

**Survey glossary**

**Outcome rate** | A measure of the proportion of units in which a housing unit or respondent provides a response to a survey question. Outcome rates are calculated by dividing the number of units that provide a response by the total number of units in the sample. Outcome rates are typically expressed as percentages. Outcome rates are used to assess the representativeness of the sample and to estimate the nonresponse bias. Outcome rates are also used to assess the quality of the data collection process. Outcome rates are typically calculated for each question in the survey. Outcome rates are typically calculated for each question in the survey. Outcome rates are typically calculated for each question in the survey.

**Response rates** | The proportion of units in which a housing unit or respondent provides a response to a survey question. Response rates are calculated by dividing the number of units that provide a response by the total number of units in the sample. Response rates are typically expressed as percentages. Response rates are used to assess the representativeness of the sample and to estimate the nonresponse bias. Response rates are also used to assess the quality of the data collection process. Response rates are typically calculated for each question in the survey. Response rates are typically calculated for each question in the survey. Response rates are typically calculated for each question in the survey.

**Refusal rates** | The proportion of units in which a housing unit or respondent refuses to do an interview or breaks-off an interview of all potentially eligible units. Refusal rates are calculated by dividing the number of units that refuse to do an interview or break-off an interview by the total number of units in the sample. Refusal rates are typically expressed as percentages. Refusal rates are used to assess the representativeness of the sample and to estimate the nonresponse bias. Refusal rates are also used to assess the quality of the data collection process. Refusal rates are typically calculated for each question in the survey. Refusal rates are typically calculated for each question in the survey. Refusal rates are typically calculated for each question in the survey.

**Contact rates** | The proportion of all units are reached by the survey. Contact rates are calculated by dividing the number of units that are reached by the survey by the total number of units in the sample. Contact rates are typically expressed as percentages. Contact rates are used to assess the representativeness of the sample and to estimate the nonresponse bias. Contact rates are also used to assess the quality of the data collection process. Contact rates are typically calculated for each question in the survey. Contact rates are typically calculated for each question in the survey. Contact rates are typically calculated for each question in the survey.

**Outlier** | An atypical observation which does not appear to follow the distribution of the rest of a dataset. Outliers are typically identified by statistical methods. Outliers are typically identified by statistical methods. Outliers are typically identified by statistical methods.

**Overediting** | Extensive editing that becomes too costly for the amount of error that is being reduced. Overediting is typically identified by statistical methods. Overediting is typically identified by statistical methods. Overediting is typically identified by statistical methods.

**Overlap in the split translations** | A compromise solution between split and full translations is to ensure that some overlap exists between materials divided among translators. The material is split up the way cards are dealt in many games, everyone getting a spread of the material. Each translator could then receive the last one or two questions of another translator's "piece". This allows the review team members to have an increased sense of differences in translating approaches between translators and their understanding of source text components at the draft production level. Overlap in the split translations is typically identified by statistical methods. Overlap in the split translations is typically identified by statistical methods. Overlap in the split translations is typically identified by statistical methods.

**Overrun** | The exceeding of costs estimated in a contract. Overruns are typically identified by statistical methods. Overruns are typically identified by statistical methods. Overruns are typically identified by statistical methods.

**Paradata** | Couper first introduced the term "paradata" into survey research methodology field (Groves & Couper, 1998) and the definition of paradata has vastly expanded since then. Paradata now refers to additional data that can be captured during the process of producing a survey statistic (Kreuter, 2013). As discussed in the 2011 International Nonresponse Workshop (Smith, 2011), two main types of paradata are available. One is process paradata, which is collected during the process of data collection, such as time stamps and keystroke data. Another type is related with observational information, such as the observed demographic information of respondents and observed neighborhood conditions. Paradata is typically identified by statistical methods. Paradata is typically identified by statistical methods. Paradata is typically identified by statistical methods.

**Pareto chart** | A bar chart that reflects the types of most errors in a process, by error type in descending order; for example, the five or six most frequent types of help desk calls from interviewers using computer-assisted interviewing. Pareto charts are typically identified by statistical methods. Pareto charts are typically identified by statistical methods. Pareto charts are typically identified by statistical methods.

**Performance measurement analysis** | A technique used in quality control to determine whether quality assurance procedures have worked. For example, analysis of routine measures of interviewer or coder performance. Performance measurement analysis is typically identified by statistical methods. Performance measurement analysis is typically identified by statistical methods. Performance measurement analysis is typically identified by statistical methods.

**Personally Identifiable Information (PII)** | Information that can be used to identify a respondent that minimally includes name, address, telephone number and identification number (such as social security number or driver's license number), but may include other information including biometric data. PII is typically identified by statistical methods. PII is typically identified by statistical methods. PII is typically identified by statistical methods.

**Pilot study** | A quantitative miniature version of the survey data collection process that involves all procedures and materials that will be used during data collection. A pilot study is also known as a "dress rehearsal" before the actual data collection begins. Pilot studies are typically identified by statistical methods. Pilot studies are typically identified by statistical methods. Pilot studies are typically identified by statistical methods.

**Pledge of confidentiality** | An agreement (typically in written or electronic form) to maintain the confidentiality of survey data that is signed by persons who have any form of access to confidential information. Pledges of confidentiality are typically identified by statistical methods. Pledges of confidentiality are typically identified by statistical methods. Pledges of confidentiality are typically identified by statistical methods.

**Portable file** | A file that is coded in a non-proprietary format such as XML or ASCII and thus can be used by a variety of software and hardware platforms. Portable files are typically identified by statistical methods. Portable files are typically identified by statistical methods. Portable files are typically identified by statistical methods.

**Post-survey adjustments** | Adjustments to reduce the impact of error on estimates. Post-survey adjustments are typically identified by statistical methods. Post-survey adjustments are typically identified by statistical methods. Post-survey adjustments are typically identified by statistical methods.

**Poststratification** | A statistical adjustment that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection. Poststratification is typically identified by statistical methods. Poststratification is typically identified by statistical methods. Poststratification is typically identified by statistical methods.

**Poststratification adjustment** | A statistical adjustment that assures that sample estimates of totals or percentages (e.g., the estimate of the percentage of men in living in Mexico based on the sample) equal population totals or percentages (e.g., the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for poststratification are formed in a similar way as strata in sample selection, but variables can be used that were not on the original sampling frame at the time of selection. Poststratification adjustment is typically identified by statistical methods. Poststratification adjustment is typically identified by statistical methods. Poststratification adjustment is typically identified by statistical methods.

**Precision** | A measure of how close an estimator is expected to be to the true value of a parameter, which is

usually expressed in terms of imprecision and related to the [variance](#) of the estimator. Less [precision](#) is reflected by a larger variance. |- | [Precoding](#) | When designing the questionnaire and survey instrument, determine [coding](#) conventions and formats of survey [items](#) (especially the [closed-ended questions](#)) based on existing coding frames or prior knowledge of the [survey population](#). |- | [Prescribed behaviors](#) | Interviewer behaviors that must be carried out exactly as specified. |- | [Pretesting](#) | A collection of techniques and activities that allow researchers to evaluate survey questions, questionnaires and/or other survey procedures before data collection begins. |- | [Primacy](#) | [Context effects](#) in which the placement of the [item](#) at the beginning of a list of [response options](#) increases the likelihood that it will be selected by the respondent. |- | [Primary Sampling Unit \(PSU\)](#) | A [cluster](#) of elements sampled at the first stage of selection. |- | [Probability proportional to size \(PPS\)](#) | A sampling method that assures that sample estimates of totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on the sample) equal population totals or percentages (e.g. the estimate of the percentage of men living in Mexico based on Census data). The adjustment cells for postratification are formed in a similar way as [strata](#) in sample selection, but variables can be used that were not on the original [sampling frame](#) at the time of selection. |- | [Probability sampling](#) | A sampling method where each element on the [sampling frame](#) has a known, non-zero chance of selection. |- | [Process analysis](#) | The use of tools such as flowcharts to analyze processes, e.g., respondent [tracking](#), computerized instrument programming and testing, [coding](#), data entry, etc. The aim is and to identify indicators or measures of the [quality](#) of products. [Process analysis](#) also is used to identify improvements that can be made to processes. |- | [Process improvement plan](#) | A plan for improving a process, as a result of [process analysis](#). A [process improvement plan](#) may result from development of a [quality](#) management plan, or as a result of [quality assurance](#) or [quality control](#). |- | [Process indicator](#) | An indicator that refers to aspects of data collection (e.g., [HPI](#), [refusal rates](#), etc.). |- | [Processing error](#) | Survey error ([variance](#) and [bias](#)) that arise during the steps between collecting information from the respondent and having the value used in estimation. [Processing errors](#) include all post-collection operations, as well as the printing of questionnaires. Most processing errors occur in data for individual units, although errors can also be introduced in the implementation of systems and estimates. In survey data, processing errors may include errors of transcription, errors of [coding](#), errors of data entry, errors in the assignment of weights, errors in disclosure avoidance, and errors of arithmetic in tabulation. |- | [Progress indicator](#) | An indicator that refers to aspects of reaching the goal (e.g., number of complete interviews). |- | [Proxy interview](#) | An interview with someone (e.g., parent, spouse) other than the person about whom information is being sought. There should be a set of rules specific to each survey that define who can serve as a [proxy](#) respondent. |- | [Public use data files](#) | An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze. |- | [Public-use data file](#) | An anonymized data file, stripped of respondent identifiers that is distributed for the public to analyze. |- | [Quality](#) | The degree to which product characteristics conform to requirements as agreed upon by producers and clients. |- | [Quality assurance](#) | A planned system of procedures, performance checks, [quality](#) audits, and corrective actions to ensure that the products produced throughout the [survey lifecycle](#) are of the highest achievable quality. [Quality assurance](#) planning involves identification of key indicators of quality used in quality assurance. |- | [Quality audit](#) | The process of the systematic examination of the [quality](#) system of an organization by an internal or external quality auditor or team. It assesses whether the [quality management plan](#) has clearly outlined [quality assurance](#), [quality control](#), corrective actions to be taken, etc., and whether they have been effectively carried out. |- | [Quality checklist](#) | A checklist for [quality](#) identifies all the steps, procedures, and controls specified to ensure required procedures have been followed and their goals met. An example of a Translation [Quality Checklist](#) is the ESS Round 7 Translation Quality Checklist (European Social Survey, 2014c). |- | [Quality control](#) | A planned system of process monitoring, verification, and analysis of indicators of [quality](#), and updates to [quality assurance](#) procedures, to ensure that quality assurance works. |- | [Quality management plan](#) | A document that describes the [quality](#) system an organization will use, including [quality assurance](#) and [quality control](#) techniques

and procedures, and recommendations for documenting the results of those procedures. corrective actions taken, and process improvements made. **Frequency profile** | A comprehensive **probability sampling** survey design that provides information data users need to assess the data. **Question by question objectives** | Text associated with some questions in an

adapting the second stage of selection.

[https://wiki.capstan.be/docs.php?id=generalreference:survey\\_glossary&rev=1561592784](https://wiki.capstan.be/docs.php?id=generalreference:survey_glossary&rev=1561592784)

**adaptation** | The deliberate technical or substantive modification of some feature of a

neither agree nor disagree

**target population** (e.g., updating language: "radio" for "wireless"; adapting an **adapting the second stage of selection** for children: "tummy" for "stomach"; or **tailoring** for cultural needs: walk several blocks versus walk 100 yards). **Quota sampling** | A non-**probability sampling** method that sets specific sample size quotas or target sample sizes for subclasses of the **target population**. The sample quotas are generally based on simple demographic characteristics (e.g., quotas for gender, age groups, and geographic region subclasses). **Random route (Random walk)** | For each randomly-chosen sampling points (e.g., urban units, small cities, or voting districts), interviewers are assigned with a starting location and provided with instructions on the random walking rules – e.g., which direction to start, on which side of the streets to walk and which crossroads to take. Households are selected by interviewers following the instructions. The routes end when the predefined number of respondents (or households) is achieved (Bauer, 2016). Since the probability of the selected household is unknown, this method is categorized as non-**probability sampling** methods (Bauer, 2016).. **Random-digit-dialing (RDD)** | A method of selecting telephone numbers in which the **target population** consists of all possible telephone numbers, and all telephone numbers have an equal probability of selection. **Ranking format** | A response format where respondents express their preferences by ordering persons, brands, etc. from top to bottom, i.e., generating a rank order of a list of **items** or entities. Example: Listed below are possible disadvantages related to smoking cigarettes. Please enter the number 1, 2, 3, or 4 alongside each possible disadvantage to indicate your rank ordering of these. 1 stands for the greatest disadvantage, 4 for the least disadvantage.   
Harmful effects on other people's health   
Stale smoke smell in clothes and furnishings   
Expense of buying cigarettes   
Harmful effects on smoker's health

