Appendix F: Statistical Review of Part B of the OMB Supporting Statement

(by Summit Consulting on behalf of the Bureau of Labor Statistics)

Independent Statistical Review: Paperwork Reduction Act (PRA) Submission: Part B

Evaluation Project: O*NET Data Collection Program

Project Summary:

The Occupational Information Network (O*NET) Data Collection Program populates and maintains a current database on the detailed characteristics of workers, occupations and skills. The O*NET Data Collection Program used a wide range of audiences, including individuals making career decisions, public agencies, educational institutions, and employers making staffing and training decisions. The O*NET Data Collection Program received initial Office of Management and Budget (OMB) clearance in 1999 for a pretest and four subsequent clearances have allowed data collection to continue without interruption since June, 2001. The Supporting Statement reviewed is a request and justification for a 3-year clearance to continue to collect data through May, 2015.

To obtain occupational information, the O*NET program surveys employees, randomly selected through a complex, multiple method data collection. The primary source of information for the database is a survey of establishments and sampled workers from within selected establishments, which is referred to as the "Establishment Method". The Establishment Method is a two stage design, wherein establishments are selected in the primary stage with probability proportional to the expected number of employed workers in the specific occupations being surveyed. In the second stage, a sample of workers is selected in the occupations within sampled businesses.

Supplement frames account for occupations that have less than expected coverage under the Establish Method. This situation may occur for new or rare occupations that are difficult to locate through Establishment Method sampling. Supplemental frames consist of either a listing of establishments highly likely to employ a particular occupation, or a listing of incumbents working in an occupation. If a supplemental frame can be obtained, the occupation is removed from the Establishment Method wave and sampled separately. One stage, stratified random samples are usually selected from supplemental frames. The supplemental frame weights are adjusted for nonresponse and combined with the Establishment Method weights. Sampling from supplemental frames is also referred to as the Occupational Expert (OE) Method.

Selected employees within an establishment are randomly assigned, with equal probability, to one of three questionnaire types concerning general work activities, work context, or knowledge. Although employees are randomly assigned to a questionnaire, analysis weights are computed for each occupation. As many of the questions are similar across the three questionnaires, the estimates for these questions are produced from all the respondents for the occupation. No sub-group

(questionnaire) estimates are produced or released. The final estimates are the mean of the Likert scales for each occupation or final percentages for items concerning work context, education and training, background items, or occupation tasks. The standard deviation will be available for each item mean as a measure of response variation among an occupation's respondents.

Analysis weights account for probabilities of establishment selection, probabilities of occupation selection, probabilities of employee selection, early termination of employee sampling activities for particular occupations because of higher-than-expected yields, multiple-sample adjustments (overlapping frames), establishment and employee non-response, and population coverage.

Submission documents reviewed by Summit:

To evaluate the adequacy of the proposed sampling methodology to meet the aforementioned objectives, we evaluated the following documents:

1. O*NET OMB Clearance Package Supporting Statement Section A
2. O*NET OMB Clearance Package Supporting Statement Section B
3. O*NET OMB Clearance Package Appendix E Non Response Analysis

General Comments:

The organization and presentation of the response was thoughtful and exhaustive. The stated survey methodology and econometric analysis appear to be adequate in addressing the evaluations objectives.

The Part B report presents a sample design that is consistent with industry practices, namely, a stratified two-stage sample in which the first stage selects establishments with probability proportional to size, and the second stage selects individuals from selected establishments in a simple random sample.

The supporting statement described steps addressing potential bias both from establishment and item non-responses. The analytical narratives both for the econometric modeling and propensity score weights are credible and considerate. Steps to minimize non-responses are consistent with industry standard practice. The data collection procedures gave careful consideration to minimize establishment burden and maximize response rates.

The complex weighting methodology was clearly presented with relevant citations, formulas, tables and figures where appropriate.

The formulas in this report are standard and correct. There is an error or unexplained omission in the footnote of page 77 within the OMB Supporting Statement Part B whereby no stratification category is included for establishments with 10-23 employees.

Recommendations:

The O*NET multiple method data collection and estimation approach is thorough and complete. No major errors exist within the statistical methodology, estimation process, or analytical framework.

Supporting Statement for Paperwork Reduction Act (PRA) Submission: Part B Review

This template provides a checklist identifying required documentation justifying the collection of information employing statistical methods (PRA Supporting Statement, Part B). Sample design components, processes, and disciplines are enumerated within sub-parts to questions (i.e., Question 1.1 corresponds to information under Part B Question 1). This checklist is the first step in Summit's review process, assuring the package meets the content requirements for submission, while providing a summary background of the design to inform further review. Upon receipt of a PRA package, the review team will summarize the statistical methodology by providing responses to the information requested below.

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
1.1 Project objective(s)	To continue the O*NET Data Collection program, which has populated and maintained a current database on the detailed characteristics of workers, occupations, and skills uninterrupted since June 2001. The current OMB clearance expires on May 31, 2012. This request will allow the continuation of data collection through May 2015.	The objectives of data collection are clearly described.
1.2 Description of sampling design or other respondent selection methods to be used.	This survey employs a multiple method data collection approach. The primary source of information for the database is a survey of establishments and sampled workers from within selected establishments, which is referred to as the "Establishment Method" (OMB P.B Pg. 73). The Establishment Method is a two stage design, wherein establishments are selected in the primary stage with probability proportional to the expected number of employed workers in the specific occupations being surveyed. In the second stage, a sample of workers is selected in the occupations within sampled businesses.	Given the vast array of occupations and industries covered by the O*NET program, this versatile multi stage sampling design appears to be adequate in capturing the desired information, minimizing burden, and maximizing efficiency. The design assumes that employees on the supplemental frame are equivalent to a random sample of all employees working in an occupation who might have been selected through the Establishment Method; this assumption appears sound and is appropriately addressed through weighting. The terminology is precise; the description of

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
	A dual frame approach is used to account for occupations that have less than expected coverage under the Establish Method. This situation may occur for new or rare occupations that are difficult to locate through Establishment Method sampling. Supplemental frames consist of either a listing of establishments highly likely to employ a particular occupation, or a listing of incumbents working in an occupation. If a supplemental frame can be obtained, the occupation is removed from the Establishment Method wave and sampled separately. One stage, stratified random samples are usually selected from supplemental frames. The supplemental frame weights are adjusted for nonresponse and combined with the Establishment Method weights.	the methodology is very clear, given the complex design. As a minor suggestion, usage of the OE Method and "supplemental frames" can be confusing as both are separate frames from the main Establishment Method. Suggest consistent references to "OE Method" in subsequent sections that refer to supplemental frames.
1.3 Description of the potential respondent universe.	The O*NET sampling universe for each occupation is generally a subset of all employees in the occupation who are working in the United States. This target population is defined by two criteria: (1) its workers represent a majority of job incumbents in the occupation, and (2) data among this set of establishments can be gathered with reasonable efficiency (OMB Pt. B Pg. 74).	The sampling frame maintains population coverage of at least 50% for each occupation, for efficiency and burden reasons (OMB P.B Pg. 74). A sensitivity study concluded that no systematic bias is introduced with the use of a population coverage minimum of 50%, compared to a minimum coverage of 80%. The study and supporting response adequately addresses coverage issues.
1.4 Sampling unit.	Within the Establishment Method, establishments are the primary sampling units	Given the O*NET objective of populating and maintaining a current database on the detailed

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
	and employees are the secondary sampling units. Employees are the primary sampling units in OE Method sampling.	characteristics of workers, occupations, and skills, these sampling units are appropriate.
1.5 Population frame: provide data on the number of entities (e.g., establishments, households, persons) in the universe covered by the collection and in the corresponding sample, for the universe as a whole, and for each strata (tabular format).	The population universe of entities is unknown.	The response provides substantial support describing data on the number of occupations and establishments in its purview. The Establishment Method frame is primarily populated by the Dun and Bradstreet national database of more than 15 million establishments (OMB Pt. A Pg. 9). The current incarnation of O*NET contains 1,122 occupations based on the Bureau of Labor Statistics (BLS) Occupational Employment Statistics (OES) codes.
1.6 Estimated sample size: provide data on the number of entities (e.g., establishments, households, persons) in the sample as a whole, and for each stratum (tabular format)	The estimated number of sampling units for the June 2012 – May 2013 year is almost 15,000 employees (OMB Pt. A Pg. 65). Under the Establishment Method, an occupation is included in the database once there are 15 valid completed questionnaires. Under the OE Method, an occupation is included in the database once there are 20 valid completed questionnaires.	While a tabular format was not used, the Establishment Method response includes numerous citations, supporting tables, and previous O*NET experience to show that a sample size of 15 questionnaires per occupation will achieve the desired level of precision to meet the needs of O*NET users (OMB Pt. B Pg. 86). The OE Method response states that 20 questionnaires enable adequate coverage of experts, occupation sub-specialties, and regional distribution. It is implicitly assumed that a total of 20 questionnaires are adequately defended given the breadth of the defense for a minimum of 15 Establishment Method questionnaires.

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
1.7 Indicate the expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rates achieved during the last collection.	Through December 31, 2010, the cumulative response rate for the Establishment Method is 76% for establishments and 65% for employees. The OE Method response rate through December 31, 2010 is 79.8%.	The response reports actual response rates achieved through December 31, 2010 for the Establishment Method and the OE Method. The response provides data collection results in tabular format, and an exhaustive summary of response rates through December 31, 2010 (OMB P.A Pg. 12).
2.1 Description of the statistical methodology for stratification.	Establishment Method: Occupation representation is the ultimate goal. The first stage sample of establishments is selected through stratification by industry grouping and establishment size. Establishments are randomly selected using a probability minimum replacement (PMR) selection procedure to ensure that establishments with the greatest likelihood of containing the desired occupations are oversampled. Employees are selected through simple random sample. Under the OE methods, the stratification differs dependent upon the supplemental frame. For establishment frames, the sample design is two stage. A stratified random sample of establishments is selected in the first stage, and employees are selected in the second stage. For a frame of incumbents or employee (experts), a stratified random sample of employees will be selected.	

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
2.2 Description of the sample selection methodology.	The Establishment Method process contains 8 steps, as described in OMB Pt. B Pgs. 77-84. The steps are enumerated in Appendix I. (Step 1) Create establishment sampling frame (Step 2) Determine industries to target for each occupation in sub-wave (Step 3) Select initial sample of establishments from frame (Step 4) Select final set of establishments for sub-wave (Step 5) Assign occupations to selected establishments (Step 6) Specify employee sample selection algorithm for each occupation/establishment (Step 7) Specify occupation-specific algorithm for random assignment of questionnaire types of sampled employees (Step 8) Randomly assign selected establishments to business liaisons for data collection Under the OE method, a frame of either establishments or incumbents working in an occupation (trade associations, etc.) will comprise the frame. In the case of establishments, a stratified random sample of establishments is selected in the first stage. The selected establishments are then included in step 6 of the Establishment	The O*NET data collection effort appears to be taking multiple adequate steps to ensure the completeness of the establishment frame while maximizing the efficiency of data collection. Step1: Establishments with 10-23 employees are not captured in the four categories discussed (OMB Pt. B Pg. 77). Step 4: The probability proportionate to composite size formula is standard and correct. The composite size measure elegantly accounts for the sampling several occupations from selected establishments. The narrative includes appropriate academic citations. Step 5: The proposed formulas appear to be satisfactory in achieving the desired results. Step 6: The methodology describes adequately controls for burden at the establishment level. Step 7: The presented method for randomly allocating questionnaires to respondents appears to be correct. The collection methodology is well described and adequately protects against bias (OMB Pt. B Pg. 82). Step 8: Administrative actions to protect against bias between the business liaison and establishments are appropriate and well thought out. The methodology presented to properly incorporate the Establishment Method and OE Method as presented is statistically accurate,

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
	Method selection above. If the supplemental frame contains individual incumbents, a simple random sample of incumbents is selected, which may be stratified by geographic region or subspecialty of occupation if the data exist.	thoughtful, and well supported through formula and text.
2.3 Description of the estimation procedure.	Estimates are calculated by dividing the sum of weighted observations by the sum of weights. (OMB Pt. A Pg. 11) Estimates are generated utilizing compensating sample weights for establishments, occupations within establishments, and employees within occupation. (OMB Pt. A Pg. 11) Estimates produced include scale means and percentage estimates whereby no subgroup or domain estimates are produced or released. Standard deviation measures are provided for each mean as a measure of response variation among an occupation's respondents. (OMB Pt. B Pg. 97)	
2.4 Description of the degree of accuracy (margin of error) required for the purpose described in the justification.	A variation of plus or minus 1 to 1.5 scale points is expected for all occupations.	The response adequately supports the expected accuracy through research citations and tabular format (OMB Pt. B Pg. 86).
2.5 Identify unusual problems requiring specialized sampling procedures.	The supporting statement thoroughly accounts for: establishment burden through Model Assisted Sampling (MAS); occupation coverage through dual frame sampling, establishment and item non-response through data collection procedures.	Given the unknown universe of respondents and establishments, the proposed methodology appears to account for the problems associated with a data collection effort of this type. The response presents an exhaustive summary of procedural tests covering questionnaire design through variance estimation.

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
2.6 Identify use of periodic data collection cycles to reduce burden.	This is a continuous data collection process, however, to reduce burden establishments are selected no more than once within a 12-month period. (OMB Pt. A Pg. 64)	The response adequately meets this objective.
2.7 Identify the procedure(s) used to estimate variance	No subgroups or domain estimates are produced or released. The standard deviation is made available for each item mean, as a measure of response variation among an occupation's respondents. Several percentage estimates for items concerning work context, education and training, occupational tasks and background items are also produced. No item imputation is conducted. Variances are estimated with the first-order Taylor series approximation of deviations of estimates from their expected values using SUDAAN. These estimates account for the combined effects of clustering (establishments), stratification (industry grouping and establishment size as used in selection of the establishment samples) and unequal weighting. Weight adjustments are implemented in three steps of Establishment Method sampling to account for establishment non-response, early termination of sampling for some occupations under MAS, employee non-response and overlapping frames. Extreme weights are trimmed and defined as any weight that exceeds the mean weight of the	The procedures described in OMB P.B Pgs. 97-98 are appropriate and well documented. The formulas are correct and include citations. The response includes a thorough discussion of analysis conducted regarding rules to suppress estimates due to very large variances or inadequate sample sizes (OMB P.B Pg. 113). Weighting adjustments use a generalized exponential model (GEM). The discussion of GEM includes references to similar studies that employ GEM. Weight trimming is appropriate given the criteria presented. The formulas presented are correct. Assumptions are clearly stated. The complex weighting scheme is discussed in an organized and very clear manner.

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
	group plus 1.5 standard deviations of the weights, or a single weight that accounts for more than 50% of the weight sum of the occupation group.	
3.1 Describe methods to maximize response rates and to deal with issues of non-response.	Methods to maximize response rates include multiple contacts, multi-mode study, incentives, refusal conversions, quality of staff, staff training, supervision/call monitoring, and adaptive total design. Non-response is discussed at both the establishment and item levels. At the establishment level, the study accounts for non-response at the verification, screening, recruiting, and sampling stage of selection. The sampling weights are adjusted for nonresponse with the use of a generalized exponential model (GEM). The GEM calibration process causes the weighted distribution of the respondents to match specified distributions simultaneously for all of the variables included in the model. Historical distributions of respondents and non-respondents across various frame attributes showed that the overall potential for nonresponse bias at both the establishment and employee level was negligible. The potential for significant nonresponse bias due to item nonresponse is negligible. (Appendix E Exhibits)	Methods to maximize response rates are exhaustive across the study design and appear to be adequate to achieve the desired response rate. Tests to determine the level of establishment nonresponse, employee nonresponse, and item nonresponse are comprehensive. The use of sampling weights adjusted for nonresponse should minimize the effects of nonresponse bias. Tangentially related, the data collection controls and MAS protocols thoroughly mitigate potential non-response throughout the collection of information.

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
3.2 The accuracy and reliability of information collected is adequate for intended uses?	The sample design was developed to provide results with a level of precision to meet the needs of general-purpose users seeking information at the occupational level.	The response meets the requirement.
3.3 Provide justification for any collection that will not yield "reliable" data that can be generalized to the universe studied.	None.	Given the proposed methodology, and continuing nature of the O*NET collection, this evaluation should be able to obtain reliable data and yield results which are generalizable to the entire population.
4.1 Describe test of procedures and methods to be undertaken to minimize burden and improve utility.	The O*NET Establishment Method minimizes burden on establishments by: 1. Asking establishments about no more than 10 occupations each. 2. Asking establishments to complete employee rosters only for five or fewer occupations. 3. Selecting establishments no more than once in a 12-month period. 4. Selecting no more than 20 employees across all occupations. 5. Defaulting to dual-frame approach or OE Method if an occupation is difficult to sample under the Establishment Method. MAS reduces the number of establishment contacts required to satisfy the random sample allocation for an occupation. Also, the collection offers incentives to employers, POCs, job incumbents, and occupation experts.	The protocols and methodology described are robust and appropriate. The response cites improved establishment response rates of 76% compared to 64% for the initial (2001-2002) data collection. Employee response rates improved from 63% to 65% over the same time period.

REQUIREMENTS	*RESPONSES*	*REVIEWER COMMENTS*
4.2 Tests must be approved if they call for answers to identical questions from 10 or more respondents. Has the proposed test been submitted for approval separately/ in combination with the collection of information?	This is an ongoing data collection effort which has operated continuously since 2001. Over the project, several tests of procedures have been implemented to improve the survey quality and efficiency of the collection, including: • Pretests comparing business reply envelopes vs. first class stamps and respondent and point of contact incentives; • Analysis of weight trimming, MAS, alternative measures of uncertainty, suppression of estimates with poor precision, alternative levels of population coverage, Adaptive Total Design to track data collection metrics, and unit and item non-response. The DOL/Employment Training Administration was notified July 30, 2008, of BLS's approval of the OMB package.	The document does not state if there are any major changes not included within the previously submitted and approved OMB package. There is no reason to believe that additional tests are necessary following the July 30, 2008 approval described in the document.
5.1 Provide contact information for individuals that consulted on statistical aspects of the design.	See page 116 of the OMB Supporting Statement Part B.	Relevant contact information was provided.
5.2 Provide contact information for the agency unit, contractor(s), grantee(s), or other person(s) who will collect and/or analyze the information.	See page 116 of the OMB Supporting Statement Part B.	Relevant contact information was provided.

Appendix I

Establishment Method sample selection steps:

1. Create establishment sampling frame: A matrix is created for each occupation; the matrix lists the industries in which the occupation is found and, for each industry, the number of employees and associated establishments by Census region and by each of four size categories (no employees (unknown), 1-9, 24-249, 250+) .
2. Determine industries to target for each occupation in sub-wave: Categorize the step 1 matrix into 4 concentration groups (high, medium, low, or none) which indicate the probability of an industry to contain the occupation. To increase efficiency in the sampling process, the sample is designed to oversample the high industries and under sample the low industries for each occupation.
3. Select initial sample of establishments from frame: A stratified simple random sample of establishments is first selected from the list of known establishments, with a sample size larger than the number that will ultimately be contracted.
4. Select final set of establishments for sub-wave: A subsample of establishments is selected with probability proportionate to a composite size measure ($S_i = w_{1i} \sum_j f_j M_{ij}$) to lessen the variation in selection probabilities and in the final analysis weights.
5. Assign occupations to selected establishments: Each establishment selected in Step 4 is assigned a maximum of 10 occupations randomly selected with probability proportional to size. Certainty occupations (because of their size) are automatically included and removed from the frame. The non-certainty occupations are sampled and placed in a random order. The sampling list is then compiled by listing the certainty occupations first and then the randomly ordered non-certainty occupations. PMR selection method is used to select occupations with PPS. A sample of up to 10 occupations for each establishment will be selected.
6. Specify employee sample selection algorithm for each occupation/establishment: A random sample of employees from the roster is conducted where the number of employees selected from each occupation is proportional to the number of times the occupation was selected in Step 5. The total number of employees selected within any single establishment never exceeds 20 and the total number in an occupation in the given establishment never exceeds 8.
7. Specify occupation-specific algorithm for random assignment of questionnaire types of sampled employees: all employees selected in Step 6 are randomly assigned to 1 of 3 questionnaire types. Assignments are made in proportion to the number of employee respondents required for each questionnaire type in a sub-wave analysis weights are computed for each occupation.
8. Randomly assign selected establishments to business liaisons for data collection: Random assignment of establishments to business liaisons for surveying.