

Supporting Document B
Variables, Rates, and Formulae for Calculating Response
Rates and Reporting Requirements: Economic Surveys
and Censuses

Version 1.6

Issued: 23 Dec 08

*Census Bureau Standard
Response Rate Definitions*

Authored by:

Nancy A. Bates (DSD)
(Team Leader)
Deborah H. Griffin (DIR)
Rita J. Petroni (ESMPD)
James B. Treat (PRED)

USCENSUSBUREAU

Helping You Make Informed Decisions



Document Management & Control¹

Version	Issue Date	Approval	Description
1.0	05 July 05	Associate Directors	Initial Release
1.1	09 Mar 06	Quality Prog. Mgr	Inserted hyperlink for main standard.
1.2	28 Mar 06	Configuration Mgr.	Reformatted Section 3.
1.3	01 Aug 06	Quality Prog. Mgr	Clarified the intent of reporting requirement 4
1.4	21 Jun 07	Quality Prog. Mgr	Removed hyperlink for main standard.
1.5	20 Jun 08	M&S Council	Updated Definitions and equations
1.6	23 Dec 08	Quality Prog. Mgr	Updated Definitions in 1.4 for q_{it} , s_{it} and in 2.2.2 for S rate

¹ **The current version of this document is maintained on the Census Bureau Intranet and may be accessed from the Methodology & Standards Council Website.**

1. Variables and Terms

The variables needed to calculate economic survey and census response rates are partitioned into four groups. The first set of variables describes eligibility status. These variables distinguish between reporting units that are known to be eligible for data collection, known to be ineligible for data collection, to have an unknown eligibility for data collection, or have their data obtained from administrative sources or an outside source, validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area. The data collection target population guides the distinction between eligible and ineligible units. The second set describes the response status for eligible reporting units. The third set of variables provides detail on nonresponse of reporting units in the sample by identifying the type of (or the reason for) the unit's nonresponse. The fourth set of variables is required for the computation of quantity response rates.

Many economic programs distinguish between two types of unit: reporting and tabulation. The reporting unit represents the original sampled unit (e.g., the company, tax entity, or establishment). Depending on the program, the reporting unit data may be used to create separate tabulation units for estimation. For example, when a reporting unit operates in more than one kind of business and is unable or unwilling to report separately for each, artificial records are created for the reporting unit in order to split the respondent's data among the different types of businesses (industries). These artificial records are the tabulation units. In other cases, a program may consolidate establishment or plant level data to the company level to create a tabulation unit. Thus, for economic surveys, unit response rates are based on the disposition of the reporting unit and item response rates (quantity and total quantity response rates) are based on the disposition of the tabulation units.

1.1 Eligibility Status Terms

The total number of reporting units selected for the sample is defined as N_R . These reporting units can be classified by their eligibility status; eligible for data collection (E), ineligible (L), unknown eligibility (U), or data obtained from qualified administrative sources (A). The target population refers to units upon which inferences (estimates) are made. Reporting units may be considered eligible in one survey or census but ineligible for another, depending upon the target population. For example, a reporting unit that was in business after October 2004 is eligible for the 2004 Annual Trade Survey, but is ineligible for the October Monthly Retail Trade Survey.

Variable	e_i
Definition	An indicator variable for whether a reporting unit selected for the sample is eligible for data collection. These include chronic refusal units (eligible reporting units that have notified the Census Bureau that they will not participate in a given program). If a reporting unit is eligible $e_i = 1$, else $e_i = 0$.

Term	E
Definition	The count of reporting units selected for the sample that were eligible for data collection.
Computation	The sum of the indicator variable for eligibility (e_i) over all the units selected for the sample. $E = \sum_{i=1}^{N_R} e_i$
Variable	l_i
Definition	An indicator variable for whether a reporting unit selected for the sample has been confirmed as not a member of the target population at the time of data collection. An attempt was made to collect data and it was confirmed that the reporting unit was not a member of the target population at that time. These reporting units are not included in the unit response rate calculations for the periods in which they are ineligible. Information confirming ineligibility may come from observation, from a respondent or from another source. Some examples of ineligible reporting units include firms that went out of business prior to the survey reference period and firms in an industry that is out-of-scope for the survey in question. If a reporting unit is ineligible $l_i = 1$, else $l_i = 0$.
Term	L
Definition	The count of reporting units selected for the sample that were ineligible for data collection in the current collection period.
Computation	The sum of the indicator variable for ineligibility (l_i) over all the units selected for the sample. $L = \sum_{i=1}^{N_R} l_i$
Variable	u_i
Definition	An indicator variable for whether the eligibility of a reporting unit selected for the sample could not be determined. If a reporting unit is of unknown eligibility $u_i = 1$, else $u_i = 0$.
Term	U
Definition	The count of reporting units selected for the sample for which eligibility could not be determined.
Computation	The sum of the indicator variable for unknown eligibility (u_i) over all the units selected for the sample. $U = \sum_{i=1}^{N_R} u_i$
Variable	a_i
Definition	An indicator variable for whether administrative data of equivalent-quality-to-reported data rather than survey data was obtained for an eligible reporting unit selected for the sample. The decision not to collect survey data must have been made for survey efficiency or to reduce respondent burden and not because that reporting unit had been a refusal in the past. These reporting units are excluded from unit response rate and the Quantity response rate calculations because they

were not sent questionnaires, and thus could not respond. But their data are included in the calculation of the Total Quantity response rate, one of the item response rates. If a reporting unit is pre-selected to receive administrative data $a_i = 1$, else $a_i = 0$.

Term	A
Definition	The count of reporting units selected for the sample that belong to the target population and were pre-selected to use administrative data rather than collect survey data.
Computation	The sum of the indicator variable for units pre-selected to use administrative data (a_i) over all the units selected for the sample. $A = \sum_{i=1}^{N_R} a_i$

The relationship among the counts of reporting units selected for the sample in the four eligibility categories is given by $N_r = E + L + U + A$.

1.2 Response Status

Response status is determined only for the eligible reporting units selected for the sample.

Variable	r_i
Definition	An indicator variable for whether an eligible reporting unit selected for the sample responded to the survey. To be classified as a response, the respondent for the reporting unit must have provided sufficient data and the data must satisfy all the edits. The definition of sufficient data will vary across surveys. Programs must designate required data items before the data collection begins. If a reporting unit responded, $r_i = 1$, else $r_i = 0$ (note $r_i = 0$ for reporting units which were eligible but did not respond and for reporting units classified as L , U , or A).

Term	R
Definition	The count of reporting units selected for the sample that were eligible for data collection and classified as a response.
Computation	The sum of the indicator variable for units that responded (r_i) over all the units selected for the sample. $R = \sum_{i=1}^{N_R} r_i$

1.3 Reasons for Nonresponse

To improve interpretation of the response rate and better manage resources, it is recommended that whenever possible, reasons for (or types of) nonresponse are measured. These terms are used to describe “unit nonresponse” and will be presented in unweighted tabulations. Five specific terms describing nonresponse reasons are defined below. The first three terms (*REF*, *CREF*, and *INSF*) define nonresponse reasons for eligible reporting units. The final two terms (*UAA* and *OTH*) define the reasons for reporting units with unknown eligibility.

Variable	ref_i
Definition	An indicator variable for whether an eligible reporting unit selected for the sample refused to respond to the survey. If a reporting unit refuses to respond $ref_i = 1$, else $ref_i = 0$.
Term	<i>REF</i>
Definition	The count of eligible reporting units selected for the sample that were classified as “refusal.”
Computation	Sum of the indicator variable for “refusal” (ref_i) over all the reporting units selected for the sample. $REF = \sum_{i=1}^{N_R} ref_i$
Variable	$cref_i$
Definition	An indicator variable for whether an eligible reporting unit selected for the sample was a “chronic refusal.” A chronic refusal is a reporting unit that informed the Census Bureau they would not participate. The Census Bureau does not send questionnaires to chronic refusals. Chronic refusals comprise a subset of refusals. If a reporting unit is a chronic refusal $cref_i = 1$, else $cref_i = 0$.
Term	<i>CREF</i>
Definition	The count of eligible reporting units selected for the sample that were classified as “chronic refusals”.
Computation	The sum of the indicator variable for “chronic refusal” ($cref_i$) over all the reporting units selected for the sample. $CREF = \sum_{i=1}^{N_R} cref_i$
Variable	$insf_i$
Definition	An indicator variable for whether an eligible reporting unit that was selected for the sample returned a questionnaire, but did not provide sufficient data to qualify as a response. If a reporting unit returned a questionnaire but fails to provide sufficient data to qualify as a response $insf_i = 1$, else $insf_i = 0$.
Term	<i>INSF</i>
Definition	The count of eligible reporting units selected for the sample that were classified as providing insufficient data.
Computation	The sum of the indicator variable for “insufficient data” ($insf_i$) over all the reporting units selected for the sample. $INSF = \sum_{i=1}^{N_R} insf_i$
Variable	uaa_i
Definition	An indicator variable for whether a reporting unit selected for the sample had a questionnaire returned as “undeliverable as addressed.” These reporting units are of unknown eligibility. If a questionnaire is returned as “undeliverable as addressed” $uaa_i = 1$, else $uaa_i = 0$.

Term	<i>UAA</i>
Definition	The count of reporting units selected for the sample that were classified as “undeliverable as addressed.”
Computation	The sum of the indicator variable for “undeliverable as addressed” (uaa_i) over all the reporting units selected for the sample. $UAA = \sum_{i=1}^{N_R} uaa_i$
Variable	oth_i
Definition	An indicator variable for whether a reporting unit selected for the sample was a nonresponse for a reason other than refusal, insufficient data, or undeliverable as addressed. These reporting units are of unknown eligibility. If a reporting unit does not respond for reasons other than refusal, insufficient data, or undeliverable as addressed $oth_i = 1$, else $oth_i = 0$.
Term	<i>OTH</i>
Definition	The count of reporting units selected for the sample that were classified as “other nonresponse.”
Computation	The sum of the indicator variable for “other nonresponse” (oth_i) over all the reporting units selected for the sample. $OTH = \sum_{i=1}^{N_R} oth_i$

1.4 Quantity Response Rate Terms

Quantity response rates and total quantity response rates show the combined effects of unit and item nonresponse on weighted estimates from the survey. These response rates are computed using tabulation units.

The total number of tabulation units in the sample is defined as N_T . These tabulation units can be classified according to the source of data obtained for the units; data reported by the respondent (R), equivalent-quality-to-reported data (Q) obtained from the various sources outlined below, or imputed data (M). The number of tabulation units N_T may differ from the number of reporting units N_R . These differences may arise because some businesses keep their records on a different basis than what is desired for a given survey. To accommodate for these differences it may be necessary to aggregate data from multiple reporting units to construct a single tabulation unit or distribute data from a single reporting unit into multiple tabulation units. Item response terms are computed for data item t based on the tabulation units in the sample.

Variable	r_{ti}
Definition	An indicator variable for whether a tabulation unit in the sample provided reported data for item t that satisfied all edits. If the unit provided such data $r_{ti} = 1$, else $r_{ti} = 0$.

Term	R_t
Definition	The count of eligible tabulation units in the sample that provided reported data for item t that satisfied all edits. This count will vary by item.
Computation	The sum of the indicator variable for reported data (r_{ti}) over all the tabulation units (N_T) selected for the sample. $R_t = \sum_{i=1}^{N_T} r_{ti}$
Variable	q_{ti}
Definition	An indicator variable for whether a tabulation unit in the sample contains equivalent-quality-to-reported data for item t . Such data can come from three sources: data directly substituted from another census or survey s (for the same reporting unit, question wording, and time period), administrative data d , or data obtained from some other equivalent source c validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (e.g., company annual reports, Securities and Exchange Commission (SEC) filings, trade association statistics). If the sample contains equivalent-quality-to-reported data $q_{ti} = 1$, else $q_{ti} = 0$.
Term	Q_t
Definition	The count of eligible tabulation units in the sample that contain equivalent-quality-to-reported data for item t . This count will vary by item.
Computation	The sum of the indicator variable for equivalent-quality-to-reported data (q_{ti}) over all tabulation units (N_T) selected for the sample. $Q_t = \sum_{i=1}^{N_T} q_{ti}$
Variable	s_{ti}
Definition	An indicator variable for whether a tabulation unit in the sample contains directly substituted data from another census or survey for item t . The same reporting unit must provide the item value (in the other program), and the question wording and time period for the substituted values must agree between the two programs. If the sample contains directly substituted data from another survey $s_{ti} = 1$, else $s_{ti} = 0$.
Term	S_t
Definition	The count of eligible tabulation units in the sample containing directly substituted data for item t . This count will vary by item.
Computation	The sum of the indicator variable for directly substituted data (s_{ti}) over all tabulation units (N_T) selected for the sample. $S_t = \sum_{i=1}^{N_T} s_{ti}$
Variable	d_{ti}
Definition	An indicator variable for whether a tabulation unit in the sample contains administrative data for item t . If the sample contains administrative data $d_{ti} = 1$, else $d_{ti} = 0$.

Term D_t
 Definition The count of eligible tabulation units in the sample containing administrative data for item t . This count will vary by item.
 Computation The sum of the indicator variable for administrative data (d_{it}) over all tabulation units (N_T) selected for the sample. $D_t = \sum_{i=1}^{N_T} d_{it}$

Variable c_{it}
 Definition An indicator variable for whether a tabulation unit in the sample contains equivalent-source data validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (e.g., company annual report, SEC filings, trade association statistics) for item t . If the sample contains equivalent-source data $c_{it} = 1$, else $c_{it} = 0$.

Term C_t
 Definition The count of eligible tabulation units in the sample containing equivalent-source data for item t . This count will vary by item.
 Computation The sum of the indicator variable for equivalent-source data (c_{it}) over all tabulation units (N_T) selected for the sample. $C_t = \sum_{i=1}^{N_T} c_{it}$

The relationship among Q_t , S_t , D_t , and C_t is given by $Q_t = S_t + D_t + C_t$.

Variable m_{it}
 Definition An indicator variable for whether a tabulation unit in the sample contains imputed data for item t . If the sample contains imputed data $m_{it} = 1$, else $m_{it} = 0$.

Term M_t
 Definition The count of eligible tabulation units in the sample containing imputed data for item t . This count will vary by item.
 Computation The sum of the indicator variable for imputed data (m_{it}) over all tabulation units (N_T) selected for the sample. $M_t = \sum_{i=1}^{N_T} m_{it}$

Variable w_i
 Definition The weight for the i th tabulation unit in the sample (includes subsampling factors and outlier adjustments. May be set = 1 when the tabulation unit is designated as an outlier).

Variable f_i
 Definition A tabulation unit non-response weight adjustment factor for the i th tabulation unit in the sample. The variable f_i is set equal to 1 for surveys that use imputation to account for unit non-response.

Variable	t_i
Definition	The tabulated quantity of the variable for item t for the i th tabulation unit in the sample. This value may be reported, equivalent-quality-to-reported data, or imputed.
Term	T
Definition	The estimated (weighted) total of data item t for the entire population represented by the sample of tabulation units. T is based on the absolute value of the data provided by the respondent, equivalent-quality-to-reported data, or imputed data. The calculation of T incorporates subsampling factors, weighting adjustment factors for unit non-response (adjustment-to-sample procedures only), and outlier-adjustment factors, but does not include post-stratification or other benchmarking adjustments.
Computation	The product of the weight (w_i) non-response weight adjustment factor (f_i), and the tabulated quantity (t_i) for the i th tabulation unit in the sample summed over all tabulation units (N_T) selected for the sample. $T = \sum_{i=1}^{N_T} w_i f_i t_i $

The relationship among the counts of tabulation units in the sample is given by

$$N_T = R_t + Q_t + M_t.$$

2. Response and Nonresponse Rates

2.1 Primary Response Rates

Rate	Unit Response Rate (<i>URR</i>)
Definition	The proportion of reporting units based on unweighted counts, that were eligible or of unknown eligibility that responded to the survey (expressed as a percentage).
Computation	$URR = [R/(E+U)] * 100$

Rate	Quantity Response Rate (<i>QRR</i>) for data item t .
Definition	The proportion of the estimated (weighted) total (T) of data item t reported by tabulation units in the sample (expressed as a percentage). [Note: Because the value of economic data items can be negative (e.g., income), the absolute value must be used in the numerators and denominators in all calculations.]

Computation	$QRR = \left[\frac{\sum_{i=1}^{N_T} w_i \times r_{ii} \times t_i }{T} \right] * 100$
-------------	--

Rate Total Quantity Response Rate (*TQRR*) for data item *t*.
Definition The proportion of the estimated (weighted) total (*T*) of data item *t* reported by tabulation units in the sample or from sources determined to be equivalent-quality-to-reported data (expressed as a percentage).

Computation
$$TQRR = \left[\frac{\sum_{i=1}^{N_T} w_i \times (r_{ti} + q_{ti}) \times |t_i|}{T} \right] * 100$$

2.2 Detailed Response and Nonresponse Rates

2.2.1 Unit Nonresponse Rate Breakdowns

The following breakdowns provide unweighted unit non-response rates.

Rate Refusal Rate (*REF* rate)
Definition The ratio of reporting units selected for the sample that were classified as “refusal” to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation $REF\ rate = [REF/(E+U)] * 100$

Rate Chronic Refusal Rate (*CREF* rate)
Definition The ratio of reporting units selected for the sample that were classified as “chronic refusals” to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation $CREF\ rate = [CREF/(E+U)] * 100$

Rate Insufficient Data Rate (*INSF* rate)
Definition The ratio of reporting units selected for the sample that were classified as “insufficient data” to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation $INSF\ rate = [INSF/(E+U)] * 100$

Rate Undeliverable as Addressed Rate (*UAA* rate)
Definition The ratio of reporting units selected for the sample that were classified as “undeliverable as addressed” to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation $UAA\ rate = [UAA/(E+U)] * 100$

Rate Other Reason for Nonresponse Rate (*OTH* rate)
Definition The ratio of reporting units selected for the sample that were classified as “other reason for nonresponse” to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation $OTH\ rate = [OTH/(E+U)] * 100$

Rate Unknown Eligibility Rate (*U* rate)

Definition The ratio of reporting units selected for the sample that were classified as “unknown eligibility” to the sum of eligible units and units of unknown eligibility (expressed as a percentage).

Computation $U\ rate = [U/(E+U)] * 100$

2.2.2 Total Quantity Response Rate Breakdowns

The following breakdowns provide weighted response rates.

Rate Equivalent-Quality-to-Reported Data Rate (*Q* rate)

Definition The proportion of the total estimate for item *t* derived from equivalent-quality-to-reported data for tabulation units in the sample (expressed as a percentage).

Computation
$$\frac{Q_t}{T} = \left[\frac{\sum_{i=1}^{N_T} w_i \times (s_{ii} + d_{ii} + c_{ii}) \times |t_i|}{T} \right] * 100$$

Rate Survey Substitution Rate (*S* rate)

Definition The proportion of the total estimate for item *t* derived from substituted other survey or census data for tabulation units in the sample (expressed as a percentage). To be tabulated in this rate, substituted data items must be obtained from the same reporting unit in the same time period as the target program, and the question wording between the two programs must agree.

Computation
$$\frac{S_t}{T} = \left[\frac{\sum_{i=1}^{N_T} w_i \times s_{ii} \times |t_i|}{T} \right] * 100$$

Rate Administrative Data Rate (*D* rate)

Definition The proportion of the total estimate of item *t* derived from administrative data for tabulation units in the sample (expressed as a percentage).

Computation
$$\frac{D_t}{T} = \left[\frac{\sum_{i=1}^{N_T} w_i \times d_{ii} \times |t_i|}{T} \right] * 100$$

Rate Other Source Rate (*C* rate)

Definition The proportion of the total estimate of item *t* derived from other source data validated by a study approved by the program manager in collaboration with the appropriate Research and Methodology area (such as company annual reports, SEC filing, trade association statistics) for tabulation units in the sample (expressed as a percentage).

Computation
$$\frac{C_t}{T} = \left[\frac{\sum_{i=1}^{N_T} w_i \times c_{it} \times |t_i|}{T} \right] * 100$$

Rate Imputation Rate (*M* rate)

Definition The proportion of the total estimate of item *t* derived from model imputes for tabulation units in the sample (expressed as a percentage).

Computation
$$\frac{M_t}{T} = \left[\frac{\sum_{i=1}^{N_T} w_i \times m_{it} \times |t_i|}{T} \right] * 100$$

3. Reporting Requirements

Reporting Requirement 1:

When a response rate is reported, the universe of inference or target population for that survey or census must also be described.

To report the universe of inference or target population, the documentation should indicate which units are included and which are excluded from the target population. For example, the Monthly Wholesale Trade Survey includes companies with employment that are primarily engaged in merchant wholesale trades in the U.S. These include merchant wholesalers that take title of the goods they sell, and jobbers, industrial distributors, exporters, and importers. Excluded are non-merchant wholesalers such as manufacturer sales branches and offices; agents, merchandise or commodity brokers, and commission merchants; and other businesses whose primary activity is other than wholesale trade.

Reporting Requirement 2:

When a response rate is reported, it must be noted whether the rate is an unweighted response rate, a quantity response rate, or a total quantity response rate.

When establishment surveys and censuses report quantity or total quantity response rates, the measure of size (i.e., quantity of a variable) used in computing the response rate must be stated. For example, the measure of size in the Monthly Retail Sales survey may be the total retail sales of an establishment.

Reporting Requirement 3:

When using any of the three response rates, surveys or censuses must report the definition of responding units. When using Total Quantity Response Rates, surveys or censuses must also report nonresponse data sources.

Many surveys or censuses allow some leeway when defining responding units. Each survey or census should establish a definition of a responding eligible unit prior to implementing the definition. Units are determined to be responding based on whether key characteristics are reported. For example, when calculating the Total Quantity Response Rate, annual payroll, first quarter payroll, or employment must be reported or obtained from administrative records.

Reporting Requirement 4:

When available, rates should be calculated using the edited data at the time of each estimate's release phase. The response rate for the final release should use the final, edited data. When the appropriate edited data are not used, note this.

When possible, the outcomes used to calculate response rates should reflect outcomes after the data has undergone any edits or other post-data collection processing. When the Census Bureau conducts surveys for other agencies, the sponsoring agency will be the source for the nonresponse rates.

Example: Reporting a response rate in an economic survey.

The dollar volume response (i.e., quantity response rate) to the Annual Wholesale Trade Survey is 89 percent for total sales and 87 percent for total inventories (*reporting requirement 2*). The survey population consists of companies with employment that are primarily engaged in merchant wholesale trades in the U.S. (*reporting requirement 1*) These include merchant wholesalers that take title of the goods they sell, and jobbers, industrial distributors, exporters, and importers. Excluded are non-merchant wholesalers such as manufacturer sales branches and offices; agents, merchandise or commodity brokers, and commission merchants; and other businesses whose primary activity is other than wholesale trade. To be eligible for tabulation, units selected into the sample with probability less than one must be on the latest available IRS mailing list for FICA taxpayers from the most recent available quarter. Units selected with a probability of one are eligible as long as there is no evidence that they were not out of business for the entire reference year(s) covered by the survey (*reporting requirement 3*).