for example, Section 134 allows the provision of assessment services to identify "the skill levels and service needs of adults and dislocated workers." Section 129 allows funds to be used to provide youth with an assessment "which shall include a review of basic skills, occupational skills, prior work experience, employability, interests, aptitudes (including interests and aptitudes for nontraditional jobs), supportive service needs, and development needs of such participant…." In Section 136, the "attainment of basic skills and, as appropriate, work readiness or occupational skills" may be included as performance indicators for youth programs.

The O*NET Career Exploration Tools™, including the O*NET Interest Profiler™ and O*NET Work Importance Locator™, are such assessment tools designed specifically to relate a person's interests and work values to the information on education and skill requirements for occupations that is provided in the O*NET database.

Section 508 of the *Workforce Investment Act* and Section 118 of the *Perkins Act* specify that persons with disabilities should have access to and use of the same information that persons

without disabilities have. The O*NET *Data Collection Program* complies with this section by providing both low-vision and text-only versions of its online application that can be used to access O*NET data and information. O*NET OnLine also provides links to several accommodation and disability resources on the Internet. Furthermore, the O*NET Consortium website, the Department of Labor's ETA O*NET website, and O*NET OnLine have all been "Bobby Approved," indicating that the Center for Applied Special Technology has deemed them accessible to persons with disabilities.

While the O*NET name is not referenced specifically in statute or regulation, O*NET information is the foundation for carrying out the responsibilities related to these sections cited above. The predecessor to the O*NET database, the DOT, is cited either directly or indirectly in a number of places as a source of occupational information in support of federal programs. The DOT is cited in the *Code of Federal Regulations* (CFR) in 30 places. These citations are generally related to determining disability, administering Department of Labor employment and training programs, or administering immigration, civil rights, and labor standards law. Department of Labor officials responsible for the O*NET Program are working with DOT users in the federal government, such as State Department officials responsible for visas, the Bureau of Apprenticeship Training, Foreign Labor Certification, and others. This work is focused on assisting these entities to make the transition from use of the DOT to use of O*NET information.

Copies of the sections of the *Workforce Investment Act*, the *Perkins Act*, and the specific CFR citations are presented in ***Appendix D***.

A.1.7 Federal Register Notice

The Federal Register notice asking for public comment prior to submission of this request to OMB is presented in ***Appendix E.***

A.2 Uses/Products and Services Based on the O*NET Program

The O*NET Program succeeds the DOT and is a powerful tool for various critical federal and state workforce development functions. The O*NET Program integrates a powerful relational database and a common language for occupational and skill descriptions into a value-added tool for businesses, job seekers, and the workforce development professionals who help bring them together. By providing information organized according to the O*NET Content Model, the O*NET database is an important tool for keeping up with today's rapidly changing world of work. The O*NET database provides:

- detailed information for more than 900 occupations.

- information on standardized descriptors of skills, abilities, interests, knowledge, work values, education, training, work context, and work activities.

- occupational coding based on the 2000 SOC.

Furthermore, O*NET OnLine, a web-based application, provides Internet access to the O*NET database and offers:

- easy-to-use interface and search menus.

- skill-searchable occupational descriptions for more exact career targeting.

- links to accommodation information to use as a counseling tool for workers with disabilities.

- links to labor market information and occupational certification requirements.

This section provides an overview and many examples of how O*NET information is used. Because the O*NET Program is an electronic system of databases and other products intended to be used by both public and private developers to build products and resources to serve customers, it is important to understand the significance of O*NET data as the underpinning for hundreds of products in the marketplace serving millions of customers (J.Wall, personal communication, June 20, 2002).

O*NET OnLine (online.onetcenter.org) is currently averaging 55,000 visits per month. The O*NET portal site (www.onetcenter.org) is averaging 16,000 visits per month. In addition, there are three other major indicators of the extent of use of O*NET information in both the public and private sectors. One is the number of individuals and firms who have submitted user certification information indicating their intent to use the O*NET database. A second major indicator is the number of websites linked to various O*NET sites available on the Internet. A third indicator is the number of users of systems that incorporate O*NET data. Each of these is discussed briefly below.

User Certifications Submitted

Since November 1999, more than 500 firms and individuals have submitted certification information indicating their intent to use O*NET data in other products. This certification form is voluntary, so the actual number is likely higher. Users were not asked to indicate their intended application of O*NET data; however, a review of job titles and company names indicates that there are 11 main categories of users (see *Exhibit A-3*). The approximate percentage share of registered users is listed by the inferred type.

While specific information is not available on how each user is using the O*NET data, many are using it to build applications. For example, career information systems incorporate O*NET data and reach millions of customers during the year. A goal of the O*NET Program has been to encourage private and public developers to build applications and deliver the information to the public, and the O*NET Program is clearly meeting that goal.

Exhibit A-3. Main Categories of O*NET Users

Inferred Type of O*NET User Submitting Certification	Approximate Share
Government Agencies	17%
Information Technology (IT) Developers	10%
Career Information System (CIS) Developers	7%
Vocational Rehabilitation Agencies, System Developers, or Individual Professionals	7%
Higher Education Institutions	7%
Other Educational Institutions	3%
Assessment Instrument Developers/Publishers	3%
International Researchers or Governments	3%
Human Resource Specialists or Consultants	2%
Law Firms	1%
Health Care/Social Service Facilities	1%
Miscellaneous Private Companies or Individuals	39%

Internet Website Linkages

Based on an informal exploratory search conducted in January 2002:

- More than 1500 sites link to O*NET OnLine (online.onetcenter.org).

- Hundreds of sites link to the O*NET portal page (www.onetcenter.org).

- Nearly 1000 sites link to the O*NET government site (www.doleta.gov/programs/onet).

The most common types are links to various O*NET websites from:

- libraries and career centers based in higher education;

- higher education institutions' schools of business, labor and industrial relations, psychology, education, and counseling;

- government agencies (primarily state labor-related agencies; some federal and local);

- public libraries (especially those offering career and job search assistance programs);

- career counselors, coaches, and recruiters (mostly private sector vendors);

- career exploration or job search assistance (both private and public sector);

- public school systems, educational associations, and secondary schools (often recommending the O*NET database as a resource for faculty, parents, and high school juniors and seniors);

- human resources management;

- lawyers specializing in immigration law;

- vocational rehabilitation/occupational medicine and health; and

- international sites in Bosnia, Turkey, Canada, the United Kingdom, and Crete.

The O*NET Program in Published Literature

The O*NET Program has influenced occupational research, as the following indicate:

- Scientific articles have been published on O*NET data, indicating that O*NET data have influenced job analysis theory and research (Campion, Morgeson, & Mayfield, 1999; Peterson et al., 2001).

- In the textbook *The Changing Nature of Work: Implications for Occupational Analysis* (National Research Council, 1999), the O*NET Program is mentioned extensively as an example of occupational analysis designed to meet the challenges of a changing world of work. Furthermore, O*NET information is built into the curriculum of some courses in I/O psychology, job analysis, human resources, and career counseling.

- O*NET data are being used in occupational research by state and private organizations. For example, a RAND study, *Characterizing the Future Defense Workforce* (Levy et al., 2001), drew heavily on O*NET data.

Specific Examples Described

Description of some specific examples of how the O*NET Program directly supports a variety of workforce development services and products are listed below:

- Alabama uses O*NET information in its web-based ACLMIS system. This system helps students and job seekers to explore a variety of occupations. O*NET information enables them to research occupations of interest and analyze the skills required, with the goal of finding employment in a field that fits their interests and abilities.

- The Boys & Girls Clubs of America developed a career preparation program to help teens make sound educational decisions, explore a variety of careers, and develop skills for success in the workplace. A system called CareerLaunch™ supports this program. This expanded career preparation website incorporates O*NET data enabling teens to find O*NET career profiles related to their interests.

- In California, O*NET information was used to help 400 displaced workers at a Boeing manufacturing plant identify their existing skills and use that information to search for other occupations requiring the same or similar skills.

- Connecticut uses O*NET data to assess current and future job opportunities in the state and the skills needed to obtain them. Using O*NET data, Connecticut identified skills, as well as the other qualities and characteristics, needed for success in each of 25 growth occupations. Using the O*NET skills database, researchers have identified 12 skills required by establishments that are common among some 25 demand occupations. Their findings are drawing interest from local community colleges, considering the development of targeted curriculum to meet this need.

- Minnesota used O*NET data as part of a statewide survey to identify new and evolving occupations and skills.

- New York's CareerZone® is an interactive career exploration resource designed for students. It incorporates information from the O*NET system and is available on the Internet.

- A One-Stop Career Center in Oklahoma is using O*NET data to help small establishments develop job descriptions. For one such employer, Center staff developed a series of job descriptions for skilled welders. These job descriptions were used as a basis for an in-house certified training program for welders.

- Nebraska uses O*NET data in the Rapid Response program at plants that are closing or facing large reductions in force. Workers first use information from the O*NET database to identify skills they have developed on the job. This information is used to find related occupations that might offer future employment.

- Texas used O*NET data to describe the requirements for 54 emerging or evolving occupations in eight industries that will impact the Texas economy and create a need for modified training programs. The O*NET database was essential to this project because it provided occupational descriptions and definitions for the knowledge, skills, and abilities necessary for these jobs.

- A number of career information delivery systems, including Choices, intoCareers (formerly the national Career Information System [CIS]), and ACT's Discover, have incorporated O*NET information in occupational reports and descriptions and in search functions.

- In New Jersey, a unique process-reengineering project brought workforce development professionals together to recommend process improvements in One-Stop service delivery to build better value for customers. This panel recommended the introduction of O*NET assessment tools in the state's service delivery process, including employer outreach programs, to improve job order accuracy and career assessment initiatives.

- Hundreds of workers in state and local One-Stop partner agencies have been trained to deploy O*NET assessment instruments for use in career counseling and training placement evaluations.

Number of Users of Systems that Employ the O*NET Database

Because the O*NET database is becoming the intelligent core of any number of technology tools serving the workforce development system and employers today, it is impossible to accurately measure the number of users of the O*NET database. However, it is clear that systems with "O*NET inside" currently serve millions of users. For example, based on past information, it is likely that at least 8 to 10 million users access career information delivery systems each year. Each of these systems uses O*NET data to present skills and characteristics information and to connect to labor market information. One system, Choices, has even computerized and integrated the O*NET assessment tools into its system, allowing users to take the assessments that then link to O*NET occupations. As shown by the services and products listed above, the O*NET database is used by students, dislocated workers, workforce development specialists, labor market specialists, career counselors, software developers, business forecasters, human resource professionals in business, and a host of other users. The O*NET database is a key resource in addressing the needs for many different audiences. Several private assessment and test developers have linked their instruments to the O*NET Program. The Program's common language serves as the means of linking various workforce development services into an integrated system for business customers, job seekers, educators, and students alike.

A.2.1 The O*NET Program and the Education and Training Needs of the American Workforce

The education and training system in the U.S. plays a significant role in preparing for the future skill needs of the American workforce. Identifying common or transferable skills needed across occupations will become increasingly important as we adapt to changes brought on by a

global, information-driven economy. Accomplishing this goal requires the ability to document the skill requirements of current and future occupations in the economy and to design curricula responsive to these needs to prepare students and workers for future work opportunities. O*NET information is used to help identify emerging and declining skills of occupations in the economy and to identify program curricula to meet these needs.

The Department of Education Office of Vocational and Adult Education (OVAE) is developing a set of Career Clusters through the work of committees of local educators, employers, and other experts. These Clusters are designed to facilitate school-to-work connections and transitions by demonstrating to students, parents, educators, and counselors the career paths related to student learning, and to help them plan further educational attainment. As part of this project, each Cluster Committee is linking its Career Cluster to both relevant O*NET occupations and relevant programs from the educational program taxonomy, the Classification of Instructional Programs (CIP). With this information, educators and school counselors can help students identify specific occupations related to skills and knowledge acquired through education. The Cluster Committees also are using O*NET information as a resource in identifying foundation skills and knowledge for the cluster and specialty career pathways within it. The committees then work to expand and validate the required skills and knowledge to help develop curriculum and tools to assess the identified skills.

The O*NET Program assists training program operators, vocational and technical education program administrators, and planners in identifying skills and knowledges their customers need to meet changes in the world of work. With this knowledge, they can then respond to customer needs and offer training programs that will be relevant to both program participants and employers. Billions of federal, state, and local government and private employer dollars are used to support occupational skill-specific training for millions of people, and the O*NET Program supports more effective use of these funds.

Perhaps most importantly, the O*NET data collection process provides a methodology for aligning data with changing conditions, providing a dynamic tool for continued alignment between employers and the educational community. Through a continuing focus on employer needs, ongoing data collection offers the prospect of continually refreshed insights into future hiring needs.

A.2.2 The O*NET Database for Career Exploration and Career Development

Career planning is a multiphase process that individuals undertake throughout their lives. It begins in childhood during the elementary school years and lasts until a person retires from the workforce. During this process, individuals internally synthesize information about themselves

and information about the world of work. O*NET information can assist individuals in this process. The O*NET database includes the following integral types of occupational and worker information for career decision making:

- Occupational definitions
- Tasks
- Abilities
- Work activities
- Knowledge
- Education and training
- Skills
- Work styles
- Work context

The O*NET database serves not only as a direct resource of important career information, but also as a source for the value-added products of commercial and noncommercial career information delivery systems and assessment developers. Career information delivery systems provide occupational and educational information for individuals in the process of career exploration and decision-making. Numerous commercial aptitude and interest inventories and instruments used in career decision making also rely on O*NET information. O*NET data serve as a resource base for tools that utilize occupational information. These systems are used by millions of individuals each year. The many products and systems that utilize O*NET information and tools serve many purposes:

- Young people exploring careers can be exposed to a wide variety of career opportunities using in-depth O*NET occupational information. It can help widen their options and help break stereotypical notions that may inappropriately limit young people.

- School counselors use O*NET information to develop career exploration activities for their students. For example, students use the O*NET Program to match their interests and values to specific occupations.

- Secondary students begin to discover more information about themselves that is crucial to their future career paths. By using the O*NET career information and assessment instruments related to O*NET data (such as the O*NET Career Exploration Tools), these students can begin to determine how their abilities, strengths, and interests match up with different occupations in the world of work. They can begin to identify the education and training requirements of careers that they might want to pursue.

- Teachers can use O*NET information in developing academic curricula that will help students meet the needs of employers and attain skill levels appropriate for future jobs.

- College students and curriculum/program planners can use O*NET information to keep abreast of the new skills and training required to pursue today's highly demanding jobs.

- Adults in career transition can use O*NET information to help identify occupations that require similar skills to those they currently possess and identify the training and education they will need to qualify for new careers.

- Individuals who become disabled during their adult years also need information to make potential career transitions. People with disabilities can use the O*NET database to assist them in identifying their skills and abilities. The accurate portrayal of skill requirements of occupations in the O*NET Program supports accurate self-appraisal but does not erect artificial or inappropriate barriers to employment for the disabled.

- Many older Americans continue to work for pleasure or to augment fixed incomes after retiring from their primary job. This group also benefits from the information collected by the O*NET Program.

How people make decisions about the work they will pursue and the education and training they seek is affected by the information available to make career decisions. The O*NET Program supplies the critical information that forms the foundation of these career decision-making processes.

A.2.3 The O*NET Database as a Tool of the Workforce Development System

Human resources, education, business, government, and training professionals use O*NET information to respond to a shifting labor market and rapidly evolving economy. The O*NET Consortium has launched a nationwide training effort to inform and train professionals from these communities about the value and uses of O*NET information and how to access O*NET resources. Since January 2001, 34 states have completed *Discover O*NET: The Language of Occupational Intelligence* train-the-trainer program. More than 600 Workforce Development Professionals have been certified as O*NET Trainers to deliver this training to colleagues within their state. As of February 2002, four additional states have scheduled *Discover O*NET* training sessions.

The regionally delivered training session, *O*NET Quick Start: How to Integrate O*NET in Local Workforce Development Initiatives*, reached 181 system integration staff from 31 states. *The Value of O*NET: Making WIA Work*, was a 1 ½-hour session delivered to 1,913 executive and managerial staff from 45 states.

Most recently, O*NET training support has been migrated to the Internet to make the effort more efficient and expand the serviceable student population to include the business customer and educational community. On O*NET Academy, students gain access to self-paced training on how to use the O*NET system and best practices for its application on the job. Live, collaborative tools, such as web-based classroom training sessions, forums, and newsletters, provide a forum for users and developers to share O*NET knowledge and gain new insights on user needs and customer satisfaction. As one of DOL's first initiatives for e-learning, O*NET Academy is already quite effective, with 600 registered users.

Using the powerful O*NET relational database, users have already created effective, timesaving systems for accessing the latest available labor market information and meeting workforce demands. Some O*NET success stories follow.

Emerging Skill Needs

An analysis of the multimedia and entertainment industries in California revealed a pressing need for a pool of talented and skilled 3-D computer artists and traditional animators. Using O*NET's survey data collection instruments and other material, California's Employment Development Department conducted an industry study, with the endorsement of the California Skillsnet Consortium, comprising establishments and educators. As a result of O*NET data, human resources personnel modified local training and education initiatives to help fill the gap between industry needs and skills possessed by local workers. Training programs defined the skills and requirements for the evolving occupations of 3-D computer artists and traditional animators. State and local leaders forged important partnerships with local establishments, educators, Workforce Investment Boards, community-based organizations, and other stakeholders to meet industry demands.

Employers

When Boeing prepared to close its aircraft manufacturing plant in Monrovia, California, managers and workers teamed with the California Employment Development Department to identify workers' skills and align those skills with available jobs. They used O*NET's skills survey tool to identify the workers' skills, and O*NET's customized reports to prepare a Skills Match report for workers. As a result, nearly 400 Boeing workers found positions at other Boeing facilities prior to layoff. With O*NET information about their skills, and how those skills transfer to other occupations, workers were able to expand the scope of their job search.

Internet Sites for Job Openings

Many electronic job banks available on the Internet, such as America's Job Bank or private industry sites, have incorporated the O*NET classification and/or O*NET data. Employers posting job openings on such sites, for example, can take advantage of this information to more efficiently post their job listings or to develop better job descriptions/posting information.

State Agencies

In Oklahoma City, One-Stop partner agencies in the Workforce Oklahoma Career Connection Center are using the O*NET Program in a broad effort to identify and meet specific employer needs in a timely, effective, and cost-efficient way. The Center uses an employer needs survey that provides a framework for customizing a service mix to meet specific employer needs. O*NET information is built into this survey to help employers clearly define their employment needs by job title and skill bundles.

Many states are also now working to integrate O*NET data into their operating systems. In particular, the Louisiana Department of Labor has integrated O*NET data into its business and operating systems. The O*NET Program supports services to employers and job seekers in the Louisiana Department of Labor's Career Search System.

Alabama has a special Dislocated Worker module, which offers a skill survey questionnaire based on O*NET data. Users can identify their skills and link them to Alabama data on wages and employment outlook as well as O*NET occupational descriptions and tasks.

Texas used O*NET data to identify the knowledge, skills, and abilities needed to obtain employment in 54 emerging or evolving occupations in eight industries. By using existing O*NET data, Texas improved the efficiency of gathering and presenting occupational data. This resulted in a savings of time and money when identifying and targeting resources to address skill gaps. Employment and training communities are then able to tailor instruction and training to match occupations in demand.

Connecting Education and Work

The Heldrich Center at Rutgers University is teaching O*NET information in its Career Development Facilitator Curriculum, *Working Ahead*. This is an approved/credentialed 120-hour curriculum for teaching career counseling and guidance to front-line staff in workforce development, community organizations, and community colleges. The O*NET Program is a

featured topic in this curriculum. The *Working Ahead* curriculum has been delivered in several organizations in Maryland and California and will be offered in Hawaii. Soon, *Working Ahead* train-the-trainer sessions will be given in five Urban League Centers across the country: Boston, MA; Oakland, CA; Louisville, KY; Columbus, OH; and Austin, TX.

Working in cooperation with a network of state departments of education and state career resource networks, the Heldrich Center is developing a training guide and student manual for training counselors and education staff in using the O*NET database as a career exploration tool by middle, high school, and college-level students.

ETA is working with the Department of Education–funded America's Career Resource Network to update the Improved Career Decision Making (ICDM) training program for counselors to integrate the O*NET Program into the ICDM curriculum. This curriculum has been used to train thousands of counselors each year since 1981 and marks an important effort by Education and Labor to introduce to the O*NET Program to more career counselors.

Several states use the O*NET database to tie their school-to-work efforts with local occupational demands:

- New York uses CareerZone, an interactive, online system for students, youth, and job seekers containing O*NET information. Students can build resumes, generate cover letters, fill out job applications, learn more about entry-level jobs, and explore future career opportunities.

- Minnesota created an innovative system using O*NET data that documents students' skills and work styles and tests these skills-based portfolios with local establishments.

- Texas uses the O*NET database as the foundation for its career exploration CD-ROM software, Occupation and Skill Computer-Assisted Researcher (OSCAR), developing links between O*NET data and skills taught in the classroom and integrating industry skill standards and industry-based job analysis into OSCAR to meet the needs of the business community.

Workforce Investment Act

The Connecticut Department of Labor (CTDOL) used O*NET data to respond to the *Workforce Investment Act* requirements that states assess: (1) current and future job opportunities in the state, (2) the skills necessary to obtain those jobs, and (3) the skills necessary to meet the economic development needs of the state. To meet these requirements, they collected and analyzed labor market information and published an extensive report entitled

Choices Today—A High Performance Workforce Tomorrow. Occupational skills information featured in the report was extracted from the O*NET database.

O*NET data was an essential part of the *Choices Today* report. The section describing skills necessary for Connecticut's high demand occupations depended on O*NET data and Content Model information, especially Skills and Skill Descriptions. Another section of the report also presented O*NET skills needed by clusters of industries that are a focus of state economic development efforts.

Choices Today was specifically designed for those involved in Connecticut's One-Stop employment and training system. Stakeholders in this system include community leaders, policymakers, planners, educators, counselors, service providers, and program operators at state and local levels. By using O*NET data, the CTDOL was able to save time and resources while meeting the planning needs of the state workforce development system.

A.3   Uses of Information Technology (IT)

The O*NET *Data Collection Program* has electronic versions of survey questionnaires available via the Internet to sampled job incumbents. The program employs the latest in IT systems and procedures to enhance the quality of the data, minimize the burden on the responding establishments and job incumbents, and reduce the overall cost of the data collection effort. Many of the benefits of the paper questionnaires were replicated in the electronic questionnaires. Specifically, users are able to start and stop multiple times without losing data. They can return to the partially completed questionnaire at any time during the survey period and continue at the point where they left off. A respondent may also review and edit previous answers if needed. In addition, an on-screen meter keeps respondents informed of their progress through the questionnaire. The programming effort to develop online data collection is a one-time expense for a product that can be used for numerous data collection cycles.

Advances in web technologies and security, and the prevalence of access to web browsers by establishments and employees have made Internet-based data collection both feasible and practical. Growth in Internet use in the past few years has been enormous and is expected to continue. The use of the web questionnaire increased in Wave 1.1, consistent with the increase in Internet use by society. In Wave 1.1, 15.2% of employee respondents used the web questionnaire, a large increase over the 3.7% who responded on the web during the 1999 pretest. The paper questionnaire cover continues to highlight the option of filling out the survey via the Internet.

The paper and web versions of the questionnaires were designed to be optimal for their respective modes of administration. The questionnaire design literature suggests that this is essential to reduce mode effects. That is, if each questionnaire is designed to minimize measurement error in its particular mode of interview, mode effects are also minimized. For example, in the paper version, multiple questions appear on a single page of the questionnaire. However, in the web version, only one question appears per screen (although the respondent can navigate at will through the instrument). The literature on web survey design (see, for example, Fuchs, 2002) suggests that one question per screen for web surveys reduces measurement error and, thus, the effects of administration mode as well. This was the only important change necessary for the web version since both instruments are self-administered. Further, to ensure comparability between the paper and web responses, the formats and wordings of the questions and response categories for the two versions are identical.

An O*NET data collection web application has been developed to support and assist with the O*NET *Data Collection Program* (onet.rti.org). This site is divided into two major sections, the "public" and "restricted access" sections. The restricted access section is further subdivided into two areas, the "online questionnaires area" and the "project management area."

Public Section

The goal of the public section is to support the establishment recruiting process. This section is accessible to the public without restrictions. The public section includes O*NET background information, endorsement letters, Frequently Asked Questions (FAQs), copies of the questionnaires, and links to other O*NET-related websites. The purpose of the public section is to provide establishments and sampled workers with readily accessible information about the data collection effort and uses of the data.

Restricted Access Section

This section contains sensitive information that is only available to certain populations, such as survey respondents, project managers, and data entry staff. This section is controlled by a user ID and password authentication scheme. The web server includes a Secure Socket Layer (SSL) certificate to allow encrypted transmission of all information over the Internet. This is the same technology used by electronic commerce websites to secure credit card numbers. Also, no "cookies" are used. A cookie is a piece of data given to a user's web browser so that the browser will hand it back to the server or website upon subsequent requests.

Online Questionnaire Area

This area contains online versions of the surveys, providing sample members with an alternative to pencil/paper. Only individuals who have been selected to participate in the survey have access to this area. Unique user IDs and passwords are assigned to each job incumbent by a central office computer system at the time they are selected into the sample; their name or other personally identifiable information is not obtained. The ID and password are provided to the sample member with other survey materials. The website confirms the validity of the ID and password and verifies that a completed survey form has not already been received before allowing access to the online questionnaire area.

Once entering this portion of the site, respondents are:

- informed that participation in the survey is voluntary.

- assured their survey responses will remain confidential.

- limited to seeing only the questionnaire they have been asked to complete.

- permitted to stop at any point and continue responding later.

- permitted to skip questions that they choose not to answer.

- permitted to go back and review/change previous responses.

On the last page of the survey, respondents confirm that they have completed the questionnaire, then exit from the questionnaire area of the site and are thanked for their participation. Their user ID is automatically deactivated at this time. Any further attempts to log in will not be possible, as the system recognizes this user as having completed the survey.

The database containing the survey data is not accessible by the Internet, but resides on a server inside the contractor's firewall. The web data collection application transmits data from the respondent's computer into the survey database. Only program staff, operating from inside the firewall, have access to the survey database.

Project Management Area

This area contains data collection management reports and information. Access accounts for this area are created for managers of the O*NET *Data Collection Program*. This portion of the website serves as an Intranet for the O*NET *Data Collection Program*, facilitating communication among data collection staff and DOL staff. Production reports are posted nightly

and include summaries of the progress of establishment recruiting, questionnaire shipment and receipt, and overall data collection status. Additional applications include a document library that allows designated data entry staff to make changes and edits to documents used in the data collection effort. This document library is the central repository for these documents, eliminating ambiguity and confusion about the validity and currency of changing documents. The project management area of the website reduces the cost of the data collection effort.

Data Collection Utilities

During Wave 1.1 of the O*NET *Data Collection Program*, enhancements were made to the Case Management System (CMS) that allowed greater flexibility and effectiveness in the communications between the contractor's staff of BLs and POCs in the sample establishment. BLs are telephone interviewers for the O*NET *Data Collection Program* who initiate and maintain telephone contact with the sampled establishments. The first enhancement was the capability to electronically fax (e-fax), upon request, written correspondence to the establishment POC. This enabled BLs to tailor their approach to best suit the POC's working style, while not adversely affecting the integrity and comparability of the written communications. The second enhancement was the creation of email accounts for the BLs. This supplemented the standard telephone and mail communications established for O*NET data collection. Again, this enabled BLs to tailor their follow-up contacts with an establishment to fit the communication style best suited to their POC. Together, these two enhancements have improved the study's effectiveness and efficiency.

A.4 Efforts to Identify Duplication

The O*NET Program provides comprehensive, reliable, and valid information about a wide range of variables for occupations in the U.S. economy. Multiple stakeholders use the database.

To avoid duplication and save on cost, several portions of the O*NET Content Model are provided from existing data sources. Specifically, as discussed in *Section A.1,* the domain of Occupational Characteristics—including information on industries, job opportunities, and pay— is provided through links to existing labor market information databases. Information about occupational licensing, certifications, or national industrial skill standards, and about related instructional programs, is provided from existing sources, such as the Manufacturing Skills Standards Council and the National Retail Federation.

The exhaustive reviews of existing labor market and occupational information conducted by the Advisory Panel and the Department of Labor's DOT Review staff, and subsequent research, identified no other comprehensive, valid, and reliable sources that could be used for the data items included in the O*NET database.

The development of the O*NET Program has also involved staff and advisors with many years of experience in labor market and occupational information who are familiar with existing data sources. In fact, many existing systems that provide detailed occupational information are actually using information based on O*NET data or the outdated DOT.

The few existing sources with similar measures are too limited to be used in the O*NET database. Some existing sources are valid and reliable—for example information from the Office of Personnel Management and the Department of Defense—but are not comprehensive, as they represent only those jobs in federal civilian employment or the military. There are also some private sources of job analysis information; however, they are based on job analyses conducted on an "as needed" basis rather than on a representative sample of employers and workers. They are therefore limited in their coverage and are not representative of the entire workforce. Furthermore, these analyses are not comparable because they do not use the prescribed O*NET common language to describe occupational requirements; it is not practical to combine them because they include dissimilar components. Finally, private sources are not available to the general public.

A.5   Efforts to Minimize Burden on Small Establishments

All sizes of establishments are represented in the O*NET estimates for most occupations targeted in a data collection wave. As discussed in *Section B.1.3*, the targeting strategy for some occupations may lead us to omit smaller or larger establishments from the sample frame for particular occupations; however, this omission occurs for a very small number of occupations. We omit smaller (or in some cases larger) establishments for some occupations when it is clear that sampling these establishments will greatly reduce the efficiency of data collection or when incumbents from these establishments are outside of the target population. In most cases, to give users of the O*NET data fully informed, unbiased occupation estimates, sampling from establishments of all sizes is preferable. Given that establishments of all sizes are represented on the sample frame for nearly all occupations, we have included specific design provisions to not overly burden small establishments.

Exhibit A-4 shows the distribution of the establishments on the InfoUSA frame of establishments. The O*NET establishment sample will be drawn randomly from this data source

for each data collection period. *Exhibit A-4* also shows the expected distribution of the O*NET sampled establishments. The data displayed in this exhibit reflect the sample of establishments worked in Wave 1.1 of the data collection effort. In future data collection waves, we plan to under-sample the smaller establishments at approximately the same rate.

Exhibit A-4.   Distribution of Frame and Sample Establishments by Employment Size

Number of Employees	Total Frame Establishments[a]	Frame Distribution	Expected Distribution of O*NET Sampled Establishments[b]
1–4	6,654,415	60.8%	18.9%
5–9	2,056,113	18.8%	17.1%
10–19	1,067,958	9.8%	13.6%
20–49	721,416	6.6%	13.1%
50–99	259,247	2.4%	6.8%
100–249	140,671	1.3%	8.9%
250–499	32,567	0.3%	5.4%
500–999	12,099	0.1%	4.6%
1,000–4,999	6,958	0.1%	9.3%
5,000–9,999	448	0.0%	1.8%
10,000+	112	0.0%	0.5%
Total	**10,952,004**	**100.2%**	**100.0%**

[a] Data based on February 2002 InfoUSA Frame of Establishments.
[b] Data based on distribution of O*NET sample in Wave 1.1 (see *Section B.1*). Future O*NET data collection waves will be designed so that smaller establishments are sampled at roughly the same rate.

The data in *Exhibit A-4* show that on the sample frame, 60.8% of the establishments have 1 to 4 employees and 18.8% have 5 to 9 employees. This compares to an expected 18.9% of the O*NET sampled establishments belonging to the 1 to 4 employee group and approximately 17.1% belonging to the 5 to 9 employee group. This extreme under-sampling of the smallest establishments minimizes burden to small establishments and improves sample design efficiency. Because the O*NET data collection effort obtains a specific number of employee respondents, it is more efficient to over-sample larger establishments because they have more employees in our occupations of interest compared to the smaller establishments. As discussed in *Section B.1.5*, we use employer size of the establishment as an explicit stratification variable during the early stages of sample selection to control for this under-sampling of smaller establishments. Sample weights created for establishments and employees reflect this disproportional sampling between strata so that O*NET estimates reflect the distribution of the sample frame and not just the sample.

In addition to reducing the number of small establishments sampled, it is estimated that the level of burden on smaller establishments will be much less than for larger establishments because of the nature of the data collection methodology. The majority of the establishment

burden lies within the sampling phase of data collection during which an establishment's POC prepares a list of all employees in the target occupations. For small establishments, the time needed to prepare this list is minimal because they only have a few employees.

A.6 Consequences of Collecting the Information Less Frequently

A.6.1 Consequences of No Data Collection

The O*NET database is the most comprehensive source of occupational information in the U.S. No other similar comprehensive, reliable, and valid source is available.

If the O*NET data are not collected, U.S. citizens and establishments will have few options that meet their needs. O*NET data are used by employers, job seekers, students, educators, and workforce development specialists to promote a healthy American labor exchange system. Citizens may continue to use the outdated DOT, as many of them have done even since the discontinuance of job analysis activity to update the DOT. These users are already missing key information about job changes during the last decade and, given its age, may question the release of data in the DOT. For example, consider the changes in the workplace in recent years resulting from IT advances. In addition to new occupations created by emerging technologies such as the Internet, most jobs in the workplace today have been significantly impacted by the use of the computer.

More than 10,000 of the 12,762 occupations in the DOT were last updated in 1977. Users who continue to rely on the DOT will not have the new information items that the O*NET database provides or the advantages of its new data structure and standard classification. They will continue to struggle with how to relate the DOT to information on the labor market, as the DOT Classification System does not match well to the SOC. They will also continue to struggle with the unwieldy DOT data, which contains more than 12,000 occupations and narrative information in a nonstandard format.

Users may also use the interim O*NET 3.1 database. As described in *Section A.1.1*, O*NET 3.1 uses existing information from the DOT, recast into the O*NET Content Model and updated based on analyst review. O*NET 3.1 is a major improvement over the DOT, as it is more current, in a standardized classification structure and format, and in an electronic form that is easy to use. However, because O*NET 3.1 utilizes source data derived from the DOT, it is based on outdated information. Without the prospect of updates to the data provided by the O*NET *Data Collection Program*, users will have to rely upon outdated information for important decisions they must make. Approximately 150 occupations will be updated as a result

of data collection in 2002, but most occupations would still be based on outdated data from the DOT.

Without the continuation of O*NET data collection, users might turn to various data sources that are neither comprehensive nor statistically based. For example, information from job openings listed with the Employment Service might be used to identify employer hiring requirements. This information, while useful, represents only the jobs the Employment Service receives. It is not representative or statistically valid and can provide only a few of the elements contained in O*NET data.

The consequence of not continuing the O*NET *Data Collection Program* is that the millions of users who need O*NET data to make the important life, business, and policy decisions described in ***Section A.2*** will have to make these decisions using information that is obsolete, incomplete, and not part of the dynamic and progressive U.S. economy. Such data are of questionable validity and reliability. These users will not have the benefit of practical results from the publicly funded research that has led to the O*NET Program. Further, O*NET users, as well as the occupational information and I/O psychology fields, will not have the benefit of continuous improvement and further research in the measurement of worker characteristics and work requirements that O*NET data provide. Millions of people are currently using O*NET data, and the number of customers is expanding quickly as more private and public developers integrate O*NET data into their products. Updating the O*NET database is crucial to providing employers, job seekers, students, educators, and counselors with the most up-to-date information about occupations.

There are no legal obstacles to reducing respondent burden. Technical issues related to response burden are addressed in ***Section B*** of this request.

A.6.2 Frequency of Data Collection

The O*NET *Data Collection Program* is a continuing data collection that populates and refreshes the database regularly, constantly reflecting changing labor market and skill requirements. We are requesting a 3-year extension of the O*NET data collection effort for the period FY2003–2005, at which point approximately 664 occupations will have been surveyed. We anticipate that in FY2005 we will seek an extension from OMB for FY2006 and FY2007 to complete the survey for the remaining 310 occupations. We provide a complete plan in this package through FY2008 (when the last set of data would be updated), so that the full data collection effort is presented to assist in the OMB review of the O*NET data collection request. The entire plan is subject to annual appropriations. The O*NET database will be updated once or twice a year beginning with the first update in February 2003, incorporating completed

occupations, providing the most current information on those occupations that have been surveyed.

A.7   Special Circumstances

The study will be conducted in a manner consistent with the guidelines in Title 5, Section 1320.6 of the CFR. There are no special circumstances that might require deviation from the guidelines.

A.8   Consultation Outside the Agency

Expert review was conducted and public comments were solicited prior to Wave 1.1. The comments from the experts and the public were used to guide the data collection design submitted to the Office of Management and Budget in December 2000. The Office of Management and Budget approved the current data collection effort on April 4, 2001 (OMB Number 1205-0421). The Terms of Clearance for the current effort required that the Department of Labor resubmit the data collection for approval after 18 months with an analysis of response rates, response bias, effectiveness of incentives, and the use of first-class postage stamps. These issues are summarized in *Appendix F.* The data collection design proposed in this package is an enhanced version of the first-year design.

Public comment and expert review comments for this 3-year OMB clearance package were solicited during April–June 2002. Responses to public comments are provided in this section. *Exhibit A-5* lists expert reviewers who have been consulted in the development of this submission package. Comments from these experts are incorporated, as applicable, into appropriate sections of this Supporting Statement.

The Federal Register notice, initiating the 60-day period for public comment, was published on April 19, 2002 (see *Appendix E*). A total of four letters were received from the public. Two came from state employment officials in Michigan and Pennsylvania, writing to express their general support for the O*NET Program and the importance of the O*NET data to their state programs. They did not have any specific comments. One letter came from a state employment official in California. The letter said that their "experience to date with O*NET has been favorable," and that the data collection is necessary, adding that "current occupational data

Exhibit A-5. Expert Reviewers

Name	Organization	Telephone Number
Department of Labor, not in Employment and Training Administration		
John Galvin	Bureau of Labor Statistics	202-691-6400
Michael Pilot	Bureau of Labor Statistics	202-691-5703
Michael Horrigan	Bureau of Labor Statistics	202-691-5701
Alan Dorfman	Bureau of Labor Statistics	202-693-3641
Frederick Conrad	Bureau of Labor Statistics	202-691-7513
Federal Government, Outside Department of Labor		
Sylvia Karman	Social Security Administration	410-965-7693
Non-Government		
Janet Wall	Sage Solutions	240-683-5824
Michael Campion	Purdue University	765-494-5909
John Campbell	University of Minnesota	612-625-9351

are needed by incumbents and subject matter experts to validate worker requirement ratings currently in the O*NET database." The letter then offered two specific comments. The fourth letter came from a private company. The writer expressed his appreciation for the O*NET program and said that "the information is both vital and necessary for America to continue to develop, and redevelop, in the future." He then offered three specific comments. The five comments from the latter two letters are summarized below, along with the response from the Department of Labor, Employment and Training Administration (ETA).

- **Comment:** The estimate of 30 minutes per questionnaire is generally accurate based on the commenter's experience. However, workers with low reading or English language skills will take longer.

  Response: The project the commenter refers to used earlier versions of the O*NET questionnaires that took significantly longer to complete than the current O*NET questionnaires. The 30-minute estimate is an average, based on timing data collected during development of the current questionnaires and the 1999 pretest. Some respondents will take less time and some more time, but we have not had any complaints about that estimate from any of the more than 12,000 respondents to date.

- **Comment:** The estimated burden time for employers of 70 minutes is unrealistic, based on experience collecting O*NET data in California on several pilot projects. The commenter listed the tasks the employers were expected to perform in these projects and concludes that "two to five hours might be a more realistic estimate of employer burden, depending on the number of occupations and how readily the occupations translate to O*NET occupations." The commenter suggested that the

employers be provided with "how-to" materials to help with these tasks, or, better yet, have an O*NET coordinator work directly with the employee's in-house coordinator to facilitate the process.

Response: The protocol followed by California in the pilot projects cited by the commenter was somewhat different and more labor-intensive for the employer than the current data collection protocol. The derivation of the 70-minute estimate is shown in *Exhibit A-6*. The time estimates for the POC contacts are based on observations of the BLs actually making the contacts. The time estimates for the POC to construct the roster (average of 20 minutes) and distribute the questionnaires (average of 15 minutes) are based on (1) discussions with the BLs, who are in frequent contact with the POCs during this process, (2) the expectation that we will select an average of only four occupations per establishment and two incumbents per occupation, and (3) the cap we have placed on POC burden by limiting the number of occupations selected in an establishment to no more than five and the number of selected employees to no more than 20. Regarding "how-to" materials and the use of an O*NET coordinator, we include considerable information and instructions in the various mailings to the POCs, and BLs fulfill the role of an O*NET coordinator.

- **Comment:** To improve the establishment response rate, send an initial "Request for Interest" letter with a self-addressed postcard to establishments and then contact only those establishments that return the postcard indicating their willingness to participate.

  Response: Rather than increasing the establishment response rate, we believe that this approach would have the opposite effect. This is somewhat similar to the "active consent" approach to obtaining parental consent in school-based education surveys, which consistently produces very low response rates. Further, without an initial telephone contact, we would not know if the establishment is eligible (still in operation at the sampled location) or who the best person is to contact at the establishment. We are not aware of any similar survey that uses this approach.

- **Comment:** Offer a higher incentive to sampled employees who respond via the Internet instead of returning the completed paper questionnaire by mail.

  Response: We do not think this approach is advisable. To implement it, we would have to pay respondents who respond via the Internet an additional incentive, on top of the $10 prepaid incentive that is included in the questionnaire package. The cost of processing this additional incentive would likely exceed the cost of keying the paper

questionnaire. Further, the offer of an additional incentive to those who complete the web questionnaire might be viewed negatively by sampled employees who do not have access to the Internet and might affect their willingness to respond.

§ **Comment:** What is the Department of Labor's intention for the future as it relates to ensuring that the interest aspect of Worker Characteristics will be updated?

Response: The O*NET Program will maintain and update the interest aspect of the Worker Characteristics. However, the interest information is not collected via incumbent surveys. Instead, the interest information contained within the O*NET database, Occupational Interest Profiles (OIPs) and Occupational Reinforcer Patterns (ORPs), is developed via occupational analysts using stimulus materials drawn from the O*NET database. OIPs and ORPs for the vast majority of occupations were recently developed in 1999 and are currently available in the O*NET 4.0 database (for a detailed description of these processes see *The Development of Occupational Interest Profiles for O*NET* [Rounds, Smith, Hubert, Lewis, & Rivkin, 1999] and *Determining the Occupational Reinforcer Patterns for O*NET Occupational Units* [McCloy et al., 1999]). Future plans call for developing OIPs and ORPs for new occupations as they are identified, as well as periodically monitoring the currency of data for all occupations to determine when updates are needed.

A.9   Payments or Gifts to Respondents

A.9.1 Incentives for the Employer and the POC

In Wave 1.1, several incentives were offered to POCs to encourage their participation in the O*NET *Data Collection Program*. At various stages of the process, the POCs received a clock, a mouse pad/calculator, and the O*NET Toolkit for Business. This Toolkit is a packet of information about the O*NET Program that managers can use for human resource planning, including a guide for writing job descriptions. The clock and the mouse pad/calculator were offered as gifts to the POC, while the Toolkit was offered primarily as an incentive for the employer because it was designed to benefit the company as a whole, rather than any specific person within the company. Although the Toolkit may appeal directly to some POCs—particularly those who work in the human resources area—it is not universally effective in its ability to create reciprocal effects to obtain POC cooperation.

For example, POCs in many small establishments have suggested that the Toolkit addresses activities and functions that are not particularly relevant for their operations. Consequently, while the Toolkit may be accepted and appreciated by the POC, it does not

provide the reciprocation effect required for sustaining continued cooperation from the POC throughout the various stages of data collection. Further, in some organizations, the POCs may have considerable latitude in unilaterally determining the level of cooperation their companies will provide to the O*NET *Data Collection Program*. In such situations, gifts viewed as beneficial to the company, but not to the POCs personally, could have little or no effect on gaining establishment cooperation.

The clock and the mouse pad/calculator, although potentially of greater personal appeal than the Toolkit, have a small perceived value when compared to the POC's responsibilities to the O*NET *Data Collection Program*. These responsibilities include:

- reading the introductory package to become familiar with the purpose of the O*NET *Data Collection Program* and the role of a POC;

- seeking permission within the company, as necessary, to participate in the O*NET *Data Collection Program*;

- making a roster of all employees at the location who work in up to five different occupations;

- participating in a sampling process that selects up to 20 employees from these occupations and maintaining this sample roster for future reference;

- distributing questionnaires to the sample persons within the company and addressing their questions and concerns about the survey; and

- distributing follow-up materials to employees, including thank you/reminder cards, replacement questionnaires, and email requests for cooperation.

Because POCs are the only link with the O*NET respondents, they must be fully committed to the data collection process. They are our representative within the establishment who communicates the importance of the O*NET Program. One of the lessons learned in Wave 1.1 is that the inexpensive office supply items, which have a total value of less than $10, may not be adequate to motivate POCs to complete the O*NET functions. Further, in situations where the Toolkit is not valued by the POCs, there is a real risk that the POCs will deliver minimally on their O*NET commitments. This usually translates into poor cooperation at the employee level.

To obtain the level of commitment from POCs necessary to maximize the establishment response rate, incentives with a higher perceived value and more universal appeal are needed. Therefore, the plan described below will be implemented for providing the POC and the employer with effective incentives to encourage their participation.

The employer will be offered the O*NET Toolkit for Business for agreeing to participate. The Toolkit provides the establishments with guides to using O*NET data and O*NET OnLine, as well as case scenarios to illustrate how O*NET data can be helpful in human resource and other management applications. Thus, the Toolkit serves a second important role in that it helps employers understand the value of the O*NET *Data Collection Program*.

Of the inexpensive desk items, only the clock with the O*NET logo will be used in Waves 2.1 and beyond. The mouse pad will be replaced by a Department of Labor Certificate of Appreciation in an attractive frame suitable for displaying in an office setting. The Certificate will be personalized with the name of the POC and signed by a high-ranking DOL official. This gift will be included in the mailing package to the POC that contains the questionnaire packets for the incumbents. The combined cost of this gift and the O*NET clock is approximately $10.

The main purpose of these incentives is to (a) establish the trust of the POC that we will deliver on our promises of gifts, (b) provide an inducement for the POC to read through the materials in the introductory package, and (c) create a sense of obligation for the POC to follow through with the early stages of the O*NET data collection process through sampling.

As a further incentive to cooperate with the initial data collection request and to instill a greater sense of obligation in the POC to fulfill all of his or her role, the POC will also be offered the choice of one of the following four incentives:

1. $20 payment to the POC
2. $20 payment to the establishment
3. $20 payment to a charity in the POC's name
4. $20 payment to a charity in the establishment's name.

All payments will be paid via money orders. The money order will be written in either the POC's name (Option 1) or the establishment's name (Option 2) because the POC may not be able to directly accept a cash gift but may be able to accept cash on behalf of the establishment.

If a monetary gift cannot be accepted by the POC or the establishment, the POC may request that $20 be donated to charity (Option 3 or 4). Here, the POC will be offered a choice from a list of approximately 10 national charitable organizations. These organizations will be selected from government-approved, Combined Federal Campaign charities. The POC may also choose to have the money donated in the name of either the POC or the establishment.

Because many federal agencies do not allow their employees to accept monetary incentives, the offer of a $20 money order or contribution to charity will not be made to federal employees. This is not true for local and state government agencies, however. For example, the

National Survey of Child and Adolescent Well-being (The NSCAW Research Group, in press) reports that the majority of local and state child welfare agency employees can accept monetary as well as non-monetary incentives.

This incentive plan is quite flexible and allows tailoring of options specifically to meet the needs of the wide variety of types and sizes of establishments that the O*NET BLs will contact. It is believed that these incentives will establish goodwill and a sense of obligation in the POC and maintain a professional, cordial relationship with that person.

A.9.2 Rationale and Justification for the POC Incentive

Several approaches were used to determine the most appropriate incentive options. First, a thorough review of the literature was conducted to determine the most effective incentives. Once this review was complete, options were discussed with a wide range of individuals working on the O*NET *Data Collection Program*, including BLs, survey methodologists, programmers, survey managers, and employees involved in the O*NET data preparation. In these meetings, the rationale of each of the incentives, as well as the logistical and cost implications of incorporating them, was discussed.

Prepaid cash and check incentives are consistently found to be the most effective form of incentive. Church's meta-analysis of 38 studies found that monetary incentives, when prepaid, improve response rates over all other types of incentives (1993). Furthermore, Kulka (1994), citing Dillman's research on mail surveys (1991), reports that "Hundreds of studies have been conducted, and review after review—both qualitative and quantitative—concludes the importance of financial incentives." In addition to typical mail surveys, Kulka argues that according to a smaller body of research, "The greatest potential effectiveness of monetary incentives appears to be in surveys that place unusual demands upon the respondent, require continued cooperation over an extended period of time, or when the positive forces on respondents to cooperate are fairly low." Other researchers have confirmed his findings (Singer, Van Hoewyk, Gebler, Raghunathan, & McGonagle, 1999).

In some establishments, a monetary incentive may not be considered appropriate. For this reason, the option of a charitable contribution in the name of the POC or the establishment will be offered. A charitable contribution will provide a way of expressing thanks for their participation. It is believed that the charitable contribution, with a known and obvious monetary value, will engender feelings of obligation to complete the task at hand in the same way as a monetary incentive (Dillman, 2000). In fact, some POCs may feel more obligated to help because their involvement with the O*NET *Data Collection Program* directly results in aid going to a needy or deserving cause.

There has been little research to date on the effect of a charitable contribution, and the results have been mixed. Robertson and Bellenger (1978) found a positive effect on response rates from offering a charitable donation, but Hubbard and Little's research (1988) determined that promised charitable donations actually resulted in lower response rates than a cash incentive, no incentive, or lottery. It should be noted that neither study was an establishment survey; both were mailed market research surveys of individuals in the household population.

To ensure that the charities that are offered to the POC are all federally approved and endorsed, the list of charities will be composed from the Office of Management and Budget's (OMB's) Combined Federal Campaign list of charities. To ensure both name recognition and the credibility of the charitable organizations on the list, the charities will be selected from the top 50 charities on the OMB list according to the volume of contributions as per the American Institute of Philanthropy.

A potential logistical challenge related to implementing a charitable contribution alternative is how to demonstrate to the POC or establishment that the charitable contribution has been made. A money order made out to the charity, along with a letter to the charity and a stamped envelope addressed to that particular charity, will be sent to the POC. That way, the POC will have the satisfaction of personally mailing the contribution and will know that our promises are legitimate. Furthermore, delays will be avoided that would be encountered if the contribution was mailed directly to the charity and data collection was delayed until the charity sent an acknowledgement to the POC.

A.9.3 Offering the Incentive to POC or Establishment

The incentive to participate will be offered to either the POC or the establishment. The O*NET BLs have indicated that sometimes the POC feels uncomfortable accepting the incentive or is not allowed to receive the incentive, according to business policy. A choice of incentives must be provided for these establishments that does not make them feel uncomfortable about participating. A money order will be sent, payable to the POC, the name of the establishment, or a charity in either the name of the POC or the establishment. Although there are no reports in the literature of this practice, it is believed that this incentive will be successful because it will allow the incentive to be offered to persons or establishments as most appropriate.

A.9.4 Incentives for the Employee

In keeping with what has been done in Wave 1.1 and subsequent waves, each employee will be offered a prepaid incentive of $10 to ensure that, once distributed, a high percentage of those job incumbents respond by completing and returning the questionnaire.

Monetary incentives have the greatest potential impact when the respondent has to exert some special effort, such as taking a test or filling out a multi-item questionnaire. The incentive serves to encourage respondents in a task requiring higher levels of involvement and commitment than the typical one-time, face-to-face interview. While the O*NET questionnaires are not tests, the cognitive demands they place on respondents are quite similar in that the respondents are asked to make assessments of the requirements of their job. Using various scales, workers must reflect on their current job to make these judgments. The monetary incentive is instrumental in impressing upon the respondent the importance of this rating task. Respondents who perceive the rating task as important are less likely to make hasty assessments that lack adequate consideration, thus improving the reliability of the data.

In addition, the monetary incentive for respondents is important because they are encouraged to complete the questionnaire on their own time rather than on the job. This minimizes the burden on employers and improves the quality of the data because a nonresponse by workers who were "too busy" to complete the questionnaire on the job could produce a bias reflecting on-job performance levels.

The monetary incentive has the potential to at least partially offset its inherent cost through efficiencies created in the data collection process as a result of higher response rates. For the job incumbent respondent especially—while they are not viewed as a difficult-to-reach population in the usual sense—considerable effort and cost will be expended to identify and reach them through the sampling process. They represent a worker in a specific occupation in a specific establishment in a specific industry. The expense of reaching that particular respondent justifies the cost of a monetary payment to ensure a high rate of response. With regard to the size of the employee incentive, payment amounts were evaluated in the pretest to determine the optimal means to maximize the response rate. Based on those data, a $10 cash incentive was used in Wave 1.1 and subsequent waves and will continue to be used in future waves.

A.9.5 Incentives for Subject Matter Experts (SMEs)

Subject matter experts (SMEs) will provide data for a small subset of occupations. Each SME will be offered a monetary incentive commensurate with that offered the incumbent respondent: $10 per questionnaire. As each SME is asked to complete all four domain questionnaires, the incentive totals $40 per SME. Details of SME data collection methodology are provided in *Section B.2.12*.

A.10 Assurance of Confidentiality

All participants in this data collection effort are assured through written materials that the confidentiality of their answers is in compliance with Section (a)(2) of the *Wagner-Peyser Act*. In addition, very little personal information about the respondent is collected and there is no identifying information, such as name or place of employment, on the survey.

The contractor for the data collection task, RTI, has extensive experience in protecting and maintaining the confidentiality of respondent data collected from surveys. RTI has drawn upon its experience in designing the data collection procedures incorporated in this program to ensure confidentiality. In addition, all research involving human subjects is reviewed and approved by RTI's Institutional Review Board (Multiple Project Assurance number M-1496) prior to study implementation.

The collection of survey data is at the employee level at selected establishments and within recruited associations. Letters to employees from the study project director and the Department of Labor, along with Participation Information Sheets for Employees, contain essential program information that enable the employee to make an informed decision regarding his/her voluntary participation in the data collection effort. Examples of the letter and Information Sheet are provided in *Appendix G*. The commitment of O*NET *Data Collection Program* staff to maintain the confidentiality of information provided by the employee is stated clearly in these documents.

As noted above, employees are asked to complete their questionnaire on their personal time, not company time. This enables the employee to select a comfortable and private setting, if desired, in which to complete the questionnaire. The employee mails the completed questionnaire directly to RTI by using a reply envelope supplied by the O*NET *Data Collection Program*. The individual responses are processed using a study ID number. All O*NET *Data Collection Program* staff are required to sign a confidentiality pledge that assures each respondent that the confidentiality of responses to the questionnaire will be maintained. Only authorized staff have access to the completed instruments and data files. The completed and processed questionnaires will remain stored in a secure document control area until federal authorization has been granted to destroy them. All computer files, including those associated with the control system, are password protected.

As discussed in *Section A.3*, the O*NET *Data Collection Program* also utilizes an Internet-based system to allow employee respondents to report data electronically. The respondents have restricted access controlled by a user ID and password authentication scheme.

The web server includes an SSL certificate to allow encrypted transmission of all information over the Internet. This is the same technology used by electronic commerce websites to secure credit card numbers. The database containing the survey data is not accessible by the Internet, but resides on a server inside the RTI firewall. The web data collection application serves as the "go-between" to transmit data from the respondent's computer into the survey database. Only O*NET *Data Collection Program* staff, operating from inside the firewall, have access to the master survey database.

A.11 Questions of a Sensitive Nature

Only one question in the O*NET questionnaire might be considered to be of a sensitive nature. In the Background section, the survey asks respondents if they consider themselves to have a disability using the disability questions developed for the 2000 Census of Population. Completion rates obtained during the O*NET pretest indicate that the vast majority of participants (97%) elected to complete the question, and preliminary analysis of the questionnaire data from Wave 1.1 suggests item nonresponse to this item will be similarly low.

The O*NET sampling strategy randomly selects participants at the individual level. The disability question, along with the demographic questions, will provide descriptive information about the sample of respondents. In addition, these data may allow for some broad comparisons with the characteristics of the general population working in the occupation, using information from the 2000 Census.

Respondents to the O*NET survey are informed that responding to all questions, which includes disability status and the other demographic characteristics, is voluntary. They complete the survey on their own time, in a private setting if they choose. Individuals' identities are separated from their response data so that no characteristics of a respondent are identifiable.

A.12 Estimates of Annualized Hour Burden

Exhibit A-6 displays the estimated burden hours by fiscal year and the average annual burden across all fiscal years. To efficiently collect data for the O*NET *Data Collection Program*, the plan is to use a 64-wave, interweaving data collection design (see *Section B.1.2*). The first four waves are being worked in FY2002, 44 waves are planned for FY 2003, 2004, and 2005, and the remaining 16 waves are planned for FY 2006 and 2007, outside the scope of this clearance request. *Exhibit A-6* shows that we plan to sample 35,216 establishments in FY2003, 34,200 in FY2004, and 34,200 in FY2005. The slight increase in the FY2003 sample is due to

Exhibit A-6. Estimate of Hour and Cost Burden by Fiscal Year for the O*NET Data Collection Program

Activities	Number of Responses per Sample Unit	Average Burden per Response (in Minutes)	Average Burden per Response (in Hours)	FY2003 Sample	Total Hours	FY2004 Sample	Total Hours	FY2005 Sample	Total Hours	Average Annual Burden Hours
ESTABLISHMENT BURDEN										
Verification Calls to Initial Contact at Establishment	1	2	0.03	35,216	1,174	34,200	1,140	34,200	1,140	1,151
Screening Call to POC	1	3	0.05	30,743	1,537	29,857	1,493	29,857	1,493	1,508
Initial Recruitment Call to POC	1	12	0.20	29,195	5,839	28,353	5,671	28,353	5,671	5,727
POC Creates Occupation Lists for Sampling	1	20	0.33	21,311	7,104	20,697	6,899	20,697	6,899	6,967
Call to POC to Sample Workers	1	10	0.17	21,311	3,552	20,697	3,449	20,697	3,449	3,484
POC Distributes Questionnaire Packets	1	15	0.25	17,084	4,271	16,592	4,148	16,592	4,148	4,189
Follow-up Calls to POC	4	2	0.03	17,084	2,278	16,592	2,212	16,592	2,212	2,234
Total Establishment Hours					25,755		25,012		25,012	25,259
Average Burden Per Responding Establishment		70	1.17							
EMPLOYEE BURDEN										
Questionnaires Returned[a]	1	30	0.50	17,097	8,549	15,793	7,897	15,793	7,897	8,114
TOTAL BURDEN HOURS					34,303		32,908		32,908	33,373
TOTAL RESPONDENTS PER YEAR[b]				83,056		79,850		79,850		
COST BURDEN										
Establishments (Assuming $31.70 per hour)					$816,420		$792,874		$792,874	$800,723
Employees (Assuming $16.72 per hour)					$142,933		$132,030		$132,030	$135,664
TOTAL COST					$959,353		$924,904		$924,904	$936,387

Note: Sample sizes and design assumptions are discussed in greater detail in *Section B.1.2* and *Section B.1.5*.

[a] Assumed 10% of returned questionnaires will be complete after data editing. See *Exhibit B-5*.

[b] Total Respondents per Year is the sum of the Verification Calls to Initial Contact at Establishment, Screening Call to POC, and Questionnaires Returned. Per hour, cost burden estimates were computed using July 2000 estimates from the Bureau of Labor Statistic's National Compensation Survey. These estimates were inflated to December 2001 levels using the BLS Employment Cost Index.

Total POC time 3+12+20+10+15+(4x2) = 68 mins + 2 verification = 70

Wave 1.4, which will be designed to complete occupations begun in Waves 1.1 through 1.3 that do not yet have the required minimum 15 completed questionnaires per questionnaire type.

The estimated hour burden displayed in *Exhibit A-6* is shown for the POC and for the employees selected within each establishment. The breakdown of burden hours by specific event is also displayed in *Exhibit A-6*. These estimates of average burden per response for the POC are based on the estimates for the current Wave 1.1 data collection effort. The estimate of the burden associated with completing employee questionnaires (i.e., the employee burden) is based on timings conducted prior to the 1999 O*NET pretest and timing data collected for the questionnaires completed via the Internet during the pretest. There has been no significant content change in the questionnaire since the pretest.

The data displayed in *Exhibit A-6* indicate the number of total respondents per year is 83,056 for FY2003, 79,850 for FY2004, and 79,850 for FY2005. This yields an annual average estimate of 80,919 respondents per year. The annual average estimate of 80,919 represents the establishments contacted, the number of points of contact, and the number of job incumbent respondents to the survey (those who actually return the questionnaire). The verification call would typically be to a receptionist and thus is different from the POC, who is generally in the human resources office. Thus we have counted these calls in our total respondents—this in essence provides the count of establishments contacted. The POC, the principal person we deal with directly, may be contacted several times but is only counted once as a respondent. However, the total burden hours do reflect the multiple contacts with a single POC.

The estimates shown in *Exhibit A-6* were created assuming the entire sample will be collected using the General Employer Sample. Early results from any wave of data collection may suggest using an alternate method of collecting data for an occupation—such as the Association List Sample methodology—or arriving at O*NET estimates using SMEs. These alternate methods of collecting data are discussed in *Sections B.1.6* and *B.1.7*. It is estimated that the burden per employee respondent for collecting data using these sources will be considerably less than that implied in this table. Consequently this table displays an estimated upper bound on the amount of burden needed to complete the data collection effort.

As described in detail in *Section B.1*, the O*NET data collection effort has been designed to minimize burden for selected establishments. In summary, these measures include the following:

- To minimize the burden within any selected establishment, all responding establishments will be asked for employment estimates in 10 occupations or fewer, and the BL will stop asking when approximately 5 positive responses are received.

This will be particularly beneficial for larger establishments that may have employees in several occupations of interest within a wave.

- Establishments will not be selected more than once each calendar year.

- Establishments, occupations, and employees within occupations will continue to be selected using a probability proportionate to size (PPS) selection technique that equalizes the burden between establishments within a design stratum and minimizes the variance-increasing effects of unequal weighting that can occur with other sample selection methodologies.

- No more than 20 employees will be selected within any establishment, regardless of the establishment's total employment.

- As mentioned above, if Wave 1 results suggest obtaining the needed data for an occupation may be too costly or burdensome (e.g., locating establishments with employees in the occupation is difficult), then an alternate data collection methodology will be considered. These alternate methodologies include the Association List Sample and the SME approach.

A.13 Annual Reporting Burden Cost

Exhibit A-6 also displays the cost burden by fiscal year. The cost burden to employee respondents is based on ½ hour of time at an hourly rate of $16.72. This rate is based on the July 2000 BLS National Compensation Survey, inflated to December 2001 rates using the Employment Cost Index.

Given that the POC will often be a human resource manager, we can derive an estimate of cost by using the July 2000 mean hourly wage of $29.95 for Personnel, Training, and Labor Relations Managers, published by the BLS. Again, the BLS Employment Cost Index is used to inflate the 2000 wage to the December 2001 mean hourly wage of $31.70.

The estimated total cost burden for employees and establishments together is $959,353 for FY2003, $924,904 for FY2004, and $924,904 for FY2005. This equates to an annual average of $936,387.

There are no respondent costs for capital or start-up, or operation, maintenance, and purchase of services. There is no cost to employers, POCs, or sampled workers other than the time it takes them to comply with the survey request. The POC will more than likely participate during his/her workday. *Exhibit A-6* indicates that individual POC participation in the O*NET *Data Collection Program* will take about 68 minutes (1.13 hours).

The establishment verification information may be supplied by the POC or may be provided by someone else at the company. The estimate of burden for this is 2 minutes per establishment, making the total burden per establishment/POC 70 minutes (or 1.17 hours).

A.14 Estimates of Annualized Cost to Government

The estimated average cost to the government for the O*NET *Data Collection Program* is $8 million per year. Operational costs involved in the survey effort, including personnel, supplies, and mailing, as well as data processing costs for sampling, survey management, and statistical analysis, were compiled to derive the estimated cost figure.

A.15 Reasons for Program Changes Reported in Sections A.13 and A.14

While the estimated total burden for the data collection effort has increased from the FY2002 projections,[3] the per-establishment estimate has decreased. For the FY2002 data collection effort, we assumed it would take 1.58 hours per responding establishment, while estimates from our current data collection effort suggest it will take only 1.17 hours. On the other hand, for the FY2002 projections, we assumed it would take roughly 40 establishments to complete each occupation.[4] Our current estimates suggest it will take 158. This disparity in the estimated establishment sample size accounts for the increase in the estimated total burden hours between the FY2002 projections and the totals displayed in *Exhibit A-6*. This increase in the establishment sample size is primarily due to the unexpectedly large number of randomly selected establishments needed to find someone in a particular occupation.[5] The most current occupation by establishment industry (SIC) data available from the BLS OES is being used to estimate the occupational distribution we would probably find in any establishment. It is worth noting that while there is a large increase in the establishments contacted, most of these are just one-time contacts in which we do not find the occupation and each additional contact adds relatively little burden but is necessary for us to successfully locate a sufficient number of respondents in the target occupation(s).

---

[3] In the OMB package submitted for the FY2002 effort, it was estimated that 23,305 burden hours would be needed per 300 occupations. This assumed an establishment sample size of 12,000 was sufficient to complete each occupation.

[4] An occupation is considered complete when we have received 15 or more useable, completed responses to each of four questionnaire types.

[5] An establishment that does not have an employee in a particular occupation when it was expected is said to be SOC-ineligible for the O*NET *Data Collection Program*. Our methods for refining the target population to increase the SOC eligibility are presented in *Section B.1.3*.

A.16 Time Schedule, Publication, and Analysis Plans

The major components of the O*NET *Data Collection Program* include sampling, data collection operations, and analysis. ***Exhibit A-7*** provides the expected schedule. The schedule assumes OMB clearance is received on or before September 30, 2002. A change from that assumption would require a recalculation of the data collection schedule. Data collection is carried out using overlapping collection waves, with each wave designed to cover approximately 48 occupations. Projected waves of data collection represent a continuation of the current data collection activities. Each wave is divided into sub-waves to allow for more efficient sampling with less burden and more completed occupations in each wave.

A.16.1 Analyses Conducted at the End of Each Data Collection Cycle

In this section the approach to data cleaning and analyses to be performed are discussed. Most of the planned analyses are conducted after the data files are closed upon completion of each data collection cycle. At the end of this section analyses that will be conducted at the end of each cycle are listed. While these are the planned analyses, there is rich potential for these data to provide information to continuously improve the O*NET *Data Collection Program* and the data collection instruments, and to inform the field of job analysis. For example, the results of the various reliability analyses may suggest the need for further investigation of the data obtained from particular domain items, or from certain occupations. Additional analyses might also be done to explore the correlation between domain items, as well as to examine the ability of the ratings to discriminate between occupations.

Data Cleaning

Paper questionnaires are manually edited so that completely blank questionnaires are dropped; responses to items that should have been skipped are blanked out; multiple responses are blanked out; and codes indicating missing data, multiple responses, and legitimate skips are inserted. Codes for legitimate skips and missing responses are also inserted in the records for questionnaires obtained through the web. In addition, an electronic check is conducted to detect duplicate responses.

Identification and Evaluation of Anomalous Cases

Analyses of the ratings are based on the assumption that raters were qualified, willing, and able to engage in the rating tasks. Unusable ratings are identified as those for which there were no tasks rated important to the rater's job, or those for which there are no ratings for the task information. Additionally, questionnaires missing ratings for more than 50% of the items are considered unusable. Potentially unusable questionnaires are flagged for further analysis.

Exhibit A-7. Data Collection Schedule

Activity Description	Schedule
Year 2002	
OMB Clearance Granted	September 30, 2002
Start Data Collection for	
Waves 2.2, 2.3, 3.1	4th Quarter, 2002
Year 2003	
Update Database 5.0	1st Quarter, 2003
Start Data Collection for	
Waves 1.4, 2.4, 3.2	1st Quarter, 2003
Waves 4.1, 3.3, 4.2	2nd Quarter, 2003
Waves 5.1, 4.3, 2.5, 5.2	3rd Quarter, 2003
Waves 6.1, 5.3, 3.4, 6.2	4th Quarter, 2003
Year 2004	
Update Database 6.0	1st Quarter, 2004
Start Data Collection for	
Waves 7.1, 6.3, 4.4, 7.2	1st Quarter, 2004
Waves 8.1, 7.3, 5.4, 8.2	2nd Quarter, 2004
Waves 9.1, 8.3, 6.4, 9.2	3rd Quarter, 2004
Waves 10.1, 9.3, 7.4, 10.2	4th Quarter, 2004
Year 2005	
Update Database 7.0	1st Quarter, 2005
Start Data Collection for	
Waves 11.1, 10.3, 8.4, 11.2	1st Quarter, 2005
Waves 12.1, 11.3, 9.4, 12.2	2nd Quarter, 2005
Waves 13.1, 12.3, 10.4, 13.2	3rd Quarter, 2005
Waves 14.1, 13.3, 11.4, 14.2	4th Quarter, 2005
Year 2006	
Update Database 8.0	1st Quarter, 2006
Start Data Collection for	
Waves 15.1, 14.3, 12.4, 15.2	1st Quarter, 2006
Waves 16.1, 15.3, 13.4, 16.2	2nd Quarter, 2006
Waves 16.3, 14.4	3rd Quarter, 2006
Wave 15.4	4th Quarter, 2006
Year 2007	
Update Database 9.0	1st Quarter, 2007
Start Data Collection for	
Wave 16.4	1st Quarter, 2007
Year 2008	
Update Database 10.0	1st Quarter, 2008

Note: Sample waves will be released for data collection on a monthly schedule. The monthly schedule is presented in Exhibits B-1a and B-1b.

These cases include questionnaires with patterns of ratings that show little or no discrimination among descriptors or with write-in job titles that do not appear to match the occupation. This analysis is conducted by experts and may include an analysis of task responses or error rates for an occupation. Responses judged to be invalid are excluded from the analysis file.

Compute Sampling Statistics

Basic sampling weights are applied to the data to make inferences to the population of incumbents for each occupation. These weights are computed as the inverses of the overall selection frequencies and the selection probabilities for each selected establishment and each individual participant. The analysis weights for the eligible sample units are adjusted to compensate for unit nonresponse, for both establishments and employees. In order to maximize comparability of O*NET estimates with estimates from other federal sources, the final sample weights will also be poststratified to occupation estimates obtained from the OES survey.

Sampling errors are computed. The analysis weights used in the sampling error computations, as noted above, have been adjusted for nonresponse and are consistent with the complex sampling design.

Nonrespondent Analyses

Frequency of nonresponse is calculated by occupation and by questionnaire to identify O*NET domains or items that respondents refused to complete. In addition, as in the pretest data analysis, establishments that do not respond to the O*NET instruments are analyzed to determine whether certain characteristics of an organization—such as size, location, and industry, or the organizational position held by the POC—predict willingness to provide O*NET data. Nonresponse analyses use logistic regression analysis with respondent/nonrespondent as the dependent variable and establishment characteristic variables as the predictors. Should such features be identified, this information would inform subsequent data collection efforts by suggesting alternative ways of contacting or interacting with the POCs at those firms. If such attempts are successful, the result will be increased response rates in subsequent data collections.

Calculation of Descriptor Values and Reliability

For each occupation, the means and standard deviations of the ratings for each descriptor are calculated. Standard errors are calculated along with the 95% confidence interval around the mean. Following standard O*NET methods (Peterson, Mumford, Levin, Green, and Waksberg, 1997), descriptors with a standard error of no more than 0.75 (a confidence interval half-width of

1.5 or less) are considered reliable estimates of the rating. Occupations are considered to be reliably represented if at least 95% of the descriptor standard errors are estimated at 0.75 or less.

A.16.2   Other Analyses

Interrater Reliability and Agreement

For each O*NET-SOC occupation, the degree of interrater reliability (the consistency of ratings across respondents) and the level of interrater agreement (the absolute level of agreement across respondents) will be calculated annually. The results of the analyses will be used to examine the potential sources of variability across respondents within a specific occupation. These results will also inform an evaluation of the O*NET occupational taxonomy, content model descriptors, and scales as part of a continuous improvement process.

Confirmatory Factor Analyses of Domains

Confirmatory factor analyses provide an empirical demonstration of the plausibility of the theoretical structure believed to underlie the O*NET content model. However, some domains in the content model were not defined by an expected empirical relationship among descriptors; rather, they were defined to maintain comparability with theoretically derived taxonomies or to represent a number of discrete descriptors. Confirmatory factor analyses will be used to model the latent structure of the O*NET content domains. The analyses will be conducted at the midpoint of the O*NET *Data Collection Program*.

Create Occupation Database

The O*NET database is scheduled to be updated once per year. Each annual update will include data for those occupations collected and analyzed during that year. Thus, an annual database update includes occupations from multiple data collection waves, depending on the number of prior waves for which analysis was completed that year. For each occupation collected, the newly calculated means data will replace existing analyst-based data in the database. Metadata will be provided to the users regarding when the data were collected and any other pertinent information that will assist the users in interpreting the data. If elements in the questionnaires have been changed since the previous database update, an analysis will be done to define the impact on the existing database. At this time, updates to the O*NET database are scheduled for the first quarter in calendar years 2003 through 2008, at which point the entire O*NET database will have been updated. This is subject to annual budget levels for the O*NET *Data Collection Program* that would allow data collection to proceed as proposed.

The O*NET database is designated with a version number denoting each annual update, e.g., from O*NET 4.0 to O*NET 5.0. The database is developed and administered using the Oracle database management system. Once the Oracle database is updated, it is used to generate the database for public release as a series of flat text files. It is accessible to the public via the O*NET Consortium website at www.onetcenter.org.

A.17 Display of Expiration Date

The expiration date will be displayed on the cover of the survey questionnaires.

A.18 Exceptions to Certification Statement

There are no exceptions.

B.1 Sampling Universe, Sampling Method, and Expected Response Rate

A multiple method approach for creating and updating the *Occupational Information Network* (O*NET) database has been developed to maximize the amount of usable information the database will contain for each occupation in a cost-efficient, less burdensome, and timely manner. The primary source of information for the database is estimates derived from a survey of establishments and sampled workers from within selected establishments. This primary source of information is referred to as the General Employer Sample. Additional secondary methods utilized include random sampling from purposely selected association member lists (Association List Sample) and selected Subject Matter Experts (SMEs). These three methods of collecting data are discussed in *Section B.1.5* (General Employer Sample); *Section B.1.6* (Association Lists Sample); and *Section B.1.7* (SMEs). The Association List Sample method and SME method have not yet been implemented.

In general, the Association List Sample will be used to supplement responses received from the establishment sample for occupations that are difficult to find using the General Employer Sample. A dual frame adjustment will be made to the sampling weights to adequately account for the coverage overlap between the two sources of information. The dual frame weight adjustment is described in *Section B.1.8*. The SME approach will be used for a very small number of occupations that are difficult to collect data on using the General Employer Sample and Association List Sample.

B.1.1 Sample Universe and Sample Size: Overview of Sampling Approach

A probability sampling methodology is used to select employees from establishments and members from associations. This methodology offers two important advantages over nonprobabilistic methods:

1. The resulting survey statistics provide linearly unbiased estimates of population parameters associated with the particular sampling frames used. The extent to which these estimates reflect all people currently working within a specific occupation depends on the coverage of the sampling frame. Establishments (and subsequently employees within establishments) are selected from a frame that contains nearly 11 million establishments. The associations from which a probability sample of members will be drawn will be selected to cover occupations where the establishment sampling frame was inefficient for the occupation.

2. As with any probability sample, randomly selecting a sample from the frame with known probabilities of selection enables measurement of the precision of the estimates (variance), in addition to estimation of the population parameters themselves.

The samples for the General Employer Sample and the Association List Sample are selected using a stratified sampling approach. The sample design and stratum definition for the Association List Sample will depend on the data available from associations. In general, preliminary contacts with several associations suggest the ability to form strata using member characteristics such as occupation, current workforce status, years working in the occupation, and possibly industry classification of the member's primary employer. Early contact with associations suggests most will be able to provide a list of all members (or a random sample of all members), so a simple random sampling approach within strata would be appropriate for virtually all association sampling designs.

In contrast to the Association List Sample, the General Employer Sample involves selecting clusters via a multistage design, because an adequate list frame of all employees by occupation is not available. The selection process for the General Employer Sample involves selecting establishments within strata at the early stages of selection and subsequently selecting employees within establishments at the latter stages of selection. The sampling design methodology used for the General Employer Sample is similar to what is used in most area household surveys designed to obtain a specified number of respondents in various demographic groups, such as groups defined by age or race. The equivalence of the O*NET establishment design and a typical area household survey design lies mainly in the notion that group membership within each cluster is not known prior to sampling. The exact occupational distribution of employees in sample establishments is unknown before the contact, similar to not knowing the demographic breakdown of a household's occupants prior to the initial contact. To overcome this sampling challenge, the best data available are used to estimate what will be observed in the establishment and select these primary sample units with probability proportionate to a size measure that reflects expectations.

In the General Employer Sample, establishments are selected with probability proportionate to a composite size measure (CSM) and with minimal replacement. The primary advantage of the CSM selection methodology is that after the final stage of sampling is complete (i.e., after employees are selected), the resulting unconditional probabilities of selection associated with employees are closer to being equal within occupations and design strata, compared to most other sample selection methodologies. Having equal unconditional probabilities of selection at the final stage typically reduces the variance associated with the final survey estimates. In addition to the potential variance reduction benefit associated with using a CSM selection methodology, the burden imposed on each cluster (or establishment) within design strata will be equalized as much as possible. This will increase the chance of participation from each establishment because one establishment will not be overly burdened, compared to other establishments, within the same design stratum. This feature is discussed in greater detail in *Section B.1.5*.

Sampling Efficiency and Establishment Burden

The General Employer Sample compares favorably to the other methods of data collection because the overall cost of data collection is lower, coverage of the occupation can be higher, and for most occupations, obtaining data directly from the employee will provide the most useful and accurate information for users of the O*NET database. Efficiency is gained because several occupations assigned to a specific wave of data collection are sampled at the same time. In other words, a sampled establishment generally is not contacted to ask for occupational information on one occupation, but rather on several occupations, depending on the set of occupations targeted in a particular data collection wave.

While it is somewhat intuitive that asking an establishment for several occupations is more efficient than asking for only one, this approach in general can lead to difficulties for other reasons. Specifically, it is undesirable to overly burden establishments by asking them for data on too many occupations. This is particularly true for larger establishments or establishments in industries with a wide variety of occupations. To achieve this balance of increasing the efficiency of the design against minimizing the burden to the establishment:

- The set of target occupations for any particular data collection wave are selected so that small clusters of the occupations are as similar as possible.[1] For example, "Fire Fighters" and "Fire Inspectors" are assigned to the same data collection wave because an establishment that has one of the occupations will have a greater chance of having the second.

- Burden to the establishment is minimized by never asking an establishment point of contact (POC) to provide information on more than 10 candidate occupations. Additionally, the Business Liaison (BL) stops inquiring about occupations as soon as a maximum of five are determined to be within the selected establishment (see *Section B.1.5*).[2]

- The sample design is continually optimized at every wave to take advantage of the information on sampling efficiency and burden learned from previous waves. This allows for rapid achievement of required questionnaires per occupation while minimizing the respondent burden.

---

[1] These small clusters of occupations will be aggregated based on their similarity (to the extent possible) to form the occupations targeted within any data collection wave.

[2] BLs are data collectors/interviewers for the O*NET *Data Collection Program*. They are responsible for initiating and maintaining telephone contact with each of the sampled establishments. In some instances, the BL will stop inquiring about occupations when less than five occupations are found in the establishment. This will occur if the frame data suggests less than five occupations are in the establishment from among the set of occupations targeted in a wave. This may also occur if one or more occupations are rare, and consequently finding them in a particular establishment is so "valuable" that sampling from any other occupations would not be desirable. These "valuable" occupations, or certainty occupations, are discussed in *Section B.1.5.3*.

Additional information about the data collection wave and the selection of occupations for an establishment are provided in subsequent sections.

B.1.2 O*NET Data Collection Waves

There are two potential difficulties in the general sampling methodology that could make data collection for specific occupations problematic. These difficulties, and the method for addressing them, are described below:

1. Specific data are not available on the occupations found in any sampled establishment. To address this, estimates obtained from the Bureau of Labor Statistics' (BLS's) Occupation Employment Statistics (OES) survey will be used to provide the average distribution of an occupation across establishment groups defined by Standard Industrial Classification (SIC). This average occupation-by-SIC matrix will be applied to all establishments on the sampling frame to provide an estimate of the number of employees in each O*NET occupation in a particular establishment. The first difficulty arises because the average can be subject to nontrivial error, particularly when applied to a single establishment. The matrix may suggest, for example, that 10% of the employees in a SIC belong to a particular occupation, but this does not necessarily mean 10% of the employees in *each* establishment classified within that SIC belong to the particular occupation. If the estimated employees within each sampled establishment differs drastically from what is expected, there may be difficulty achieving the desired number of respondents for this particular occupation.

2. If the observed response and eligibility rates vary greatly for a particular occupation from the design-predicted rates, then there may be difficulty achieving the desired number of respondents for this particular occupation.

Various types of error (e.g., frame errors) occur in all surveys and can have a dramatic impact on data collection efficiency by reducing sample yield and increasing cost. To directly address these errors and maintain the efficiency of the data collection effort, each occupation is sampled in sample waves, where the early waves are designed to (1) obtain a fraction of the desired number of respondents within each occupation and (2) provide more accurate estimates of design parameters that, in turn, can be used to design more efficient subsequent sample waves. The subsequent waves are designed to obtain the remaining desired number of respondents within each occupation.

A summary of the data collection wave strategy being employed is depicted in *Exhibits B-1a and B-1b*. These exhibits display the same information and differ only with respect to the ordering of the rows in the tables. *Exhibit B-1a* displays the wave strategy ordered by sample wave, and *Exhibit B-1b* displays the same information ordered by the start of data collection. Waves 1.1, 1.2, 1.3, and 2.1 are data collection efforts that have recently been completed or are

Exhibit B-1a. O*NET Sample Wave Design (by Wave Number)

Wave	Primary/ Secondary Occupations Worked	Employee Respondents Expected[a]	Selected Employee Sample[b]	Selected Establishment Sample	Start of Data Collection (Month-Year)
TOTAL		65,981	144,803	153,799	
1.1-1.3	201/0	8,200	16,831	15,514	Jun-01
1.4		780	1,601	3,866	Jan-03
Wave 1 Total		**8,980**	**18,432**	**19,380**	
2.1	98/0	2,903	5,574	4,682	Aug-02
2.2		1,842	3,655	2,850	Oct-02
2.3		1,842	3,655	2,850	Nov-02
2.4		1,842	3,655	2,850	Feb-03
2.5		67	137	330	Sep-03
Wave 2 Total		**8,496**	**16,676**	**13,563**	
3.1	48/50	1,325	2,544	2,850	Dec-02
3.2		1,382	2,741	2,850	Mar-03
3.3		1,382	2,741	2,850	May-03
3.4		100	205	496	Dec-03
Wave 3 Total		**4,189**	**8,233**	**9,046**	
4.1	48/50	1,325	2,544	2,850	Apr-03
4.2		1,382	2,741	2,850	Jun-03
4.3		1,382	2,741	2,850	Aug-03
4.4		100	205	496	Mar-04
Wave 4 Total		**4,189**	**8,233**	**9,046**	
5.1	48/50	1,325	2,544	2,850	Jul-03
5.2		1,222	2,424	2,520	Sep-03
5.3		1,382	2,741	2,850	Nov-03
5.4		100	205	496	Jun-04
Wave 5 Total		**4,028**	**7,915**	**8,715**	
6.1	49/50	1,325	2,544	2,850	Oct-03
6.2		1,141	2,265	2,354	Dec-03
6.3		1,382	2,741	2,850	Feb-04
6.4		100	205	496	Sep-04
Wave 6 Total		**3,948**	**7,756**	**8,550**	
7.1	48/50	1,325	2,544	2,850	Jan-04
7.2		1,141	2,265	2,354	Mar-04
7.3		1,382	2,741	2,850	May-04
7.4		100	205	496	Dec-04
Wave 7 Total		**3,948**	**7,756**	**8,550**	
8.1	48/50	1,325	2,544	2,850	Apr-04
8.2		1,141	2,265	2,354	Jun-04
8.3		1,382	2,741	2,850	Aug-04
8.4		100	205	496	Mar-05
Wave 8 Total		**3,948**	**7,756**	**8,550**	
9.1	48/50	1,325	2,544	2,850	Jul-04
9.2		1,141	2,265	2,354	Sep-04
9.3		1,382	2,741	2,850	Nov-04
9.4		100	205	496	Jun-05
Wave 9 Total		**3,948**	**7,756**	**8,550**	
10.1	48/50	1,325	2,544	2,850	Oct-04
10.2		1,141	2,265	2,354	Dec-04
10.3		1,382	2,741	2,850	Feb-05
10.4		100	205	496	Sep-05
Wave 10 Total		**3,948**	**7,756**	**8,550**	
11.1	49/50	1,325	2,544	2,850	Jan-05
11.2		1,141	2,265	2,354	Mar-05
11.3		1,382	2,741	2,850	May-05
11.4		100	205	496	Dec-05
Wave 11 Total		**3,948**	**7,756**	**8,550**	

continued

Exhibit B-1a. O*NET Sample Wave Design (by Wave Number) – Continued

Wave	Primary/ Secondary Occupations Worked	Employee Respondents Expected[a]	Selected Employee Sample[b]	Selected Establishment Sample	Start of Data Collection (Month-Year)
12.1	48/50	1,325	2,544	2,850	Apr-05
12.2		1,141	2,265	2,354	Jun-05
12.3		1,382	2,741	2,850	Aug-05
12.4		100	205	496	Mar-06
Wave 12 Total		**3,948**	**7,756**	**8,550**	
13.1	48/50	1,325	2,544	2,850	Jul-05
13.2		1,141	2,265	2,354	Sep-05
13.3		1,382	2,741	2,850	Nov-05
13.4		100	205	496	Jun-06
Wave 13 Total		**3,948**	**7,756**	**8,550**	
14.1	48/50	1,325	2,544	2,850	Oct-05
14.2		1,141	2,265	2,354	Dec-05
14.3		1,382	2,741	2,850	Feb-06
14.4		100	205	496	Sep-06
Wave 14 Total		**3,948**	**7,756**	**8,550**	
15.1	48/50	1,325	2,544	2,850	Jan-06
15.2		1,141	2,265	2,354	Mar-06
15.3		1,382	2,741	2,850	May-06
15.4		100	205	496	Dec-06
Wave 15 Total		**3,948**	**7,756**	**8,550**	
16.1	49/50	1,325	2,544	2,850	Apr-06
16.2		1,141	2,265	2,354	Jun-06
16.3		1,382	2,741	2,850	Jul-06
16.4		100	205	496	Mar-07
Wave 16 Total		**3,948**	**7,756**	**8,550**	

[a] Within each wave, **Employee Respondents Needed** is approximately 105% of the actual respondents needed to complete an occupation (an occupation is considered complete when at least 60 complete and valid responses are received: 15 for each of four questionnaire types.) It is assumed that the final employee respondent sample size will be greater than what is required for some occupations in order for them to be considered complete. This excess is due to errors in the estimates of the occupation distribution of employees within establishments prior to the initial contact.

[b] **Selected Employee Sample** reflects what will be used in the design of the survey. This reflects anticipated employee response and employee eligibility, as well as a sample error rate to account for errors in the estimation of design parameters. Therefore, **Selected Employee Sample** and **Employee Respondents Expected** cannot be used to compute anticipated response rates.

Exhibit B-1b. O*NET Sample Wave Design (by Start of Data Collection)

Wave	Primary/ Secondary Occupations Worked	Employee Respondents Expected[a]	Selected Employee Sample[b]	Selected Establishment Sample	Start of Data Collection (Month-Year)
TOTAL		**65,981**	**144,803**	**153,799**	
1.1-1.3	123/0	8,200	16,831	15,514	Jun-01
2.1	98/0	2,903	5,574	4,682	Aug-02
FY01-02 Total		**11,103**	**22,405**	**20,196**	
2.2	98/0	1,842	3,655	2,850	Oct-02
2.3	98/0	1,842	3,655	2,850	Nov-02
3.1	48/50	1,325	2,544	2,850	Dec-02
1.4	78/0	780	1,601	3,866	Jan-03
2.4	98/0	1,842	3,655	2,850	Feb-03
3.2	48/50	1,382	2,741	2,850	Mar-03
4.1	48/50	1,325	2,544	2,850	Apr-03
3.3	48/50	1,382	2,741	2,850	May-03
4.2	48/50	1,382	2,741	2,850	Jun-03
5.1	48/50	1,325	2,544	2,850	Jul-03
4.3	48/50	1,382	2,741	2,850	Aug-03

Exhibit B-1b. O*NET Sample Wave Design (by Start of Data Collection) – Continued

Wave	Primary/ Secondary Occupations Worked	Employee Respondents Expected[a]	Selected Employee Sample[b]	Selected Establishment Sample	Start of Data Collection (Month-Year)
2.5		67	137	330	Sep-03
5.2	48/50	1,222	2,424	2,520	Sep-03
FY03 Total		**17,097**	**33,727**	**35,216**	
6.1	49/50	1,325	2,544	2,850	Oct-03
5.3	48/50	1,382	2,741	2,850	Nov-03
3.4		100	205	496	Dec-03
6.2	49/50	1,141	2,265	2,354	Dec-03
7.1	48/50	1,325	2,544	2,850	Jan-04
6.3	49/50	1,382	2,741	2,850	Feb-04
4.4		100	205	496	Mar-04
7.2	48/50	1,141	2,265	2,354	Mar-04
8.1	48/50	1,325	2,544	2,850	Apr-04
7.3	48/50	1,382	2,741	2,850	May-04
5.4		100	205	496	Jun-04
8.2	48/50	1,141	2,265	2,354	Jun-04
9.1	48/50	1,325	2,544	2,850	Jul-04
8.3	48/50	1,382	2,741	2,850	Aug-04
6.4		100	205	496	Sep-04
9.2	48/50	1,141	2,265	2,354	Sep-04
FY04 Total		**15,793**	**31,024**	**34,200**	
10.1	48/50	1,325	2,544	2,850	Oct-04
9.3	48/50	1,382	2,741	2,850	Nov-04
7.4		100	205	496	Dec-04
10.2	48/50	1,141	2,265	2,354	Dec-04
11.1	49/50	1,325	2,544	2,850	Jan-05
10.3	48/50	1,382	2,741	2,850	Feb-05
8.4		100	205	496	Mar-05
11.2	49/50	1,141	2,265	2,354	Mar-05
12.1	48/50	1,325	2,544	2,850	Apr-05
11.3	49/50	1,382	2,741	2,850	May-05
9.4		100	205	496	Jun-05
12.2	48/50	1,141	2,265	2,354	Jun-05
13.1	48/50	1,325	2,544	2,850	Jul-05
12.3	48/50	1,382	2,741	2,850	Aug-05
10.4		100	205	496	Sep-05
13.2	48/50	1,141	2,265	2,354	Sep-05
FY05 Total		**15,793**	**31,024**	**34,200**	
14.1	48/50	1,325	2,544	2,850	Oct-05
13.3	48/50	1,382	2,741	2,850	Nov-05
11.4		100	205	496	Dec-05
14.2	48/50	1,141	2,265	2,354	Dec-05
15.1	48/50	1,325	2,544	2,850	Jan-06
14.3	48/50	1,382	2,741	2,850	Feb-06
12.4		100	205	496	Mar-06
15.2	48/50	1,141	2,265	2,354	Mar-06
16.1	49/50	1,325	2,544	2,850	Apr-06
15.3	48/50	1,382	2,741	2,850	May-06
13.4		100	205	496	Jun-06
16.2	49/50	1,141	2,265	2,354	Jun-06
16.3	49/50	1,382	2,741	2,850	Jul-06
14.4		100	205	496	Sep-06
FY06 Total		**13,327**	**26,214**	**28,996**	

Exhibit B-1b. O*NET Sample Wave Design (by Start of Data Collection) – Continued

Wave	Primary/ Secondary Occupations Worked	Employee Respondents Expected[a]	Selected Employee Sample[b]	Selected Establishment Sample	Start of Data Collection (Month-Year)
15.4		100	205	496	Dec-06
16.4		100	205	496	Mar-07
FY07 Total		**200**	**411**	**991**	

[a] Within each wave, **Employee Respondents Needed** is approximately 105% of the actual respondents needed to complete an occupation (an occupation is considered complete when at least 60 complete and valid responses are received: 15 for each of four questionnaire types.) It is assumed that the final employee respondent sample size will be greater than what is required for some occupations in order for them to be considered complete. This excess is due to errors in the estimates of the occupation distribution of employees within establishments prior to the initial contact.

[b] **Selected Employee Sample** reflects what will be used in the design of the survey. This reflects anticipated employee response and employee eligibility, as well as a sample error rate to account for errors in the estimation of design parameters. Therefore, **Selected Employee Sample** and **Employee Respondents Expected** cannot be used to compute anticipated response rates.

scheduled to be launched this fiscal year. Data collection for Waves 11.4, 12.4, 13.3, 13.4, 14.1-14.4, 15.1-15.4, and 16.1-16.4 are scheduled to occur beginning in October 2006—outside the range for which OMB clearance is currently requested. These waves are included in this discussion in order to illustrate the plans for completing the entire set of 974 O*NET occupations.

The general pattern exhibited in Waves 3.1 through 16.4 (**Exhibit B-1a**) shows that sets of approximately 50 *primary occupations* and 50 *secondary occupations* are being targeted using a basic four-sub-wave approach. The *primary occupations* assigned to a wave are those occupations that will drive the design of the survey in that wave; i.e., the data collection effort in an X.1-X.4 sub-wave set will be designed to complete all of the primary occupations assigned to the wave. In order to optimally utilize establishments that are willing to participate but that may have fewer than the expected number of employees in the primary occupations, we will also assign up to 50 *secondary occupations* to the wave. The secondary occupations will be chosen based on the industry concentration of the establishments selected in the wave. Secondary occupations are "second priority" and will become the primary occupations in later waves. Any completed questionnaires obtained for a secondary occupation will be subtracted from the desired amount for the wave in which that occupation becomes primary.

The first wave (i.e., the X.1 wave) in a four-wave set (i.e., X.1-X.4) is designed to obtain 33% of the desired employee sample within each occupation. Preliminary data from the first month of data collection from Wave X.1 will be used to design a much more efficient second wave of data collection. The second wave will be designed to obtain another 33% of the desired employee sample. Similarly, Wave X.3 will use results from the X.1 and X.2 data collection experience to efficiently collect data for the remaining 33% of the employee sample. A fourth

wave is assumed and scheduled to start after the third wave is complete. The fourth wave will be designed to complete any occupations still requiring additional sample.[3]

One important feature of the data collection schedule presented in **Exhibit B-1b** is the overlap of the sample waves. For example, in FY2003 we will work Wave 2.2 in October 2002, then in the next month work Wave 2.3, then 3.1, then 1.4, then 2.4 and so on. This interweaving of sample waves enables a trained staff of experienced interviewers to be maintained while allowing sufficient time between waves for the analysis of the preceding wave results.

To minimize burden to any single establishment, an establishment will never be selected for more than one wave in the same calendar year. To control this, the sample for each wave in a year will be selected from among those not sampled in previous waves. The methodology for controlling this is discussed in detail in the Stage 1 selection process described in **Section B.1.5**.

The pattern of the proposed sample wave structure for Waves 2.2 through 16.4 is a refinement of Waves 1.1, 1.2, 1.3, and 2.1 based on the desire to obtain data for as many completed occupations as possible, as early as possible within the annual funding constraints established within the Department of Labor. The proposed wave structure for Waves 2.2 through 16.4 resulted from the Wave 1.1 experience. It is worth noting that the disproportionate sample allocation seen in Waves 1.1 to 1.3 will not necessarily bias estimates produced from the sample waves compared to estimates that will be generated in subsequent waves. Sampling between waves will be done independently and appropriate adjustments to the sampling weights will be made when pooling the data.

B.1.3 Sample Universe

The central goal of the O*NET *Data Collection Program* is to provide data for each of 974 occupations that are prevalent to varying degrees in different industries in the U.S. Estimates from this program are designed to assist users in distinguishing among occupations and are not necessarily designed to capture all the subtle differences between jobs in different industries. With this in mind, the O*NET sampling universe for each occupation will generally be a subset of all employees in the occupation working in the U.S. This subset, or target population for the occupation, will be defined using two criteria: (1) its workers represent a substantial majority of job incumbents in the occupations and (2) data among this set of establishments can be gathered with reasonable efficiency. Given these criteria, the O*NET

---

[3] As is discussed in **Section B.1.4**, the General Employer Sample will be designed to obtain a minimum of 15 respondents per occupation and questionnaire type. There are four questionnaire types. Consequently, if at the end of a third wave (Wave X.3) it is determined that additional respondents are needed for an occupation and questionnaire type (for example), then Wave X.4 will be designed to obtain these additional respondents.

target population for each occupation will cover, on average, approximately 80% of the total employees in the occupation, with appreciable gains met in both the quality of data collected and the efficiency with which it is gathered, compared to 100% coverage design. These points are discussed in greater detail below.

Defining Target Population from Which to Sample Representative Workers

As described in ***Section A.1.1***, the O*NET Program is a comprehensive system for collecting, organizing, describing, and disseminating information on occupational characteristics and worker attributes. The O*NET database will be used by a wide variety of people, including human resources personnel, career counselors, students, and state and local officials to assist people in identifying alternate career choices, given a set of interests, skills, abilities, training, and experience. It is therefore important to users of the database that the information be reflective of the occupation, in general. Because O*NET data are reported as a mean of all respondents in an occupation, it is important that the data reflect what is generally encountered in an occupation, not unduly influenced by outlier respondents.[4] For example, if a small proportion of librarians works in television broadcasting (0.01%), their responses represent occupational characteristics that may not be typical of the occupation in general and may unduly influence mean estimates, because the respondents' jobs are outliers. For the General Employer Sample, each occupation will be considered individually, and establishments in a selected set of industries and perhaps employer size categories will define the target population for the occupation.
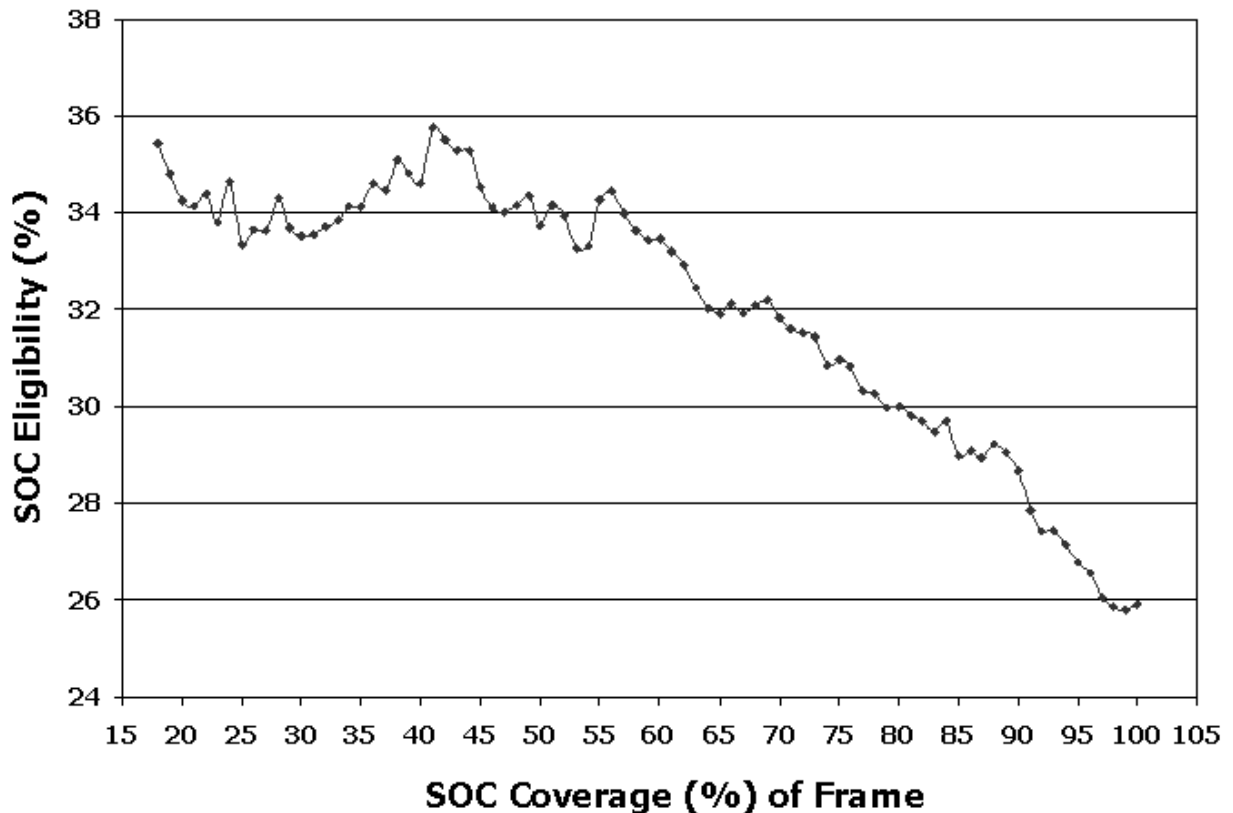
Defining Target Population to Increase Efficiency of Data Collection Effort

Wave 1.1 showed that trying to build a sampling frame that covers 100% of an occupation is inefficient and poses undue burden for some establishments. For example, during the design of Wave 1.1, the industry-occupation matrix data suggested that a very small number of bricklayers could be found in establishments in the "hospitals" SIC. However, asking a POC within a hospital about bricklayers led to some difficulties. In addition to being unduly burdensome, often the BL lost credibility when a POC was asked about occupations probably not associated with his or her establishment, such as bricklayers in hospitals. Additionally, there may be a number of false negative responses from the establishment POC because they simply do not know if some rare occupations exist in their establishment. This would be particularly true for larger establishments. To address these concerns, the target population will be defined so that it will include establishments in industries and size categories where the occupation is most prevalent.

---

[4] In addition to controlling the effect of outlier respondents through effective targeting of SICs, we will also examine the final respondent data to check for outliers.

Exhibit B-2 shows the expected efficiency gained if the coverage of an occupation decreases from 100%. This graph was based on preliminary results from Wave 1.1 and shows, on average, the percent of time an occupation is found in a randomly selected establishment asked about the occupation. So, for example, the graph suggests a job incumbent in an occupation will be found in roughly 26% of the establishments asked, if a frame is used that covers 100% of the occupation. If the percent coverage decreases to 80%, then the occupation can be expected to be found in roughly 30% of the establishments asked. The increase from 26% to 30% suggests an expected increase in sampling efficiency of about 15%, meaning the survey can be designed with 15% fewer establishments to obtain a responding employee sample of the same size. This suggests that targeting an occupation in establishments and associations where it is most prevalent will make the survey more meaningful to the establishment respondents and increase the perceived credibility of the interviewer, while dramatically increasing the efficiency of the survey by lowering the establishment burden and data collection costs.

Exhibit B-2. Estimated Standard Occupational Classification (SOC) Eligibility by Coverage of Sampling Frame

Note: Data based on unweighted preliminary results from the O*NET Wave 1.1 data (January 26, 2002); **SOC Eligibility** is the percentage of establishments that have 1 or more employees in an occupation, among those asked in Wave 1.1; **SOC Coverage of Frame** is based on employment estimates by occupation and industry obtained from BLS's OES Survey (2000).

To further increase the efficiency of the sample and reduce unnecessary burden to the establishment, targeting specific industries will be done to an even greater extent in Waves X.2, X.3, and X.4. The Wave X.1 data will supplement the OES estimates to help determine an adequate targeting in Waves X.2-X.4. Note that even though Waves X.2-X.4 will cover a smaller portion of the occupation than Wave X.1, the pooled estimate will still cover the same population covered by Wave X.1 because the samples are independent. Appropriate adjustments to the sampling weights will be made when pooling the Waves X.1, X.2, X.3, and X.4 data.

Because the target population for each occupation will vary and, consequently, the sampling universe for each occupation may vary, estimated occupational coverage rates will be published along with the estimates produced from the survey.

B.1.4 Employee Sample Size

As described in *Section A.2*, O*NET data will be used for a wide variety of purposes. The broadest general purpose uses are for career development counseling, workforce development, curriculum development, program administration, and labor market and occupational information development. More specialized uses relate to activities such as vocational rehabilitation and human resources activities.

The more general purpose uses of occupational data often involve sorting occupations on specific criteria of interest to a job seeker or individual making a career decision, such as identifying occupations with high ratings on descriptors such as abilities, work context, and interest scales. This use may also involve comparing the results of an assessment, such as an interest inventory, to the ratings of occupations on the relevant scales. Another approach often used is to provide descriptive information for a specific occupation selected from a list of occupation titles. Some general purpose users may also cluster occupations to broader levels of aggregation than provided by the O*NET occupations.

A key issue in sample design is the level of precision required in the resulting data and the cost of producing a particular level of precision, in terms of both dollars and respondent burden. The O*NET sample design has been developed to provide results with a level of precision that should be adequate to meet the needs of general purpose users (those seeking information at the occupation level). The following comments were received in personal communication from one of the professional reviewers of this OMB clearance package:

> The level of accuracy in the O*NET data is considered to be sufficient for the human resource functions … [at the occupation level]. For example, when the data are used on the web page to identify occupations with similar skills to those

of a job seeker, the computer program compares the candidates' self-rated skills to occupational profiles. The means on the skills collected on the O*NET surveys are used for the occupational profiles, and the occupations with the skills exceeding a cut-off score are selected (such as skills exceeding the 3.5 level on the 5-point importance scale). The accuracy of the data are fully sufficient for this purpose. …(M. Campion, personal communication, March 22, 2002).

Users are reminded that for organization or job specific uses, additional information should be gathered to supplement the O*NET data (e.g., O*NET Toolkit for Business, 2000). As described in *Section A.16.1*, release of O*NET data will be accompanied by sampling error measures, allowing users with more specialized purposes to evaluate whether the level of precision is sufficient to meet their needs. Statistical analysis indicates that achieving significant increased precision would require a major expansion of the sample size, with accompanying increases in costs and response burden.

The final technical report of Peterson, Mumford, Levin, Green, and Waksberg (1997) stated that 15 to 30 incumbents typically provide sufficient interrater reliability for describing occupations, given the types of measures the O*NET Program will use to describe people's jobs. This report also presented means and standard deviations for 7-point level scales for the descriptors within Skills, Knowledge, Generalized Work Activities, Abilities, and Work Styles. Statistics were computed separately using the reported data for each of six occupations. The standard errors presented in that report were computed for each descriptor within each study domain using the simple random sampling formula, where each of the subjects responded by choosing one of the points of the scale. Means and standard deviations for the 5-point importance scales are also shown in the report for each descriptor in Skills, Knowledge, Generalized Work Activities, Work Context, Abilities, and Work Styles and for the same six occupations.

The data in these tables indicate that when 15 responses per descriptor are obtained, the mean values for virtually all of the 5-point and the 7-point descriptors will be estimated within 1 to 1.5 scale points, with 95% confidence, for all occupations. And, as these are maximal values, the mean values of most descriptors would be estimated more precisely (i.e., 68% of the descriptors would be estimated within 0.5 to 0.75 scale points). Also, Mumford, Peterson, and Childs (1997, p. 3-8) cited Fleishman and Mumford (1991) as support that variation of 1 to 1½ scale points on a 7-point scale "is typical of that found for well-developed level scales." The data reported by Peterson, Mumford, Levin, Green, and Waksberg (1999) indicated that the average accuracies (averaged across descriptors) for the four domains will usually be even higher than estimations within 1 and 1 ½ scale points. For example, 68% of the time, using an average of 18 raters, skill items were estimated within ±0.47 scale points on the 7-point level scales, or

within ±0.29 scale points on the 5-point importance scales (M. Campion, personal communication, March 22, 2002).

The instrumentation for incumbent worker respondents consists of four distinct questionnaires. Each questionnaire takes approximately 30 minutes to complete, and each sample employee will only complete one of the questionnaires. The same instrumentation is used for all occupations. Consistent with the 1999 O*NET methods pretest results and the earlier results from Fleishman and Mumford (1991), the O*NET *Data Collection Program* considers an occupation complete and ready for inclusion in the final O*NET database when at least 15 respondents are obtained for each of the four questionnaire instruments. The sampling design and sample sizes proposed in this report were created based on this requirement.

B.1.5 O*NET General Employer Sampling Design

The primary sample selection methodology for the O*NET *Data Collection Program* is the General Employer Sample methodology. This arm of the design involves a multistage sample process where establishments are selected at the first stages of selection and employees at the later stages. To summarize, the sample selection process will proceed as follows:

1. **Create establishment sampling frame.** A sampling frame of establishments, covering nearly 11 million establishments in the U.S., will be constructed from the InfoUSA list of all U.S. establishment locations. Several commercial list frames have been evaluated, and the InfoUSA frame is the least costly source of data that meets the study requirements; that is, it has the essential establishment-level information, including industry type and location-specific employment.

2. **Determine industries (SICs) to sample.** Current occupation by industry (SIC) employment estimates generated from the OES program will be merged with the InfoUSA frame information. These data will be merged by SIC and used to determine which SICs to target for each occupation being sampled in a particular wave. In general, occupations will be clustered into the sample waves (see ***Exhibit B-1a*** in ***Section B.1.2***) in a manner that maximizes their similarity. This will allow as much usable information as possible to be obtained from each establishment contact. A subset of industries will be selected for each occupation, using several pieces of information:

   ▪ Those SICs that contain a higher proportion of the occupation will be selected. For example, a SIC that contains 25% of an occupation will be more efficient from which to sample than a SIC that contains 2% of an occupation.

   ▪ Those SICs that have a higher concentration of the occupation will be selected. For example, if 3% of an occupation is located in two SICs and comprises 10% of the total employment in SIC #1 and 75% of the employment in SIC #2, then

sampling from the second SIC group will be more efficient than sampling from the first.

- Adjustments to the over-sampling of larger establishments (establishments with more employees), based on the percent of the occupation that are self-employed, may be warranted. For some occupations, the percent self-employed may be an indirect measure of how the occupation may be distributed across establishment size. For example, if a high proportion of exterminators are self-employed, it might be more efficient to over-sample smaller establishments for this occupation (i.e., establishments with fewer numbers of employees). On the other hand, there may be several occupations where going to a smaller establishment may be inefficient.

As described in *Section B.1.3*, a targeted sampling frame will be constructed for each occupation that covers, on average, 80% of the total employees in the occupation. Sampling will proceed in five stages that are summarized below and explained in detail in *Sections B.1.5* through *B.1.6.3*. First, a sample for a full year of data collection will be drawn from InfoUSA. Second, establishments will be sampled for each sample wave. Third, occupations will be selected for each establishment. Fourth, employees will be selected from establishments. Fifth, questionnaires will be randomly assigned to employees.

1. **Stage 1 of Selection—Selecting establishments for Stage 2 frame**. The initial InfoUSA sampling frame will be stratified by 6-digit SIC and total number of employees. A very large sample of establishments will be selected from this sampling frame, using a stratified, simple random sampling technique. At this first stage of selection, a sample of roughly 450,000 establishments will be selected and, from this, subselected for each sample wave being worked in the calendar year.

   At this first stage of selection, only a very minimal amount of information will be kept for each establishment on the datafile. These data are:

   - 6-digit SIC;
   - total number of employees; and
   - InfoUSA establishment identifier.

   This stage of selection is an element of the overall design for two important reasons. First, it will minimize costs associated with the selection process. The cost-per-record of obtaining this minimal amount of information on each establishment is considerably lower than the cost-per-record of obtaining address, telephone number, etc. for each establishment. Second and even more important, subselecting from the resulting Stage 1 sample allows control of the number of times any single establishment is contacted in a calendar year. Establishments with a very large number of employees and establishments in SICs with a large percentage of employees in the occupations of interest within a wave will be selected with certainty and therefore would probably be selected more than once in a year if this requirement were not imposed. To minimize burden on any single establishment within a calendar year, an establishment will be selected no more than once.

The first stage of selection is discussed in detail in **Section B.1.5.1**.

2. **Stage 2 of Selection—Selecting establishments**. For each sample wave under consideration, a composite size measure (CSM) will be created for each establishment in the set of 450,000 that reflects the desired sample in each occupation of interest. This CSM is a function of the desired sampling rates for the occupations of interest within the wave, as well as the expected number of employees an establishment will have in each of the occupations. An establishment's exact number of employees by occupation is not available on the establishment sampling frame. To estimate this distribution, data from the occupation by SIC matrix provided by the OES survey will be used.

   Establishments will be selected with probability proportionate to their CSM and with minimal replacement, using Chromy's selection procedure (Chromy, 1979). Selection of the establishments will be done within strata, defined by the total number of employees in the occupations of interest and by industry groupings. Intuitively then, by selecting establishments with probability proportionate to their CSM, those establishments with a higher proportion of occupations of interest will have a larger chance of being selected into the sample. This is an important feature of the CSM methodology. If establishment A has more people in the occupations of interest than establishment B, then establishment A will have a greater chance of being selected in the sample, even though establishment B may have a greater total number of employees across all occupations. Additional information on the CSM technique is described in Folsom, Potter, and Williams (1987).

   A list of all selected establishments will be sent to InfoUSA to obtain detailed information on the sample of establishments. This detailed information will not be used for subsequent stages of selection; however, it will be used by the data collection staff (BLs). The additional data include address, telephone number, and other such contact information.

Additional detail on this second stage of selection is provided in **Section B.1.5.2**.

3. **Stage 3 of Selection—Selecting 10 occupations**. Among those establishments selected for the General Employer Sample at Stage 2, each establishment will have 10 (or fewer) occupations randomly selected with probability proportionate to the sampling rate times the establishment's estimated number of employees within the occupation. After removing the initial set of certainty occupations (and recording the expected number of times these occupations are selected, which could be greater than one), sampling will be done without replacement and the order in which occupations are selected will be retained. The results of this process will be an ordered list of occupations that were randomly ordered proportional to their "value" to the O*NET wave under consideration. Therefore, those occupations that are rare (have larger sampling rate) and those occupations that are believed to be highly prevalent in an establishment will have a greater chance of appearing near the top of the ordered list. The set of 10 occupations, as well as the number of times each occupation is selected (which can be greater than one for certainty occupations), will be retained on the control system for each establishment.

The list of the 10 (or fewer) occupations selected for each establishment will be entered into the control system for this project. When a POC is identified within each establishment, the BL will move down the list of occupations and ask the POC to provide an estimate of the total number of employees in the occupations. Each time the interviewer receives a response greater than zero, a counter within the control system will increment by the number of times the occupation was selected at Stage 3. If this counter reaches five before the BL completes the list, the BL will stop. Note that this methodology accomplishes three important things: (1) selecting 10 occupations without replacement and retaining their order of selection on the list read by the BL means that if the interviewer stops at any point in the list, the sample of occupations up to that point will be a random sample with known probabilities of selection; (2) there will never be more than five positive responses on the list; and (3) the perceived burden of this data collection effort is minimized for the POC. Regarding this third point, if the company is large, for example, and happens to have employees in all 10 occupations, stopping after 5 minimizes the perceived burden to the POC, in contrast to asking for employment estimates for all 10 occupations and then subselecting 5. After the five (or fewer) occupations are identified, the POC will be asked to compile a roster of employees in each of these occupations.

The third stage of sample selection is discussed in detail in *Section B.1.5.3.*

4. **Stage 4 of Selection—Selection of employees**. At this stage of selection, employees will be randomly selected from the roster of employees. The number of employees selected from each occupation will be roughly proportional to the number of times the occupation was selected. In order to further minimize burden, the total number of employees selected within any single establishment will never exceed 20.[5] Additionally, the number of employees selected from an establishment within the same occupation will never exceed 8. Employees will be randomly assigned at Stage 5 to one of the four questionnaire types, so we expect that no more than 2 (8 ÷ 4) respondents will be selected from within the same establishment for the same occupation.

The fourth stage of sample selection is discussed in detail in *Section B.1.5.4.*

5. **Stage 5 of Selection—Randomly assigning employees to questionnaire types.** The last randomization stage needed for the data collection effort is the assignment of selected employees to questionnaire types. The survey is designed to obtain 15 respondents for each occupation to each of four different questionnaire types (Skills Questionnaire, Generalized Work Activities Questionnaire, Work Context Questionnaire, Knowledge Questionnaire). Obtaining a minimum of 15 respondents for each questionnaire is critical for this data collection effort based on previous results that examined the expected accuracy of estimates produced from the survey. At this last stage of selection, all employees selected at Stage 4 will be randomly assigned to one of the four questionnaire types, proportional to the number of employee respondents needed for each questionnaire type.

---

[5] The maximum 20 selected employees chosen per establishment was based on observed results from the Wave 1 experience.

B.1.5.1 Stage 1 of Selection—Selecting Establishments for Stage 2 Frame

The first stage of selection for the General Employer Sample will be the selection of a large sample of establishments from the InfoUSA list of all U.S. establishment locations. This selection procedure is summarized in *Exhibit B-3*.

Exhibit B-3. General Employer Sample, Stage 1 Sample Selection Summary

Sampling Frame	InfoUSA list of approximately 11 million establishments
Stratification	6-digit industry code* by total number of employees in the establishment The employee strata will be: Employee Strata (Number of Employees) 1–4 5–9 10–19 20–49 50–99 100–249 250–499 500–999 1,000–4,999 5,000–9,999 Over 10,000 At this stage, the employee size strata reflect the total employees in the establishment and not the total employees in the occupations of interest within any particular wave.
Selection Methodology	Simple random sampling within strata. As mentioned in *Section A.5*, establishments with a larger number of employees will be over-sampled, and establishments with the largest number of employees will be selected with certainty or near certainty within the industry codes of interest.
Sample Size	Roughly 450,000 establishments.

* The 6-digit industry codes comes from InfoUSA and is based on OMB's Standard Industrial Classification.

After this stage of selection is complete, let

$w_i^{(1)}$ =   Stage 1 sampling weight for establishment i. This weight will be equal to the inverse probability of selection for establishment i from the initial InfoUSA list of establishments. Note that

$$\sum_{\substack{i \\ \textit{Across All SIC} \\ \textit{And Emp Size} \\ \textit{Strata}}} w_i^{(1)} \quad = \quad$$ Total establishments in the initial InfoUSA file among the industry codes of interest.

At this stage of selection, this sampling weight will not depend on occupation or employee.

B.1.5.2 Stage 2 of Selection—Selecting Establishments

The second stage of selection for the General Employer Sample will be the subselection of establishments from among those establishments selected at the first stage. This subselection process will be done with probability proportionate to a CSM and with minimal replacement. A summary of this stage of selection is provided in ***Exhibit B-4***.

Exhibit B-4. General Employer Sample, Stage 2 Sample Selection Summary

Sampling Frame	Sample of ~450,000 establishments selected at Stage 1.
Stratification	Stratify establishments by the expected number of employees in the occupations of interest for a particular wave and by industry (SIC) groupings.
Selection Methodology	Select establishments with probability proportionate to the CSMs and with minimal replacement. Frame will be sorted by 4-digit SIC and ordered by total number of occupation employees for selection purposes.
Sample Size	Depends on the wave under consideration.

A detailed discussion of the general CSM methodology is provided in Folsom et al. (1987). The CSM assigned to each establishment has several unique features that make it particularly attractive to the design of the General Employer Sample:

▪ Selecting establishments with probability proportionate to a CSM within strata will allow elimination of a substantial portion of the variation in the unconditional employee-level sampling weights within stratum and occupation. Some variation within an occupation by stratum domain will still result from the inability to know the precise number of employees in an occupation within an establishment. Recall, at this stage of selection for each establishment, the number of employees within each occupation will be approximated using occupation by SIC matrix provided by BLS.

▪ The CSM will allow better equalization of the burden of selected sample persons per establishment. The survey will be designed so that no more than 15 employees will be selected to respond per establishment.

▪ The CSM methodology minimizes the number of establishments that need to be selected to achieve the expected respondent sample within the occupations of interest.

Establishments will be selected within strata, using a minimal replacement selection technique discussed in Chromy (1979) and Williams and Chromy (1980). This selection technique offers the advantage of minimizing variance in estimates produced from the survey by ensuring that the actual selection frequencies are different from the expected frequencies by less than one. In summary, this selection procedure involves ordering the sampling frame by some set of variables; dividing the ordered list into equal size intervals or zones based on the CSM and the desired sample size; and selecting one unit from each zone. In a manner similar to explicit stratified random sampling, the variance of estimates can be further reduced with this selection procedure by virtue of the implicit stratification implied by the sort variables. In order to get adequate representation of the establishment sample, the Stage 1 sample within strata will be ordered by 4-digit SIC and subsequently ordered by total employees in the occupations of interest.

Within each of the strata, let

N_{ij} = Total expected employees in establishment i, occupation j.

N_j = Estimate of the total employees in occupation j on the frame. Note that this will equal roughly 80% of the total employees in the occupation in the U.S., since the frame will be built to cover (on average) 80% of an occupation. This estimate will be derived from the Stage 1 sample as $\sum_{i \in Stage\,1} w_i^{(1)} N_{ij}$.

ε_j = Expected person-level completion rate for occupation j. In summary, this rate will account for people selected within an occupation who were later determined to be ineligible for the study and/or occupation. For example, a "cashier" mistakenly fills out a questionnaire that was sent for a "computer analyst." This will also account for employees who did not respond to at least 50% of the questionnaire items.

ρ_j = Expected person-level response rate for occupation j.

o_j = Expected eligibility rate for occupation j. This eligibility rate accounts for establishments with no people currently in an occupation, even though the occupation is represented in the establishment's SIC. These rates tend to be low because the only information about occupational eligibility available for the design of this study is the BLS SOC by SIC matrix which provides eligibility at the SIC level. Thus, the matrix may indicate some percentage of the employees in an SIC are eligible for occupation j. However, not all establishments in this SIC will have at least one employee in occupation j. The eligibility rate, o_j, accounts for this. Note

that by targeting the occupation in only certain SICs (i.e., creating a more targeted frame that covers on average 80% of an occupation), this rate is maximized, compared to what would have been observed if 100% coverage of an occupation were desired.

s_j =   Design parameter correction factor. This accounts for error in the design parameter estimation process (e.g., errors in the predicted response and eligibility rates).

r_j =   Desired number of responding persons from the General Employer Sample in occupation j. In general, this will equal 33% of the desired total sample in an occupation in the first three waves (i.e., $15 \times 4/3 = 20$ people). For Wave X.4, this will equal any additional sample that is needed for an occupation.

n_j =   Desired number of selected persons from the General Employer Sample in occupation j. In other words, $n_j = \dfrac{r_j}{\varepsilon_j \cdot \rho_j \cdot o_j \cdot s_j}$.

f_j =   Sampling rate for occupation j. In other words, $f_j = \dfrac{n_j}{N_j}$.

n^* =   Desired total number of people selected from within an establishment. This number will vary by stratum and will never exceed 20.

L =   Expected eligibility rate for establishments within each stratum. This accounts for selected establishments that have gone out of business or became otherwise ineligible prior to data collection.

θ =   Expected response rate for establishments within each stratum.

The expected results from the above variables are shown in ***Exhibit B-5***.

m =   Desired number of selected establishments in the stratum. Note that the expected number of successfully screened establishments (eligible and responded) is anticipated to be $m \cdot L \cdot \theta$.

$$S_i \;=\; w_i^{(1)} \sum_j f_j N_{ij} = w_i^{(1)} \sum_j \frac{n_j}{N_j} N_{ij} = w_i^{(1)} \sum_j n_j \frac{N_{ij}}{N_j}.$$

This is the CSM for establishment i. As mentioned earlier, notice that this composite measure is a linear function of both the sampling rate, f_j, and the total number of employees in

the occupations of interest, N_{ij}. Establishments will be selected proportional to the S_i; consequently, those establishments with occupations that have a greater sampling rate (i.e., greater f_j) and those establishments with more expected employees in the desired occupations (i.e., greater N_{ij}) will have a greater chance of being selected.

Exhibit B-5. Waves 2.2–16.4 Design Assumptions

Design Parameter	Current Assumption
Mean Establishment Eligibility Rate	84.84%
Mean Establishment Response Rate	64.33%
Mean Occupation Eligibility Rate	27.00%
Mean Employee Response Rate	62.58%
Mean Employee Completion Rate	90.00%
Sampling Design Parameter Correction Factor	84.00%
Mean Number of People Selected per Establishment and Occupation	2
Mean Number of Occupations asked per Establishment (Wave X.1, X.2, X.3)	3
Mean Number of Occupations asked per Establishment (Wave X.4)	1
Number of Complete Employees Desired per Occupation	60

Note: Expectations displayed above are equivalent to the Wave 1.1 experience except for the Mean Occupation Eligibility Rate. Wave 1.1 experience yielded an occupation eligibility rate of 21.78%. As indicated in **Section B.1.3**, we assume that the increased targeting will raise this to 27.00%.

Note that the sum of the CSMs over all selected establishments from Stage 1 within a stratum and data collection wave is equal to the total employee sample size. In other words:

$$S_+ = \sum_i S_i = \sum_i \sum_j w_i^{(1)} f_j N_{ij} = \sum_j n_j \frac{\sum_i w_i^{(1)} N_{ij}}{N_j} = \sum_j n_j = n = n^* m$$

The sample of establishments will be selected with probability proportionate to their composite size and with minimal replacement. With this type of selection, the probability of selecting any particular establishment i is equal to:

O*NET Data Collection Program

$$\text{Prob}\{\text{Selecting Establishment i}\} = \frac{S_i}{\frac{1}{m}S_+} = \frac{mS_i}{S_+} = \frac{S_i}{n^*}.$$

After this selection procedure is complete, a list of all sampled establishments will be sent to InfoUSA in order to obtain address information. This address information will be used by field staff to locate the sampled establishments.

B.1.5.3 Stage 3 of Selection—Selecting 10 Occupations

A design feature of the O*NET *Data Collection Program* that will be implemented to minimize respondent burden at the establishment level is the rule that a POC will never be asked for employment estimates for more than 10 occupations of interest. Additionally, the interviewer will stop asking for employment estimates as soon as 5 (or fewer) positive responses are received. To account for this rule, those establishments that were selected in Stage 2 will first have occupations of interest selected without replacement. A summary of this selection process is presented in *Exhibit B-6*.

Exhibit B-6. General Employer Sample, Stage 3 Sample Selection Summary

Sampling Frame	Establishments selected in Stage 2
Stratification	Establishment
Selection Methodology	Within each establishment, occupations that should be selected with certainty will be chosen first and placed on the ordered list of occupations. Subsequent occupations will be selected with probability proportionate to size (PPS), without replacement and placed on the ordered list in the order in which they were selected.
Sample Size	The number of certainty and without replacement selections will never exceed 10.

Suppose establishment i has J_i occupations associated with it. For each occupation $j = 1,...,J_i$, define the size measure

$$O_{ij} = f_j N_{ij}.$$

Then $O_{i+} = \sum_j O_{ij} = \sum_j f_j N_{ij} = \frac{1}{w_i^{(1)}} S_i$.

In other words, the sum of the occupation size measures across all occupations within an establishment is equal to the CSM assigned to the establishment at Stage 2 divided by the Stage 1 sampling weight.

In a PPS sample of size 5, the expected number of times an occupation j is selected is equal to:

$$E\{j \text{ is selected}\} = \frac{5 \cdot O_{ij}}{O_{i+}}$$

Note within any establishment i, it is possible for this quantity to be greater than 1, suggesting the occupation should be selected with certainty. For these cases, those occupations will be selected with certainty and given an occupation factor (or "number of times selected") equal to:

$$\alpha_{ij} = Randomly\ Rounded\left(\frac{5 \cdot O_{ij}}{O_{i+}}\right).$$

These certainty selections will appear at the top of the list used by the interviewers when they inquire about whether an occupation exists in a sampled establishment.

From among the remaining occupations, a PPS will be selected, without replacement sample, of up to 10 occupations using the Hanurav-Vijayan procedure described in Fox (1989), Golmant (1990), and Watts (1991). Occupations will be placed on the ordered list used by the interviewer in the order in which they are selected from the frame. For these occupations, the probability of selecting the occupation is

$$\text{Prob}\{\text{Selecting occupation j} \mid \text{establishment i}\} = [10 - \omega_i] \cdot \frac{O_{ij}}{\displaystyle\sum_{\substack{j=1 \\ \text{Noncertainty}}}^{J_i} O_{ij}}$$

where ω_i = total number of certainty occupations for establishment i.

For these occupations, the occupation factor will be set at $\alpha_{ij} = 1$.

When the interviewer contacts the appropriate person at an establishment and begins asking the POC for estimates of the total number of employees in each occupation on the ordered

list, the interviewer will stop after the sum of the corresponding α_{ij}'s equals 5. This step is incorporated in the data collection methodology to minimize the POC's perceived burden of the data collection effort.

Suppose the interviewers asks for employment for occupations in $\overline{J}_{Cert,i}$ certainty occupations and $\overline{J}_{Noncert,i}$ occupations. Then $\overline{J}_{Cert,i} + \overline{J}_{Noncert,i} \leq 10$ and the probability of selecting occupation j within establishment i is equal to:

Prob{Selecting occupation j|establishment i} = 1 for the certainty occupations and

$$= \frac{\overline{J}_{Noncert,i} O_{ij}}{\sum_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'}} \text{ for the noncertainty occupations.}$$

B.1.5.4 Stage 4 of Selection—Selecting Employees

After five (or possibly fewer) nonzero responses are received at Stage 3 of the design, the POC at the establishment will be asked to roster all employees within each of the selected occupations. The interviewer will then instruct the POC to systematically select people from these rosters. The sample selection at this stage of the selection is summarized in ***Exhibit B-7***.

The sample sizes for the certainty and noncertainty occupations noted above will result in a desired, approximate self-weighting sample within occupations. Additionally, the total number of employees selected within the establishment will approach n^* as the observed number of employees in each occupation approaches the frame estimates. To see this, note the following.

From earlier discussions, suppose:

Stage 1 probability of selection is equal to $\dfrac{1}{w_i^{(1)}}$,

Stage 2 probability of selection is equal to $\dfrac{w_i^{(1)} O_{i+}}{n^*}$, and

Stage 3 probability of selection is equal to

　　　1 for the original certainty occupations and

Exhibit B-7. General Employer Sample, Stage 4 Sample Selection Summary

Sampling Frame	Establishments
Stratification	Establishment by occupation (SOC)
Selection Methodology	Employees will be systematically selected from within each occupation of each establishment.
Sample Size	For occupations selected with certainty, $\left(\dfrac{f_j \widetilde{N}_{ij}}{O_{i+}} \cdot n^* \right)$ employees will be selected. For the noncertainty occupations selected PPS without replacement at Stage 3, $\left(\dfrac{f_j \widetilde{N}_{ij}}{O_{ij}} \right) \left(\dfrac{\sum_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'}}{O_{i+}} \cdot n^* \right) \left(\dfrac{1}{\overline{J}_{Noncert,i}} \right)$ employees will be selected. Here $\widetilde{N}_{ij}$ is the estimated number of employees in establishment i, occupation j obtained from the POC. These occupation sample sizes may be scaled up or down in order to maximize the number of employees selected in the establishment without exceeding 20 for the establishment and 8 for a single occupation within the establishment.

$$\frac{\overline{J}_{Noncert,i} O_{ij}}{\sum_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'}} \text{ for the noncertainty occupations.}$$

Then for occupations selected with certainty, the unconditional probability of selection for each employee is equal to the product of the probabilities of selection at each stage, or:

$$\frac{1}{w_i^{(1)}} \cdot \frac{w_i^{(1)} O_{i+}}{n^*} \cdot 1 \cdot \frac{\left(\dfrac{f_j \widetilde{N}_{ij}}{O_{i+}} \cdot n^* \right)}{\widetilde{N}_{ij}}$$

$$= f_j$$

And for occupations selected with noncertainty, the unconditional probability of selection for each employee is equal to:

$$\frac{1}{w_i^{(1)}} \cdot \frac{w_i^{(1)} O_{i+}}{n^*} \cdot \frac{\bar{J}_{Noncert,i} O_{ij}}{\sum\limits_{\substack{j=1 \\ \text{Noncertainty}}}^{J_i} O_{ij}} \cdot \frac{\left(\frac{f_j \tilde{N}_{ij}}{O_{ij}}\right)\left(\frac{\sum\limits_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'}}{O_{i+}} \cdot n^*\right)\left(\frac{1}{\bar{J}_{Noncert,i}}\right)}{\tilde{N}_{ij}}$$

$$= f_j.$$

Additionally, if the observed number of employees in an occupation is equal to what is expected, in other words $\tilde{N}_{ij} \approx N_{ij}$, then the total number of employees selected in the certainty occupations are $\left(\frac{f_j N_{ij}}{O_{i+}} \cdot n^*\right)$, the total number of employee selections in the noncertainty

occupations are $\left(\frac{f_j N_{ij}}{O_{ij}}\right)\left(\frac{\sum\limits_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'}}{O_{i+}} \cdot n^*\right)\left(\frac{1}{\bar{J}_{Noncert,i}}\right)$, and these would sum to:

$$\sum_{\substack{Certa\,\text{int}\,y\,j}} \left(\frac{f_j N_{ij}}{O_{i+}} \cdot n^*\right) + \sum_{\substack{Noncerta\,\text{int}\,y\,j \\ Selected}} \left(\frac{f_j N_{ij}}{O_{ij}}\right)\left(\frac{\sum\limits_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'}}{O_{i+}} \cdot n^*\right)\left(\frac{1}{\bar{J}_{Noncert,i}}\right).$$

Since $\left(\frac{f_j N_{ij}}{O_{ij}}\right) = 1$, this would equal

$$\frac{n^*}{O_{i+}}\left(\sum_{\substack{Certa\,\text{int}\,y\,j}}(f_j N_{ij}) + \sum_{\substack{j'=1 \\ \text{Noncertainty}}}^{J_i} O_{ij'} \sum_{\substack{Noncerta\,\text{int}\,y\,j \\ Selected}}\left(\frac{1}{\bar{J}_{Noncert,i}}\right)\right).$$

Since there are $\bar{J}_{noncert,i}$ noncertainty occupations selected, the term

$$\sum_{\substack{Noncerta\,\text{int}\,y\,j \\ Selected}}\left(\frac{1}{\bar{J}_{Noncert,i}}\right)$$ equals 1 and the above equals

$$\frac{n^*}{O_{i+}}\left(\sum_{\substack{Certa\text{int }y\ j}}\left(f_j N_{ij}\right)+\sum_{\substack{j'=1\\ \text{Noncertainty}}}^{J_i} O_{ij'}\right)$$

$$=\frac{n^*}{O_{i+}}\left(\sum_{\substack{Certa\text{int }y\ j}} O_{ij}+\sum_{\substack{j'=1\\ \text{Noncertainty}}}^{J_i} O_{ij'}\right)$$

$$=\frac{n^*}{O_{i+}}\left(O_{i+}\right)$$

$$=n^*.$$

In summary, the above shows that if the frame estimates of employment by establishment and occupation are close to what is reported by the POC, then the within-establishment sample size will be close to the design parameter of n^*. Additionally, the unconditional sampling weights within each occupation and stratum will be approximately equal. Equalizing the sampling weights within occupations will result in more precise estimates produced from the data collection effort.

B.1.6 O*NET Association List Sampling Design

For selected occupations, respondents will be recruited from professional and trade association member lists. To be selected for O*NET data collection, an association must (a) represent the O*NET occupation in the nature of the work performed by its members, (b) contain a high percentage of the total occupational employment, and (c) be willing to provide a list of its members in usable form to be used as an O*NET sampling frame. Professional associations, licensing authorities, and commercial companies will be contacted for possible inclusion in the Association List Sample. This sampling methodology is appealing for several reasons:

- Higher response rates and higher eligibility rates, compared to the General Employer Sample, are anticipated.

- In general, most associations will have membership lists available at the individual member level. Therefore, respondents can be sampled using a comparatively simple, single stage design.

- Members will be contacted directly, thereby eliminating any gatekeeper effect associated with the establishment contact.

The sample selection procedures will vary across associations, depending on the type of information available on association members. In general, association lists will be sampled using a single stage, stratified simple random sampling approach.

For some occupations, obtaining the full desired respondent sample size by questionnaire type from appropriate associations may be the most efficient approach, particularly if an association covers a high proportion of an occupation. However, in most cases, the Association List Sample will be used in conjunction with the General Employer Sample to complete an occupation.[6] In order to take full advantage of the single-stage sample selection techniques associated with the Association List Sample while simultaneously ensuring adequate coverage of the occupation using the General Employer Sample, a dual-frame sampling approach will be used. For these occupations, a proportion of the sample will be allocated and selected using the Association List Sample methodology and the remaining sample will be allocated and selected using the General Employer Sample methodology. A question about membership in the association is included in the survey instruments for those employees selected from establishments using the General Employer Sample methodology. Similarly, a question will be added to the survey instrument for those employees selected from associations asking about their current work status (currently working, retired, unemployed, etc.). This information will allow computation of the joint probability of being selected from both the Association List Sample and the General Employer Sample. This joint probability will be used to adjust the sampling weights for employees who could have been sampled from each frame.

B.1.7 Collecting Data from Subject Matter Experts (SMEs)

SMEs will be used to obtain the data for a very small number of occupations in which the General Employer Sample and Association List Sample methodologies are problematic. Examples include occupations with very small employment, new and emerging occupations that do not yet have industry employment data, and those whose incumbents are in remote locations that are difficult to access. For those occupations included in the General Employer Sample, an occupation will be assigned to the SME method upon evidence of substantial underperformance by the General Employer and Association List methods (i.e., low eligibility, poor representation by available association samples). This method shall therefore be viewed as an infrequent alternative to establishment-based and association-based data collection. In the case of new and emerging occupations, the SME methodology will be the preferred method for data collection.

---

[6] Recall that an occupation is considered complete for the O*NET *Data Collection Program* when 15 respondents are obtained for each of four questionnaire types.

To decide which sampling method to use for each occupation, O*NET staff will compare the advantages and disadvantages of each potential method. For each of the occupations, information on the predicted eligibility rate and the predicted response rates will be used to quantify the efficiency of sampling the occupation through a General Employer Sample. The projected coverage afforded by association lists of employees, when available, will be used to quantify the potential benefit of sampling from such lists. Since a sample represents only the portion of the universe covered by the sampling frame, it is usually not worthwhile to use frames that cover a small part of the population of interest. For this reason, lists covering less than 50% of the employment for an occupation will not normally be sampled in O*NET surveys.

B.1.8 Weighting and Standard Errors

Most estimates generated from the resulting O*NET data will be computed using the appropriate sampling weights in order to reflect the complex nature of the sampling design. The basic sampling weights will be computed as the inverse of the unconditional selection probabilities for each selected establishment and each individual participant. The analysis weights for the eligible sample units will be adjusted to compensate for nonresponse, for both establishments and employees. Weight adjustments or imputations to compensate for item nonresponse within completed instruments are not planned.

Unit nonresponse adjustments to the sampling weights will be computed using a model-based approach described most recently in Folsom and Singh (2000). The modeling approach has been used in recent years to generate nonresponse adjustments because of the potential increase in bias reduction that can be achieved over the commonly used weighting class approach. This increase in bias reduction results from the modeling approach, which allows use of more significant main effect and lower-order interactions terms in the adjustment, compared to what can be used with a weighting class approach.[7] This is particularly appealing for the sampling weights generated for O*NET employee respondents, since the respondent sample sizes within each occupation can be very small (minimum 60 respondents). The Folsom and Singh modeling approach is based on a simple generalization of constrained models first suggested by Deville & Särndal (1992). These models allow the user to impose pre-determined constraints on the resulting model-based adjustment in order to minimize the effect the weight adjustment has on variance. This bias and variance reduction property of the weight adjustment, coupled with the relatively minimal amount of time and resources that are needed to make the adjustment, makes this approach particularly appealing for the O*NET sampling weights.

---

[7] If the response propensity model contains all main effect and interaction terms for a set of categorical variables, then the modeling approach to deriving weighting adjustments is equivalent to the weighting class approach. Consequently, one may also view the modeling approach as a generalization of the weighting class approach.

*O*NET Data Collection Program*

If applicable, sample weights will be corrected for employees who were selected from a corporation instead of just the physical location of the establishment that was selected. In other words, some establishments may keep their employee/occupation data at the corporate level, and it would be difficult for the POC to identify those employees who work only at the physical location of the sampled establishment. In these cases, the BL will ask the POC to sample from the POC's corporate records, and the sample weights of the resulting selected employees will be appropriately adjusted for the frame duplicity.

Weights also will be adjusted to correct for duplication between the general and association frames (dual frame samples) for applicable occupations. A similar weight adjustment will be applied to the wave-specific samples for occupations targeted in several waves. If the sampling weights were not adjusted, then some portion of the occupation's population would be counted more than once in the pooled sample. A standard dual frame adjustment that reweights the sampling weights by their sample sizes will be applied to that portion of the sample represented multiple times in the pooled set of respondents. This method of adjusting estimates produces the minimum variance on the pooled estimate, provided the standard error components between the samples are equivalent.

After the nonresponse and dual-frame adjustment to the sample weights have been applied, a final post-stratification weight adjustment will be created so that weight sums by occupation agree with other federal data sources, specifically the OES survey. The post-strata control totals will be adjusted using the coverage rates estimated from the sample frame. For example, if the target population for an occupation covers 80% of an occupation, then the control total obtained from the OES will be multiplied by 0.80.

All estimates generated from the O*NET database will also have accompanying estimates of the standard errors, in order to give users of the data some measure of the precision of the questionnaire estimates. Standard errors will be computed using the first-order Taylor Series approximation of deviations of estimates from their expected values. These standard error estimates will account for the basic properties of the complex O*NET sampling design, including stratification, clustering, and the often variance-increasing effects of unequal weighting.

B.1.9 Expected Response Rates and Sample Yields

The O*NET *Data Collection Program* recognizes the importance of obtaining high response rates to control nonsampling errors and minimize nonresponse bias. One of the lessons learned from prior O*NET field experiences, noted by Peterson, Mumford, Levin, et al. (1997), was the difficulty of obtaining high response rates in the O*NET field survey application. The

response rates obtained in the earlier O*NET field surveys, especially the employee response rate, were quite low. Even though establishments had agreed to cooperate, it appears that many of them ceased to cooperate when they received the survey materials to distribute to the sampled employees. Completed, usable questionnaires were returned for only 16.1% of the eligible employees sampled in the American Institutes for Research (AIR) baseline survey (Peterson, Mumford, Levin, et al., p. 2-65); a response rate this low has the potential to seriously compromise the parameter estimates.

The eligibility and response rates experienced in the 1999 O*NET pretest were also somewhat lower than desired. However, the primary purpose of the pretest was to experimentally determine those survey design factors that would maximize O*NET response rates. The *O*NET Report: Results of Statistical Analysis of Pretest* (see **Appendix A**) concluded that, had the "best design" been employed in the entire pretest sample, a 69% employer response rate and 55% employee response rate would have been obtained.

As shown in **Exhibit B-8,** final response rates of 64% were observed for employers and 63% were observed for employees in Wave 1.1. These response rates are assumed for future waves, for purposes of estimating sample sizes and yields. While a 63% employee response rate is sufficient to meet O*NET information needs, a number of enhancements to the survey procedures have been identified and developed as a result of experience gained in Wave 1.1. These improvements are anticipated to lead to higher response rates in future waves. These methods are described in **Section B.3**. The impact of these enhancements will be continually monitored, and further refinements will be made through a process of continuous improvement, so that an expectation of achieving higher response rates for both employers and employees is reasonable. As these higher response rates are achieved, the sample design assumptions will be adjusted accordingly for subsequent waves.

The eligibility and response rates presented in **Exhibit B-8** are presented by survey phase, reflecting contact with the establishment POC (Verification, Screening, Recruiting, and Sampling survey phases) and response/eligibility at the employee level. These various phases of data collection are discussed in greater detail in the next section. These eligibility and response rates presented in **Exhibit B-8** were computed using the following definitions:

- Establishment eligibility rates at each survey phase were computed as the total number of eligible establishments at the conclusion of the phase, divided by the total number of establishments that started the phase. An establishment is considered eligible for the data collection effort provided it is in operation, has at least one employee, is located at the physical location address as listed on the sample frame, and its primary activity agrees with the sample frame SIC. Due to the subsetting

Exhibit B-8. Response and Eligibility Rates from Wave 1.1 Data Collection Effort and Projections for Future Waves

Survey Phase	Wave 1.1 Results	Projected for Waves 2.2–16.4
Establishment Sample		
Verification Stage[a]		
Establishment Eligibility Rate	n/a	90.0%
Establishment Response Rate	n/a	97.0%
Screening Stage		
Establishment Eligibility Rate	87.3%	97.0%
Establishment Response Rate	95.0%	97.9%
Recruiting Stage		
Establishment Eligibility Rate	97.2%	97.2%
Establishment Response Rate	75.1%	75.1%
Sampling Stage		
Establishment Eligibility Rate	99.8%	99.8%
Establishment Response Rate	90.5%	90.5%
Total Establishment Rates		
Eligibility Rate	84.8%	84.8%
Response Rate	64.3%	64.3%
Employee Sample		
Employee Response Rate	62.6%	62.6%
Employee Completion Rate	90.0%	90.0%

[a] Beginning in Wave 1.2, the Wave 1.1 screening stage will be split into a verification stage (a very brief conversation made with the first employee contacted at an establishment) and screening stage (a more thorough introduction of the data collection effort with someone in the establishment more knowledgeable about the establishment's occupations). The verification contact may or may not be the same employee as the screening contact.

nature of the survey phases, by definition, those establishments that start any phase of data collection are equivalent to the respondents at the previous stage. So, for example, Verification respondents move forward to the Screening phase of data collection, and therefore make up the denominator of the eligibility rate for Screening.

▪ Establishment response rates at each survey phase were computed as the total number of responding establishments at the conclusion of the phase divided by the total number of eligible establishments at the conclusion of the phase. Note that the denominator of the response rate at a phase is equivalent to the numerator of the eligibility rate at the same phase. An establishment is considered a respondent if the POC responds to all critical questions asked. If an establishment agrees to participate and does not have employees in any of the occupations inquired about, then the establishment is considered a respondent at every establishment phase of data

collection. For each occupation, a selected employee is considered a respondent if the employee responds to any part of the questionnaire, either the web instrument or the paper questionnaire form. A questionnaire is considered complete and usable, provided the responses indicate the respondent matches the profile of the occupation sampled for and provided at least 50% of the questionnaire items have been answered. Given these definitions, the employee response rate is defined as equal to the total number responding employees, divided by the total number of selected employees.

- The employee completion rate is equal to the total number of completed employees divided by the total number of employee respondents.

B.2 Procedures for the Collection of Information

Data collection activities are conducted from the O*NET Operations Center under the management of RTI. BLs are responsible for recruiting sampled establishments to participate in data collection. The data collection procedures were first tested in the O*NET pretest and have been modified as a part of ongoing continuous improvement efforts.
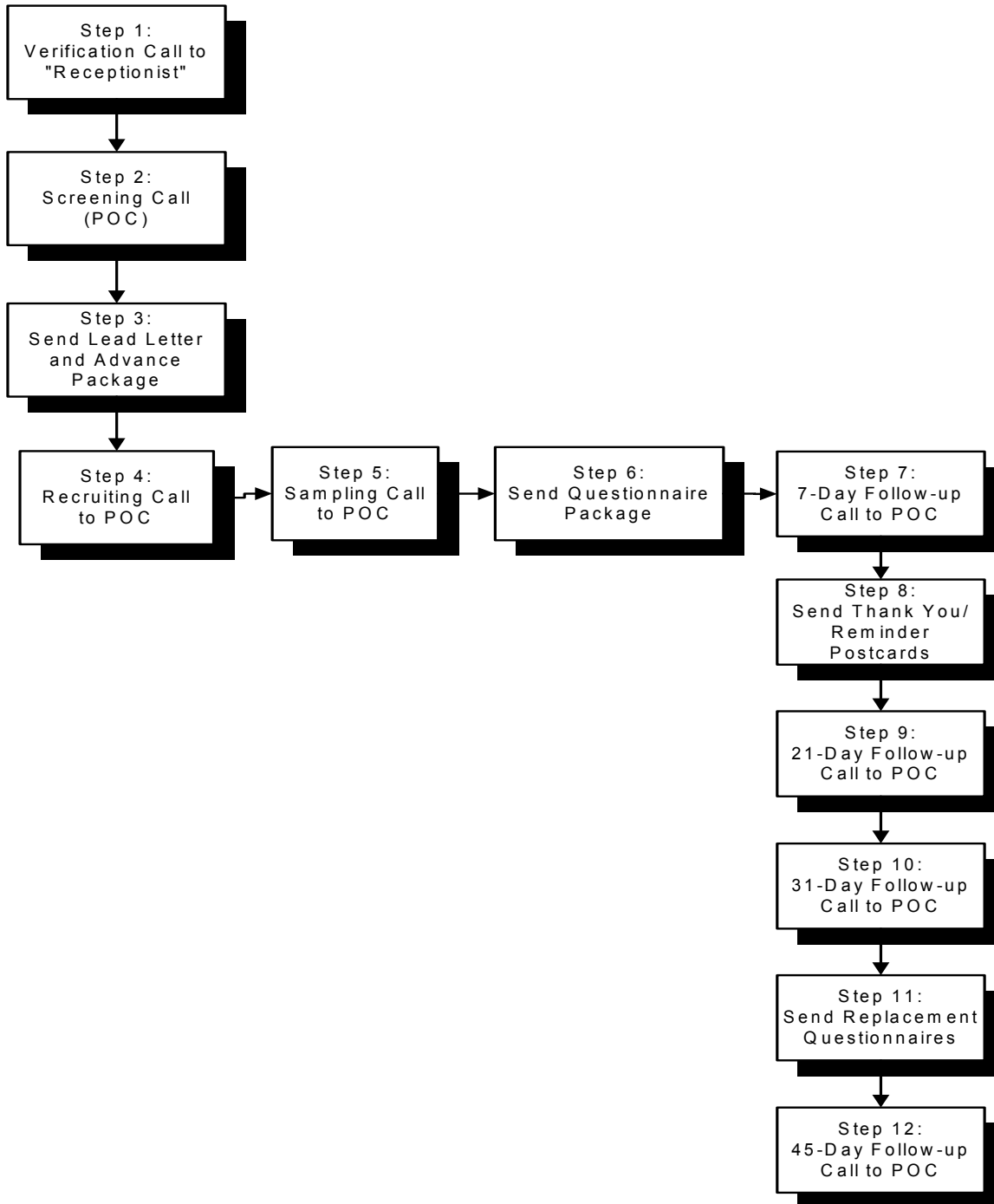
Exhibit B-9 shows a flowchart of the establishment contacts designed to collect information from O*NET participants.

B.2.1 The Establishment Recruiting Staff

The BLs interact with establishments during all steps of the data collection process to secure and maintain cooperation. An Operations Center Manager has oversight of all BLs. A core group of experienced BLs assists operations by acting as coaches for new BLs. New BLs are well qualified and thoroughly trained to communicate effectively with participating establishments. The BLs' communications with establishments is not scripted, but rather, the BL treats each establishment individually. Only one BL is assigned to each establishment.

A sophisticated, user-friendly, web-based computer control system has been developed to track all establishment information gathered by the BL, including the current disposition of each sampled establishment. This information will be used to prompt future events, such as making follow-up calls to the establishment POC, generating personalized letters and occupational profiles, and producing program status reports.

Exhibit B-9. Information Collection Flowchart

Step 1:
Verification Call to
"Receptionist"

Step 2:
Screening Call
(POC)

Step 3:
Send Lead Letter
and Advance
Package

Step 4:
Recruiting Call
to POC

Step 5:
Sampling Call
to POC

Step 6:
Send Questionnaire
Package

Step 7:
7-Day Follow-up
Call to POC

Step 8:
Send Thank You/
Reminder
Postcards

Step 9:
21-Day Follow-up
Call to POC

Step 10:
31-Day Follow-up
Call to POC

Step 11:
Send Replacement
Questionnaires

Step 12:
45-Day Follow-up
Call to POC

Packaged scripts are not used during any of the telephone calls to establishments. Guidelines are used to outline the necessary information to be obtained from each establishment. BLs are trained to use their skills and abilities to listen and interact effectively with the POC. All information gathered is entered into the control system by each BL after each conversation with the POC.

B.2.2 The Verification Telephone Call to Sampled Establishments

The BL calls each sampled establishment to verify the establishment's eligibility by verifying the establishment's address and type of work (***Exhibit B-9***, Step 1). If it has a different name but does the same work at the same location, it is considered eligible. The verification call is generally completed with the person who answers the phone at the establishment. During the call, the BL:

- explains the reason for the call;

- obtains verification of address and number of employees;

- obtains the name and position of the verification informant;

- obtains the name of a knowledgeable POC candidate; and

- requests the verification informant to transfer the call to the POC candidate.

B.2.3 The Screening Call to Determine Standard Occupational Classification (SOC) Eligibility

The BL makes a screening call to contact a knowledgeable POC and determine whether the establishment has any of the occupations sampled for that establishment (***Exhibit B-9***, Step 2). Specifically, the BL:

- explains the reason for the call;

- verifies that the person is an appropriate POC; if not, asks to be referred to such a person;

- confirms presence of target occupations (in establishment up to five "yes" responses);

- describes the advance package; and

- explains the establishment incentive (O*NET Toolkit for Business).

B.2.4 Advance Package of Information Sent to the POC

After the screening call, an advance package is sent to the POC by U.S. priority mail and contains the following:

- letter from the Department of Labor;

- O*NET brochure;

- project information sheet pamphlet;

- selected occupations list;

- initial gift to the POC as an incentive to read the advance package (clock);

- O*NET list of association endorsements; and

- information about the POC incentive options if the POC decides to participate (see *Section A.9*).

Examples of the information listed above are found in *Appendix G*.

B.2.5 Recruitment Call to the POC

The BL confirms that the package has been delivered using the U.S. Postal Service website and the package tracking number. After this, the BL allows 2 to 3 days for that POC to receive the package in interoffice mail and review the materials. Then the BL calls the POC to recruit them into the study (*Exhibit B-9*, Step 4). At this stage, the BL secures:

- confirmation of occupational match for one or more target occupations;

- a commitment to participate;

- agreement to prepare a roster of employees for each non-zero occupation;

- the POC's choice of incentive options;

- an exchange of email addresses; and

- an appointment for the next contact (sampling call).

When asking the POC about the presence of target occupations, the BL refers the POC to the occupation definitions on the selected occupations list that was included in the advance package. If the POC does not have the list handy, the BL faxes or e-mails the list to the POC.

The BL cannot proceed to ask about the target occupations until the POC has the definitions of the target occupations available for reference.

If the POC is still uncertain about the presence of an occupation after reviewing the occupation definition, the BL refers the POC to the ID Profile for the occupation. This document includes typical tasks associated with the occupation as well as alternate job titles and exclusionary titles (i.e., titles that relate to another occupation). The ID Profiles are accessible through the project website. If the POC does not have web access, the BL can fax or e-mail the ID Profile to the POC for the occupation in question.

Once the presence of target occupations is confirmed, the BL asks the POC to prepare a roster of employees for each occupation. The BL asks the POC to include all employees in the occupation at that establishment location, including employees who are:

- telecommuters or field agents who are administratively assigned to the sampled establishment location, even though they are not physically located there.

- on vacation, sabbatical, leave, or travel status.

- on temporary assignment at another location.

- new to the company.

- not direct employees of the establishment (e.g., consultants, temporary employees, or contract employees).

- activated military reservists who are temporarily on military duty.

The only employees who are ineligible are those who are retiring or leaving the company permanently within the next 60 days, and active-duty military employees at a military facility.

The BLs are trained to work with POCs to determine whether or not the POC's establishment employs any incumbents in the target occupations. In order to make these determinations, BLs are trained to use a target occupation's description and tasks to help the POC decide whether any employees' jobs in their establishment are a good match with the target occupation. To facilitate this process, the BLs have access to online Identification Profiles for each occupation, which can be e-mailed or faxed to POCs if necessary. These Identification Profiles display key information about the occupations, including the occupational code, title, description, tasks, alternate titles for the target occupation, and occupation titles that are not to be considered equivalent to the target occupation.

The BLs are thoroughly trained to answer questions, overcome barriers to participation, and convince the POC of the importance of the O*NET Program and of the POC's participation. Gaining the participation of the establishment and POC may take several calls. The BL works around the POC's schedule and moves at a pace that is comfortable for the POC. Ultimately, however, if the POC refuses to participate, the BL thanks the person for his/her time and terminates the call amicably. The case is then referred to a specially trained "converter" BL, who calls the POC back after several days to attempt to secure the person's participation. If this attempt is unsuccessful, the BL asks the POC if there is another person in the establishment that can be contacted to serve in the POC role.

B.2.6 Sampling Call to the POC

After a successful recruiting call, the next contact with the POC is the sampling call (*Exhibit B-9*, Step 5). Prior to this call, the POC has compiled a list of employees' names for each identified occupation (this step may not be necessary in smaller establishments). The POC tells the BL how many names are in each list. The BL enters the numbers into the control system, which uses an algorithm to select line numbers. The line numbers represent employees to be sampled. The POC notes the line numbers on his/her lists. The POC has sole possession of the lists and is asked to retain the lists to assist in the distribution and follow-up of questionnaires.

B.2.7 Mailing the Questionnaire Package

Following the sampling call, a questionnaire package containing an individually sealed and labeled survey packet for each sampled employee is sent to the POC by priority mail for distribution. The POC is instructed to link the line number found on each packet label to the occupation lists that the POC has maintained. Each individual survey packet contains a letter from the RTI project director, the assigned questionnaire, a return envelope, an information sheet for completing the questionnaire, and a $10 cash incentive. Examples of the letters and the information sheet are provided in *Appendix G*. The questionnaires are provided in *Appendix B*.

Also included in the questionnaire package is a cover letter to the POC and the cash incentive that the POC selected during the earlier recruiting or sampling call. The O*NET Toolkit is sent to the POC as an additional incentive in a separate package after the questionnaire mailing so as to not confuse the POC into thinking the Toolkit is to be used in the data collection. The POC is asked to communicate to all selected employees, in whatever manner works best, that the company supports the O*NET *Data Collection Program* and encourages their voluntary participation.

If the POC prefers to provide the names of employees and the employees can receive mail at work, the survey packet is shipped directly to each selected employee. This is based on the POC's assurance that internal policy permits the release of selected employee names and that they may receive mail at the company. In this situation the POC still receives the POC incentive. However, no name information will appear on returned questionnaires.

B.2.8 Employee Participation in the O*NET *Data Collection Program*

The selected employees are encouraged to complete the questionnaire on their own time, not company time. On average, the questionnaire takes 30 minutes to complete. Each selected employee is asked to return the completed questionnaire directly to RTI within 2 weeks.

Sampled employees have the option of completing an online version of the assigned questionnaire by accessing the O*NET data collection web page. This is discussed in detail in *Section A.3*.

B.2.9 Follow-Up with the Participating POCs

The BLs conduct a series of follow-up contacts with POCs at establishments with sampled employees (*Exhibit B-9*, Steps 7-12). These contacts are critical to maximizing employee response rates, as well as supporting the ongoing efforts of the POC. Following are details regarding each of the (up to) four follow-up calls to participating establishments.

7-Day Call

The BL calls each of his/her assigned participating POCs 7 calendar days after the shipment of the survey packets, regardless of whether they were shipped to the sampled employee or to the POC. The call verifies that the questionnaires arrived and have been distributed to employees. Soon after this follow-up call, thank you/reminder postcards are shipped to the POC, which he/she distributes to each sampled employee. The postcard includes both a thank you for those employees who have already returned the survey and a request to fill out and return the questionnaire for those who have not yet done so. The cover letter asks the POC to distribute the postcards to the selected employees.

21-Day Call

The BL calls the POC 21 calendar days after the questionnaire shipment date to ask the POC to distribute the reminder postcards.

31-Day Call

Thirty-one days after the initial mailing, the BL calls the POC to update the POC on the status of questionnaire return activity for the POC's establishment. At this time, the BL informs the POC that replacement questionnaires have been sent for any nonresponding employees and the BL secures the POC's cooperation in distributing these. The BL also introduces an email-based memo, which the POC can use when distributing the replacement questionnaires.

45-Day Call

The BL calls the POC again 45 days after the initial mailing of packets. The BL thanks the POC for his or her participation and provides an update of the status of employee participation. The BL confirms both receipt and distribution of the replacement questionnaires.

Additional Calls

In addition to the scheduled calls described above, other contacts are made, as necessary and appropriate. These include refusal conversion calls to "gatekeeper" POCs who do not follow through with the distribution of the employee questionnaire packets.

B.2.10   Post-Collection Data Processing

Approximately 8 weeks is allowed following the shipment of replacement survey packets for all completed questionnaires to be received at the data collection contractor. All questionnaires received by that time are processed (edited and keyed) and the data file is developed for later analysis. All questionnaires are stored in a locked document control area at the data collection contractor until federal authorization to destroy them is issued. In the future, questionnaires will be optically scannable.

B.2.11   Data Collection by Association Lists

For a small number of the occupations, a more direct approach to reach survey respondents will be through professional associations. This may be in addition to the employer-based approach or in place of the employer-based approach, as was discussed in *Section B.1.1*. Associations will be offered the same incentives as establishments. As yet, this methodology has not been implemented.

Prior to sample selection, a predicted eligibility model will be used to identify occupations for which the general employer sample method is likely to produce low eligibility

rates. These occupations will be examined to determine if there are associations whose membership includes a significant portion of the occupation's job incumbents. For such occupations, either an association list frame or a dual-frame sample (involving either a general employer sample or a targeted employer sample) will be used, and the cooperation of the association(s) will be sought for selecting a sample of their members.

In the process of obtaining professional and trade association endorsements, some associations may be identified whose membership represents a high coverage of the employment for an occupation. In these cases, the cooperation of the association(s) will be sought for selecting a sample of their members. If cooperation is achieved, data collection for the occupation may be done using only the sample of association members, depending on the eligibility estimates for that occupation, or a dual-frame approach may be used, as described above. Once the associations are identified through rigorous guidelines (listed below), they will be asked to assist with data collection by making their membership lists available to the O*NET Program for distributed surveys.

Research has been conducted to determine appropriate associations to target for data collection. The focus will be on those associations whose members comprise a large percentage of workers in the occupation, whose membership represents a broad range of work performed within the occupation, and who have the capacity to survey their membership.

The following criteria will be used for evaluating the potential utility of association lists as sampling frames for particular occupations:

- The professional association's membership represents a substantial percentage of the nation's total workforce for that occupation.

- Adequate information about the members must be made available—at a minimum, name, address, and occupation. Ideally, information offered by the association would include telephone number, employment status, age, gender, race, and ethnicity, for use with lists that contain retired persons or individuals from more than one O*NET occupation.

- The professional association is either willing to make its membership list available for sampling purposes or will select a sample for the study with assistance. This factor is critical.

For occupations where association members are surveyed, the questionnaire will be modified to include questions concerning the respondent's employment status. Respondents who are not employed are ineligible for the survey and will be asked to return the questionnaire without completing the occupational questions. Also, where the dual-frame sample is used,

employees sampled through contacts with employers will be asked whether they are members of the associations for which members are sampled. This information will be used to compute correct sampling weights, as employees who are association members potentially could be selected in both sampling frames.

B.2.12   Subject Matter Experts (SMEs): Additional Source for O*NET Data Collection

SMEs will serve as the source of data for a small percentage of occupations, including (1) those with an extremely small population of job incumbents (e.g., with fewer than 5,000 job incumbents in the national economy, such as Marine Architects [approximately 1,000]); (2) those without industry and employment data (e.g., new and emerging occupations, such as Website Developers, not currently included within the SOC); and (3) those whose incumbents are determined to be highly inaccessible (e.g., Ordinary Seamen and Marine Oilers). Professional associations, academic/educational organizations, labor unions, licensing bodies, and other professional groups will be used to identify SMEs representing selected occupations, resulting in a convenience sampling frame from which approximately 28 experts will be selected. Each participating expert will complete the four O*NET questionnaires (Generalized Work Activities, Knowledge areas, Skills, and Work Context factors). The goal will be to collect data from at least 21 experts.

Identifying SMEs

Candidate SMEs will be sampled either from professional associations or from educational institutions offering instructional programs in the target O*NET occupation. These potential SME source organizations will be chosen based on representativeness in relation to the occupation of interest to ensure data quality and to provide an adequate sampling frame. Each referral source will be asked to compile a list of qualified candidates from within its membership or employment, from which potential SMEs will be randomly selected. Judgment of SME qualification will be based on the candidate's experience within the occupation, including capacities in which he or she has served (e.g., incumbent, supervisor, trainer). Occupation-related education and member classification within a professional association or group will also be considered, to the extent that it informs qualification to respond for the target occupation. Emphasis will be placed on representativeness of the sample, according to key variables identified by the referral source as potentially impacting the ratings for an occupation (e.g. region, industry, company size). The sample size of 28 assumes approximately an 70% response rate from the experts, yielding at least 20 complete responses.

Contacting SMEs

O*NET BLs will contact sampled SMEs by telephone. In certain cases, a professional association/group may assist with the contact effort. The initial contact will include an overview of the program and a confirmation that the nominated SME meets the eligibility criteria described above. Upon such confirmation, the BL will discuss the candidate's role in the program (e.g., serving as a national expert, completing a series of questionnaires). When the SME agrees to participate, the survey schedule will be discussed (e.g., date they will receive packets, expected turnaround time). The BL will also provide a contact to assist with any questions or concerns the respondent may have. Candidates who defer the decision to participate pending additional information will be recontacted at a later date, once they have had an opportunity to review the additional information.

Administration of Questionnaire Packets

Each SME will receive a set of four O*NET domain questionnaires to complete (Generalized Work Activities, Knowledges, Skills, and Work Context), along with a background questionnaire and a task questionnaire. Order of questionnaires in the packet will be randomly varied to control for order effects. The questionnaire package, sent to SMEs via overnight mail, will include additional program summary information, general instructions for completing the questionnaires, and return materials. Each SME will receive a follow-up phone call 1 week following package mail-out to confirm package arrival. Additional follow-up phone calls may be made at the discretion of the BL and the operations supervisor.

SME Incentive

Each participating SME will be offered a prepaid incentive of $10 per questionnaire. Commensurate with the incumbent incentive plan, it allows for a total incentive of $40 per SME, as each expert will be asked to complete all four domain questionnaires.

B.3   Methods to Maximize Response Rates

The O*NET *Data Collection Program* is designed to maximize the response rates through the method of continuous improvement. The goal of the continuous improvement method is to reduce nonresponse by continuously updating and improving the process using information gathered about the source of nonresponse during the data collection process. This information is used to evaluate the effectiveness of the current procedures and to guide enhancements. Three sources of information have been used: the results of the 1999 O*NET pretest, problems encountered during Wave 1.1, and a series of focus groups conducted with the BLs during the Wave 1.1 data collection process. The primary findings and implications for

enhancements to the data collection process for Wave 2.2 and future waves are summarized below. See *Section B.2* for a complete discussion of the data collection procedures.

1999 O*NET Pretest

1. The National O*NET Consortium conducted an outreach campaign targeting more than 150 associations representing the first 50 preselected occupations to be tested. The campaign was a great success, returning 40 endorsements from key associations, including the "umbrella" organizations such as the National Alliance of Business and the U.S. Chamber of Commerce. The National O*NET Consortium continued the outreach campaign through Wave 1.1 data collection to further increase awareness of the O*NET Program in the business and employer community. Along with awareness building, the Consortium strives to demonstrate the importance and possibilities of the O*NET Program, while securing meaningful endorsements from key industry and association leaders. The endorsement list appears in the advance package to the POC, described in *Section B.2.4*. The outreach campaign will continue in future waves.

2. The effectiveness of a noncash incentive in obtaining the cooperation of establishments in the O*NET *Data Collection Program* was demonstrated in the pretest. The O*NET Toolkit for Business will continue to be offered to participating establishments. The details of and rationale for this incentive are described in *Section A.9.1*.

3. Monetary incentives had a significant impact on employee response rates. The $10 prepaid cash incentive was demonstrated to be the most effective and efficient employee incentive and will be continued in future waves. The details of and rationale for this incentive are described in *Section A.9.4*.

Resolution of Problems Encountered During Wave 1.1 and Enhancements to Survey Process

1. Gatekeeper POCs had a negative impact on employee response rates by refusing to distribute the questionnaire materials to the selected employees. A team of experienced, successful BLs were trained as Gatekeeper Specialists to work the cases with reluctant POCs. This team will be maintained in future waves.

2. Materials were developed for training the BLs in Strategies and Tactics for Averting Refusals (STAR). Focus groups were conducted with the BLs to identify specific arguments and statements made by reluctant POCs that could lead to refusals. A training session was developed and administered with materials tailored to overcoming POC refusals. The STAR training sessions will be used in future waves.

3. Many questionnaires (514) were returned without the Task Questionnaire. This increased the rate of nonresponse because the Main Questionnaire data cannot be analyzed without the accompanying Task Questionnaire. To reduce this type of

nonresponse, the questionnaire booklet has been reformatted so both questionnaires are bound and formatted as one document.

4. Many of the establishments were willing to participate but were unable to yield completed employees surveys in the expected time frame. A 45-day follow-up call was added to the data collection procedure in order to provide the POCs with an additional opportunity to encourage the employees to respond. The additional call will be used in future waves.

Focus Groups with Business Liasons

1. The reminder postcard was sent to the POCs to be forwarded to the nonresponding employees in the late stages of data collection. The BLs reported in the focus groups that the timing of the postcard is not effective. This is consistent with the recommendations in the literature, which call for the reminder postcard to closely follow the initial mailing (Dillman, 2000). In future waves, the postcard will be sent to all POCs immediately following the 7-day call.

2. The BLs reported that POCs were often intimidated by the size of the advance package, stating that it was imposing and appeared to be excessively long. The advance package was streamlined to reduce the amount of paper the POC receives as well as the POC's perception of burden. The contents of the POC advance package are described in *Section B.2.4*.

3. The BLs reported that the POCs were often confused by the advance package because they received it without an advance contact by a BL. In addition, many of the advance packages were initially sent to the wrong contact at an establishment. For these reasons, the POCs will be contacted before the advance package is sent to verify that they are the appropriate contacts and to inform them of the contents of the package.

The Office of Management and Budget approved the current data collection effort on April 4, 2001 (OMB Number 1205-0421). The Terms of Clearance for the current effort required that the Department of Labor resubmit the collection for approval after 18 months with an analysis of response rates, response bias, effectiveness of incentives, and the use of first-class postage stamps. These issues are summarized for Wave 1.1 in *Appendix F*.

B.4   Tests of Procedures

A split-sample, random assignment experiment was used in Wave 1.1 to evaluate the relative benefits of using business reply envelopes or envelopes with first-class stamps. Preliminary results from the 1999 O*NET pretest demonstrated that there was no statistically significant difference between the two techniques. However, the BLs reported anecdotal evidence that the first-class stamps yielded shorter response times than the business reply envelopes. The results of the Wave 1.1 study support the pretest findings. The response rate for

employees assigned the business reply envelope is 52.2%, compared to 50.3% with the first-class postage.[8] The average response time for the business reply envelopes was 48.1 days, compared to 47.7 days for the first-class postage, which is not a statistically significant difference.

These results are not consistent with the literature, which suggests higher response rates when using first-class stamps (Dillman, 2000). This is probably due to design differences between the O*NET survey and the standard mail survey, which calls for reminder postcard 1 week after the initial questionnaires are sent. Dillman argues that the effect of the stamp is to make the postcard reminder more effective because respondents are more likely to keep a stamped envelope on hand and throw out a business reply mail envelope. The Wave 1.1 design did not offer a fair test of Dillman's hypothesis, since the postcards were sent more than a week after the questionnaires were sent and were only sent to the respondents who had not returned the questionnaire. As stated above, the postcards will be sent after the 7-day call in all future waves.

No tests of procedures have been identified at this time for the O*NET *Data Collection Program*, but opportunities for evaluating methodological enhancements to the program will continue to be monitored.

B.5   Statistical Consultants

The individuals shown in ***Exhibit B-10*** have consulted on the statistical aspects of the O*NET *Data Collection Program* sample design and/or will be involved in the data collection and analysis. Those individuals, shown in sections 1 and 2 of the table, reviewed a near-final draft of the OMB Clearance Package, and the package has been revised based on their comments. No additional external review of the present package has been conducted by these individuals.

The DOL/ETA officials responsible for the O*NET *Data Collection Program* are Jim Woods (202-693-3641) and Pam Frugoli (202-693-3643).

---

[8] At the time this request for OMB clearance was prepared, employee data collection was still in progress for Wave 1.1. These results are preliminary and are based on questionnaire return rates as of February 15, 2002.

Exhibit B-10. Statistical Consultants

Name	Organization	Telephone Number
(1) Non-Federal Statisticians and Researchers		
Michael Campion	Purdue University	765-494-5909
John Campbell	University of Minnesota	612-625-9351
Janet Wall	Sage Solutions	240-683-5824
(2) Federal Government		
Alan Dorfman	Bureau of Labor Statistics	202-693-3641
Frederick Conrad	Bureau of Labor Statistics	202-691-7513
Sylvia Karman	Social Security Administration	410-965-7693
John Galvin	Bureau of Labor Statistics	202-691-6400
Michael Pilot	Bureau of Labor Statistics	202-691-5703
Michael Horrigan	Bureau of Labor Statistics	202-691-5701
(3) Data Collection/Analysis Contractors		
Michael Weeks	Research Triangle Institute	919-541-6026
Joe Eyerman	Research Triangle Institute	919-541-7139
Michael Witt	Research Triangle Institute	919-541-8346
Paul Biemer	Research Triangle Institute	919-541-1261
Robert Mason	Research Triangle Institute	919-848-8577
Babu Shah	Research Triangle Institute	919-541-6879
Qing Yao	Research Triangle Institute	919-316-3384
Sutapa Das	Research Triangle Institute	919-541-6129
Robert J. Morris	Research Triangle Institute	919-485-5619
Ye-Ying Cen	Research Triangle Institute	919-541-8711

References

Bureau of Labor Statistics (BLS). (2000). *Occupational and Employment Statistics (OES) Survey*. Retrieved April 2, 2002, from http://www.bls.gov/oes/oes_dl.htm.

Bureau of Labor Statistics (BLS). (2000, January). *Employment and Earnings*. Washington, DC: Government Printing Office.

Campion, M.A., Morgeson, F.P., & Mayfield, M.S. (1999). O*Net's theoretical contributions to job analysis research (pp. 297-304). In N.G. Peterson, M.D. Mumford, W.C. Borman, P.R. Jeanneret, & E.A. Fleishman (Eds.), *An Occupational Information System for the 21st Century: The Development of O*NET*. Washington, DC: American Psychological Association.

Chromy, J.R. (1979). Sequential sample selection methods. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 401–406.

Church, A.H. (1993). Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly*, *57*(1), 62–79.

Deville, J.C., & Särndal, C.E. (1992). Calibration estimation in survey sampling. *Journal of the American Statistical Association, 87,* 376–382.

Dillman, D. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: John Wiley & Sons.

Dillman, D. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York: John Wiley & Sons.

Dillman, D.A. (1991). The design and administration of mail surveys. *Annual Review of Sociology*, *17,* 225–249.

Fleishman, E.A., & Mumford, M.D. (1988). The ability requirement scales. In S. Gael (Ed.), *The Job Analysis Handbook for Business, Industry, and Government*. New York: Wiley.

Fleishman, E.A., & Mumford, M.D. (1991). Evaluating classifications of job behavior: A construct validation of the ability requirements scales. *Personnel Psychology*, *44*, 523–575.

Folsom, R.E., Jr., & Singh, A.C. (2000). A generalized exponential model of sampling weight calibration for extreme values, nonresponse and poststratification. *Proceedings of the American Statistical Association*, Section on Survey Research Methods, 598–603.

Folsom, R.E., Potter, F.J., & Williams, S.R. (1987). Notes on a composite size measure for self-weighting samples in multiple domains. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 792–796.

Fox, David R. (1989). Computer selection of size-biased samples. *The American Statistician*, *43*(3), 168–171.

Fuchs, M. (2002). Paper and pencil versus web survey: Measurement effects and mode differences. Paper presented at the annual conference of the American Association for Public Opinion Research, St. Petersburg Beach, FL.

Golmant, J. (1990). Correction to 'Computer selection of size-biased samples.' *The American Statistician, 44*(2), 194.

Hubbard, M., McCloy, R., Campbell, J., Nottingham, J., Lewis, P., Rivkin, D., & Levine, J. (2000, October). *Revision of O*NET Data Collection Instruments*. Raleigh, NC: National O*NET Consortium, Employment Security Commission. Retrieved March 25, 2002, from http://www.onetcenter.org/dl_files/Data_appnd.pdf.

Hubbard, R., & Little, E.L. (1988). Promised contributions to charity and mail survey responses: replication with extension. *Public Opinion Quarterly, 52*(2), 223–230.

Kulka, R.A. (1994). *The use of incentives to survey 'hard-to-reach' respondents: A brief review of empirical research and current practice*. Research Triangle Park: RTI.

Levy, D.G., Thie, H.J., Robbert, A.A., Naftel, S., Cannon, C., Ehrenberg, R., & Gershwin, M. (2001). *Characterizing the future defense workforce*. Santa Monica, CA: RAND, National Security Research Division.

McCloy, R.A., Waugh, G., Medsker, G., Wall, J., Rivkin, D., & Lewis, P. (1999). Determining the occupational reinforcer patterns for O*NET occupational units: Volumes I (Report) and II (Appendix). Raleigh, NC: National Center for O*NET Development, Employment Security Commission of North Carolina.

Mumford, M.D., Peterson, N.G., & Childs, R.A. (1997). Chapter 3: Basic and cross-functional skills: Evidence for the reliability and validity of the measures. In N.G. Peterson, M.D. Mumford, W.C. Borman, P.R. Jeanneret, E.A. Fleishman, & K.Y. Levin (Eds.), *O*NET Final Technical Report* (p. 3-1 to 3-36). Salt Lake City, UT: Utah Department of Workforce Services, through a contract with the American Institutes for Research.

National Research Council. (1999). *The Changing Nature of Work: Implications for Occupational Analysis*. Washington, DC: National Academy Press. 364 pp.

NSCAW Research Group, The. (in press). Methodological lessons from the national survey of child and adolescent well-being: The first three years of the USA's first national probability study of children and families investigated for abuse and neglect. *Children and Youth Services Review*.

Paxson, M.C., Dillman, D.A., & Tarnai, J. (1995). Improving response to business mail surveys. In B.G. Cox (Ed.), *Business Survey Methods*. New York: John Wiley & Sons.

Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A. (1995, September). *Development of Prototype Occupational Information Network (O*NET) Content Model*. Utah Department of Workforce Services, through a contract with the American Institutes for Research.

Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., & Fleishman, E.A., Campion, M.A., Levin, K.Y., Mayfield, M.S., Morgeson, F.P., Pearlman, K., Gowing, M.K., Lancaster, A., & Dye, D. (2001). Understanding work using the occupational information network (O*NET): Implications for practice and research. *Personnel Psychology*, *54*, 451–492.

Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., Fleishman, E.A., & Levin, K.Y. (Eds.) (1997, September). *O*NET Final Technical Report*. Salt Lake City, UT: Utah Department of Workforce Services, through a contract with the American Institutes for Research.

Peterson, N.G., Mumford, M.D., Levin, K.Y., Green, J., & Waksberg, J. (1997). Chapter 2: Research method: Development and field testing of the content model. In N.G. Peterson,

M.D. Mumford, W.C. Borman, P.R. Jeanneret, E.A. Fleishman, & K.Y. Levin (Eds.), *O*NET Final Technical Report* (pp. 2-1 to 2-55). Salt Lake City, UT: Utah Department of Workforce Services, through a contract with the American Institutes for Research.

Peterson, N.G., Mumford, M.D., Levin, K.Y., Green, J., & Waksberg, J. (1999). Research method: Development and field testing of the content model. In N.G. Peterson, M.D. Mumford, W.C. Borman, P.R., Jeanneret, & Fleishman, E.A. (Eds.), *An occupational information system for the 21st century: The development of O*NET*. Washington, DC: American Psychological Association.

Peterson, N.G., Owens-Kurtz, C., Hoffman, R.G., Arabian, J.M., & Whetzel, D.C. (1990). *Army synthetic validation project*. Alexandria, VA: U.S. Army Research Institute for the Behavioral Sciences.

Research Triangle Institute (RTI). (2000, October). *O*NET report: Results of statistical analysis of pretest*. Research Triangle Park: Author.

Robertson, D.H., & Bellenger, D.N. (1978). A new method of increasing mail survey responses: Contributions to charity. *Journal of Marketing Research, 15*, 632–633.

Rounds, J.B., Jr., Smith, T., Hubert, L., Lewis, P., & Rivkin, D. (1999). The development of occupational interest profiles for the O*NET: Volumes I (Report) and II (Appendices). Raleigh, NC: National Center for O*NET Development, Employment Security Commission of North Carolina.

Roth, P.L., & BeVier, C. (1998). Response rates in HRM/OB survey research: Norms and correlates, 1990–1994. *Journal of Management*, *24*(1): 97–117.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., & McGonagle, K. (1999). The effect of incentives on response rates in interviewer-mediated surveys. *Journal of Official Statistics, 15*(2), 217–230.

Tulp, D.R., Jr., Hoy, E., Kusch, G., & Cole, S. (1991). Nonresponse under mandatory vs. voluntary reporting in the 1989 survey of pollution abatement costs and experiments (PACE). *Proceedings of the American Statistical Association, Survey Research Methods Section,* 272–277.

U.S. Department of Labor (USDOL) (2001, October). *Quality Information—Informed Choices: Advancing the Workforce Information System, Secretary of Labor's Workforce Information Plan for Federal Fiscal Years 2002–2005*. Retrieved March 25, 2002, from http://www.workforceinfocouncil.org/plan.htm). Washington, DC: Author.

Watts, D.L. (1991). Correction to 'Computer selection of size-biased samples.' *The American Statistician, 45*(2), 172.

Williams, R. L., & Chromy, J.R. (1980). *SAS sample selection macros*. Paper presented at the SAS Users Group Annual Conference, Texas.

Worden, G., & Hoy, E. (1992). *Summary of nonresponse studies conducted by industry division, 1989–1991*, unpublished paper. Washington, DC: U.S. Bureau of the Census.