



National Construction Industry Wage Rate Survey in British Columbia

Methodology Report

June 28, 2001

**Prepared by:
Business Survey Methods Division
and
Small Business and Special Surveys Division
Statistics Canada
R.H. Coats Building 3-C
Ottawa, Ontario
K1A 0T6**

Table of Content

1 - BACKGROUND INFORMATION	1
2 - METHODOLOGY	1
2.1 QUESTIONNAIRE.....	1
2.2 SAMPLE DESIGN AND SELECTION	2
2.2.1 Sampling frame	2
2.2.2 Stratification	2
2.2.3 Sample Selection	2
2.3 DATA COLLECTION.....	3
Count.....	3
2.4 DATA PROCESSING.....	4
2.4.1 Verification of Data from Influential Respondents	4
2.4.2 Wage Rate Calculation	5
2.4.3 Editing.....	5
2.4.4 Outlier Detection	5
2.4.5 Imputation.....	5
Description.....	6
2.5 ESTIMATION AND VARIANCE CALCULATION.....	6
2.5.1 Adjustment of sample counts.....	6
2.5.2 Estimation, Variance Calculation and Publication	7

National Construction Industry Wage Rate Survey (British Columbia)

Methodology Report

1 - BACKGROUND INFORMATION

The National Construction Industry Wage Rate Survey (British Columbia) was conducted by Statistics Canada on behalf of the client, the Labour Program of Human Resources Development Canada (HRDC). The main objective of this study was to produce statistical information on wages for a set of occupations supplied by the client at the provincial, regional and unionized/not unionized level in the province of British Columbia.

2 - METHODOLOGY

The survey was conducted as a sample survey using a computer-assisted telephone interview (CATI) methodology.

2.1 QUESTIONNAIRE

Wage survey questionnaires previously developed for other provinces by Statistics Canada were modified based on consultation with unions and employer associations from British Columbia and adapted for the CATI application. In the final questionnaire, respondents were asked to provide the following information for up to 6 occupations in their establishment. The information requested from respondents was only about their activities in the commercial or institutional construction sector.

Variable Name	Description
nq14	Usual number of hours worked per week for full-time employees
q15h	Starting hourly wage for full-time employees
q16h	Usual hourly wage for full-time employees
q18h	Maximum hourly wage for full-time employees
q92h	Most frequently paid hourly wage for full-time employees
nq1	Number of full-time employees employed by the establishment at any time during the last 12 months
nq3a	Indicates if the employees for the occupation are unionized
nq13	Number of full-time employees employed for the occupation at any time during the last 12 months

2.2 SAMPLE DESIGN AND SELECTION

2.2.1 Sampling frame

The National Construction Industry Wage Rate Survey (British Columbia) used the June 2000 version of Statistics Canada's Business Register (BR) as its sampling frame. This database is constructed and adopted, using various types of tax records from the Canada customs and Revenue Agency and is also updated regularly, using feedback from other business surveys. The BR contains the universe of establishments in Canada. Establishments in the construction industry with 6 or more employees were considered to be in scope for this survey.

2.2.2 Stratification

Stratification for the survey is defined using three variables: region, industry and occupation.

The region component of the stratification was defined using Economic Regions (ER). Three regions were used for stratification purposes in British Columbia. They were the Greater Vancouver-Centre-South area, the Island-Coast-Northern region and the South-Interior region. These groupings were necessary to increase the likelihood of obtaining publishable results for all of the occupations being targeted.

The industry component of the stratification is defined using the 3-digit level of the 1980 Standard Industrial Classification (SIC) which are in fact the sub-sectors of the construction industry (14 sub-sectors).

The occupation component of the stratification was defined using the list of occupations developed according to the needs identified by the client and the consultation groups and to the feasibility of finding enough information about them in the survey. The sample covered 37 occupations in British Columbia.

2.2.3 Sample Selection

Targeting for occupations was done using census tabulations that show the number of people employed in a specific province by 3-digit level SIC and by 4-digit level National Occupation Classification (NOC) combinations. In each 3-digit level SIC, only the top 12 occupations in terms of percent coverage were selected. In cases where fewer than 12 occupations had employment in a 3-digit SIC according to 1996 Census data, tabulations were collapsed to the 2-digit SIC level to fill the remaining list of 12 occupations. If the list of occupations was still fewer than 12 due to there being fewer than 12 occupations with employment at the 2-digit SIC level, then we would go with a list of fewer than 12 occupations. Establishments within the sample were asked, from this list of occupations, to identify up to six of them for which they have had employees in the last 12 months. As a result, in order to keep the response burden low, questions were only asked for a maximum of six occupations.

2.3 DATA COLLECTION

Data collection started in early January and ended in early March. Some units for which there were no valid phone numbers went through tracing in order to locate the establishment. In some cases, the establishments themselves could not give the information requested and referred us to the head office.

The final collection status by establishment was as follows:

Completion Status	Count
Complete	1588
Refusals	105
Out-of-scope	1132
Out-of-business	713
Unable to locate	206
Other non-response	198
Duplicate	25
Total sample selected	3967
Response Rate	75.7%

Response Rates are computed as follows:

Completes / (# Total Sample Units Selected - # Out-of-scopes - # Out-of-businesses - # Duplicates)

Each establishment included in the sample received a letter in advance of the survey, containing information describing the purpose of the survey, questions that were going to be asked and occupations that were being targeted. This mail-out undoubtedly contributed to keeping the refusal rate fairly low.

The large number of the out-of-scope group is composed of responding establishments that reported they did not have employees doing commercial or institutional construction work, or they did not have any of the occupations listed. The main cause was either a lack of information on the survey frame, identifying establishments doing commercial or institutional construction work; or the nature of the survey itself since we are sampling establishments but ask questions on occupations.

Some establishments, falling into the out-of-business category, could not be located because they are no longer in operation. The Business Register keeps track of businesses' births and deaths, but can never be absolutely current. Other establishments could not be contacted for many reasons, including incorrect phone numbers (unable to locate) and no answer or answering machines or language barrier (other non-response). Furthermore, some duplicates occurred when two records in the sample corresponded to the same establishment.

The response rate of 75.7% is considered fairly good for a voluntary survey.

2.4 DATA PROCESSING

The data was converted from an “establishment-level” format to an “occupation-level” format. Each occupation record was examined to ensure that at least one wage question had a response. Any records that did not have at least one response to a wage question were removed from the file and were not processed further.

2.4.1 Verification of Data from Influential Respondents

Some respondents reported wage rates that potentially had a large impact on the estimates. Consequently an important step to ensuring estimates of good quality was to verify the data collected from some of the respondents.

Preliminary wage rate estimates were calculated for the most frequently paid wage rate variable prior to the verification and the impact that each respondent had on the estimates was determined. Respondents were then called back to verify that the data collected was correct if any of the following criteria was satisfied:

1. If the number of employees reported by the establishment was at least 100 more than the number of employees according to the BR.
2. When the total number of employees reported in each of the occupations was considerably more than the establishment reported they had in total.
3. If there was at least 100 employees reported by the establishment and there was no response for the number of employees in a particular occupation.
4. If the weight assigned to an establishment for an estimate was at least 33% of the total weights for the estimate.
5. If an establishment had at least a \$0.25 impact on the wage estimate for a given occupation.

If an establishment was selected for verification, then the data was verified for all the occupations that the establishment reported. In total, the data was verified for 121 establishments.

The interviewers conducting the verification clarified whether or not the establishment was in scope for the survey (i.e. commercial and institutional construction), that the occupations were classified properly, that the number of employees reported were correct and that the wages reported were correct (i.e. made sure wages reported excluded vacation pay and benefits).

In most of the cases, the respondent indicated that the data originally collected was correct. In some cases, however, the respondent indicated that their establishment was actually out of scope for our survey, when probed further by the interviewer conducting the verification. In other cases the respondent made corrections to the number employees or the wages they had initially reported.

The data collected from the verification was then used in place of the original data that was collected for those establishments, in all subsequent data processing steps.

2.4.2 Wage Rate Calculation

Wage rates were provided by respondents according to the time period of their choice: hourly, daily, weekly, biweekly, twice per month, monthly or annually. Wages reported on an other-than-hourly basis were converted to hourly rates using information on the usual number of hours worked per week, which was provided for each occupation (variable nq14). Daily rates were multiplied by 5, to give a weekly wage, then divided by the usual number of hours worked per week. Weekly rates were divided by the usual number of hours worked per week reported for the occupation. Biweekly rates were divided by 2 times the usual number of hours worked per week. Twice per month rates were multiplied by 24, to give the annual salary, then divided by 52 times the usual number of hours worked per week. Monthly rates were multiplied by 12, to give an annual salary, then divided by 52 times the usual number of hours worked per week. Salaries reported on an annual basis were divided by 52 times the usual number of hours worked per week. Almost all respondent answers were reported on an hourly basis.

2.4.3 Editing

After calculating the hourly wages, they were reviewed for anomalies. The first test was to examine any case where the hourly wage was below the provincial minimum wage (\$7.60 in British Columbia) or over \$100.00. The second test was to examine cases where the usual wage rate was not between the starting wage rate and the maximum wage rate an employer was willing to pay. The last test was to examine cases where the most frequent wage rate paid was greater than the maximum wage an employer was willing to pay. The minimal set of imputations required for the wage variables to pass the edits was then determined and the corresponding wage variables were flagged for imputation.

Consistency edits were performed on the two “number of employees” variables. The respondent was asked to report the “total number of employees in the establishment” and “the number of full-time employees in each specific occupation”. A first test was if the “number of full-time employees in the occupation” was greater than the “total number of employees in the establishment”. A second test compared the number of employees reported by the establishment to the number of employees that the establishment had according to the BR. These cases were reviewed and either fixed, sent to imputation or accepted as is.

2.4.4 Outlier Detection

Outlier detection was performed on the four wage questions in order to detect values that are improbable or influential. This detection was done on the raw data and the method used was the Quartile Method. It is a univariate, non-parametric, outlier detection method that uses the sample median and the sample quartile ranges. Values were also identified as outliers if the wages were less than half or more than twice the median wage. Values flagged as outliers were set to missing to be imputed at the next stage of processing. Very few outliers were actually observed.

2.4.5 Imputation

Imputation was done whenever an expected wage value was missing or whenever one of the two “number of employees” variables were missing or flagged for imputation.

For the wage variables, the first step in the imputation process was to calculate the relationships (trends) between all non-missing wage variables that were not flagged for imputation. For this, each occupation was treated separately. The next step was to identify a “base” variable to be used as the basis of the imputation. The base variable is the first non-missing variable that was not flagged for imputation when looking at wage variables one by one in a pre-determined order. For the current survey, the pre-determined order of wage variables for imputation was q92h, q18h, q16h and q15h.

Missing values were imputed, by multiplying the corresponding trend factor by the base wage variable. If the resulting value was below the minimum wage of \$7.60, then the minimum wage was imputed. Also, if the resulting value would have been detected as an outlier, then the minimum or maximum wage that would not have been detected as an outlier was imputed, depending on whether the value was lower or higher than the allowable range respectively. After imputation, edits were re-run on the imputed records and adjustments were made to the imputed values, if necessary.

A similar approach was used to impute the number of employees’ variables. First, if the total number of employees in the establishment was missing, we imputed the number of employees that was available from our frame. Then, the other number of employee variable was imputed based on the total for the establishment using a trend again.

The table that follows gives the imputation rates for all the wage and employee questions

Variable Name	Description	Imputation Rate
q15h	Starting hourly wage for full-time employees	3.2%
q16h	Usual hourly wage for full-time employees	2.3%
q18h	Maximum hourly wage for full-time employees	3.5%
q92h	Most frequent paid hourly wage for full-time employees	1.8%
nq13	Number of full-time employees employed for the occupation at any time during the last 12 months	0%

2.5 ESTIMATION AND VARIANCE CALCULATION

2.5.1 Adjustment of sample counts

Before estimation and weighting, sample counts were adjusted for non-response. The

adjustment of counts is based on the assumptions that the non-response is random and that a non-respondent is not characteristically different from a respondent.

2.5.2 Estimation, Variance Calculation and Publication

Estimates were produced using GES; the Generalized Estimation System developed at Statistics Canada. GES is a SAS-based application for producing estimates for domains of a population.

Records were assigned weights based on the number of employees in the occupations meaning establishment with more employees for an occupation had a larger contribution to the estimates for that occupation. A weighted-average, hourly wage rate was estimated for every occupation pertaining to full-time employees and for the four different wage concepts (variables q15h, q16h, q18h and q92h). Estimates were also produced by sub-provincial area and union status. The technique used is known as domain estimation, in which records from different strata are grouped together in a “domain”. In this case, a domain regroups all records for a given occupation for the purposes of calculating statistics, regardless of industry and, in some cases, region. All records coded as out-of-scope, or out-of-business, are excluded from the domain estimates.

Since all the estimates produced from this survey are based on sample results, they are subject to sampling error. This error can be expressed as a coefficient of variation (CV). The CV is a percentage that expresses the size of the standard error as a proportion of the estimate to which it is related. For example, a CV of 10% indicates that the standard error is 10% of the estimate. If a wage rate estimate is \$9.50 per hour, with a CV of 10%, then the standard error is \$0.95. In the output reports, CVs have been converted to a code that can be interpreted using the following table:

Value of CV	Code	Rating
0 to 5%	A	Very Good
5.01% to 15%	B	Good
15.01% to 33%	C	Good to Poor-- use with caution
33.01% and over	D	Very poor -- may not be acceptable

Most of the wage-rate estimates fall into the Very good category while a few others are in the Good category.

All estimates created with four or fewer sample units – or establishments – have not been released for reliability and/or confidentiality purposes.

At the provincial level, the estimates were released for all 37 occupations of interest. It was also determined that there was insufficient information to publish many of the estimates broken down by sub-provincial area and union status.