CME

Data Collection Tools and Methods under Clinical Studies

*Sada Nand Dwivedi**

To generate quality evidence regarding informed decisions for clinical/ public health practice, quality data collection is an integral part of every clinical study, regardless of the involved research questions, hypotheses and objectives. As such, data collection is the systematic, purposeful process of information gathering, measuring, and analyzing accurate, targeted information from various sources to answer research questions, test hypothesized hypotheses, and support decision-making. The process of data collection might broadly consist of two components: tools of data collection (e.g., questionnaires) and methods of data collection (e.g., surveys, interviews, observations, focus group discussions, and others). The present write-up briefly describes them and their related aspects to make the readers aware of them while planning any clinical study.

Access this article online

Website: www.cijmr.com

DOI: 10.58999/cijmr.v5i01.376

Keywords: Data collection tools Data collection methods, Questionnaires, Surveys, Interviews, Observations, Reliability, Validity

Background

The data collection is the systematic, purposeful process of information gathering, measuring, and analyzing accurate, targeted information from various sources to answer research questions, test involved hypotheses, and support decision-making. Regardless of involved research questions, hypotheses and objectives, to generate quality evidence regarding informed decisions for clinical/ public health practice, quality data collection is an integral part of every clinical study.[1,2] The process of data collection might broadly consist of two components: tools of data collection (e.g., questionnaires) and methods of data collection (e.g., surveys, interviews, observations, focus group discussions, and others). These components may vary according to the planned type of research (e.g., quantitative research, qualitative research, or both).

Quantitative research aims to quantify the burden of a specific public health issue, its distribution and association with certain variables. Often, structured questionnaires are used to collect required data. On the other hand, qualitative research aims to identify and explore a number of related aspects that give insight into the nature and causes of certain problems for those affected with a considered public health problem.[1,3] This research provides qualitative information, which is often recorded in narrative form. Under qualitative research, flexible techniques, such as loosely structured interviews using open-ended questions, focus group discussion; in depth interviews and participant observation are used. Sometimes, a study may consider both quantitative and qualitative research. As such, the focus of the present write-up is on the quantitative research.

Due precautions need to be ensured to collect quality data. For example, the scale of measurements of various variables involved in the study needs to be decided optimally to avoid unnecessary biases in the collected data.[4,5] Further, intra-observer as well as inter-observer variation needs to be minimized through involvement of appropriate quality training/checks/ supervision. In other words, the reliability and validity of the collected data may ensure reliable and valid study results (described in the next section) through appropriate analysis of the collected quality data.

Keeping in view the above-mentioned points, the present write-up briefly describes data collection tools as well as methods, and their related aspects to make the readers aware, helping them while planning any clinical study.

International Centre for Health Research, RD Gardi Medical College, Ujjain, Madhya Pradesh, India

Correspondence to: Sada Nand Dwivedi, International Centre for Health Research, RD Gardi Medical College, Ujjain, Madhya Pradesh, India. E-mail: dwivedi7@hotmail.com

Submitted: 18/03/2026 Revision: 30/03/2026
Accepted: 15/04/2026 Published: 20/04/2026

How to cite this article: Dwivedi SN. Data Collection Tools and Methods under Clinical Studies. Central India Journal of Medical Research. 2026;5(1):4-9.

Reliability and Validity of the Study Results

Regardless of the considered research design, as a primary concern under a clinical study, the derived conclusions of the study need to be reliable and valid.[6-8]

Reliability of the Study Results:

As a measure of the consistency, stability, and reproducibility of research findings, the reliability of the study results refers to the degree same methods applied in similar circumstances produce the same results. In other words, someone else using the same method in the same circumstances should be able to obtain the same findings. To ensure its optimum level, it is obvious to use the best design of the study[6,7] and standardized procedures. As indicated earlier, its lack may also be because of the lack of reliability in the measurements/ observations.[9]

Validity of the Study Results

This is the degree to which the inference drawn from a study may be generalisable, extending beyond the study sample. To ensure this, one has to take into account the appropriate steps under research methodology, including data collection tools and methods, and the representativeness of the study sample in terms of the comparable characteristics of the study population from which it is drawn. In other words, the drawn conclusions from the study sample may be generalisable not only to the targeted study population but also to different settings/regions/ times with similar characteristics. Hence, this may further be categorized as follows:

Internal Validity

Internal validity refers to validity inside the study. For example, under a study involving comparison between two groups, the study design and appropriate analytical methods should be used in such a way that, apart from sampling error, the differences observed between the study groups may be attributed only to the hypothesized exposure under investigation. In other words, internal validity may be achieved mainly through controlling non-sampling errors (i.e., inaccuracies, biases and mistakes). To be more specific, internal validity of a study using a representative sample from the study population refers mainly to technical, structural, and methodological accuracy, deriving true results for the considered sample, leading to its Generalizability among the target study population.

External Validity

A study is said to be externally valid if it can produce unbiased inferences for the targeted study population and generalisable beyond the subject in the study. For example, a study conducted among a male sample might or might not be generalisable to all people (including female subjects). Along with internal validity, the evaluation of Generalizability also involves subject matter judgment. In other words, for assessing external validity, one has to determine whether internally valid observed study results may be applicable to different populations, regions and times.

Instruments for Measuring Data

There are mainly three instruments to measure the data.

Human

The required information may be asked/observed and recorded by a person with little or no involvement of apparatus. For example, grading spleen enlargement, and taking a patient's social/ clinical history.

Apparatus

Under clinical studies, a mechanical device is often used to measure the required data. For example, the use of weighing scales to measure weight; use of a thermometer to measure temperature during fever.

Combination of Human and Apparatus

Sometimes recording of data involves both human as well as apparatus. For example, the use of chest X-ray films to investigate tuberculosis, and the reading of blood films to microscopically evaluate stained blood cells.

Tools of Data Collection

For data collection as per planned objectives, various types of forms/ proformae are used, ensuring their appropriateness to provide complete, accurate, consistent, reliable and valid data. They are broadly referred to as:

Interview Schedule

An interview schedule consists of a meticulously planned set of questions aiming to provide the intended information completely and accurately.

Questionnaire

A questionnaire consists of a predetermined set of questions used to collect data on required aspects (e.g., socio-economic status, occupational group, clinical data, and others). In contrast to an interview schedule, this term is often used to specify a self-completed survey instrument.

Survey Instrument

The interview schedule, questionnaire, medical examination record form, and other forms used in a survey are generally referred to as a survey instrument.

Questions Helpful In Designing Tools for Data Collection

While designing a tool for data collection, one has to have answers to all relevant questions, like:

- According to the planned objectives and variables identified during proposal writing, what
- exactly does one want to know?
- Is questioning the right approach to obtain all answers?
- Does one need additional approaches like observations or analysis of records?
- Who will be the interviewer/ respondent, and what approach will be used?
- Does it require only structured questions involving predefined response options (i.e.,
- closed-ended), or only unstructured questions allowing free-form answers (i.e., open-
- ended); or both?
- Do the respondents include illiterates? If yes, the self-administered questionnaires may
- not be a feasible option.
- What should be the required minimum sample size to answer the planned research
- questions? Under studies involving a large number of respondents, the use of shorter and
- highly structured questionnaires may be a better option. On the other hand, a number of
- open-ended questions may be used under smaller studies.

Types of Involved Questions under the Tools for Data Collection:

As indicated earlier, there are mainly two types of questions involved in finalizing a survey instrument (e.g., questionnaire):

Closed Questions:

While answering the closed questions, the respondents have to choose a correct answer out of a set of listed possible answers for each question. The listed answers/ options under a closed question remain exclusive and mutually exclusive. As such, closed questions may be useful if the range of possible responses is known, for example**:**

(i) What is your educational status? a. Illiterate; b. Literate & up to primary; c. above primary and up to senior secondary; d. above senior secondary, diploma and up to under-graduation, including professional degrees; e. above under- graduation, including professional degrees.
(ii) The use of closed questions might be helpful in optimal utilization of the time of the respondent as well as the interviewer through avoiding the collection of more information than is needed. For instance, to know the consumption of milk products by an individual, one may simply ask: Did you consume any of the following items yesterday?

• Milk	Yes	No
• Curd	Yes	No
• Ghee	Yes	No
• Paneer	Yes	No
• Others (Kheer, etc.)	Yes	No

(iii) Using closed questions, the respondents may be asked to express their opinions by choosing rating points on an ordinal scale (i.e., Likert-type item). For example, MBBS students finishing a community medicine course may be asked, do you agree with the statement: Overall, the covered contents under the community medicine course are very useful. Further, the student may be advised to choose one of the five ratings listed below:

- Strongly disagree
- Disagree
- Sometimes agree
- Agree
- Strongly Agree

Data collection using closed questions may be comparatively quicker. Further, Analysis of collected data remains to be easier. However, it remains difficult for illiterate respondents. Also, comparatively, it may involve a chance of bias and respondents may lose interest after answering a few questions. To ensure quality and accurate data collection, as in the case of open-ended questions, pre-testing of the developed survey tool is necessary. Further, an instruction manual for data collection is a must. Also, its use may be combined with open-ended questions (explained in section 5.3) and sequencing of the questions might also be useful.

Open-Ended Questions

Responses under open-ended questions are recorded in the respondent's own words, that is, as the respondents say. In contrast to closed questions, under open-ended questions, the respondent is not given any possible options to choose from. Such questions often provide newer information. To be more specific, it reveals facts

with which the researcher may not be very familiar. Further, open-ended questions are often used to seek their opinion, attitudes, and suggestions on sensitive issues like: What do you think are the reasons for a low contraceptive adoption in villages? What would you do if you noticed that one of your children has become a drug addict while still in school?

In contrast to closed questions, open-ended questions may sometimes provide knowledge about unknown issues. It further provides quicker response, which comparatively remains more valid too. It also provides additional information for inclusion in the related reports. However, data collection using open-ended questions needs professional interviewers. Further, analysis of collected data remains comparatively complex and needs experienced data analysts/ advanced computer programs. To ensure quality, accurate data, pre-testing of the finalized questionnaire and also specialized training of involved interviewers are a must. To be more specific, experienced interviewers may be more appropriate for such data collection.

Combination of Closed and Open-Ended Questions

Sometimes an open-ended question may be built up under a closed question. This allows collecting quantitative data along with qualitative data. In other words, first, data is collected on specific facts, and then on deeper qualitative insights. For example, why did you not go to a health facility regarding pregnancy registration for antenatal check-ups?

- Lack of knowledge A
- No health worker visited B
- Family did not allow C
- Too far from village D
- Poor quality service E
- No time to go F
- No transport G
- Too much cost H
- Not necessary I
- Not customary J
- Others (Specify) K

To be more specific, such a combined question will be able to provide specific facts about the listed points from "A to J" along with related deeper insights of the respondent under point "K".

Steps in Finalizing a Questionnaire for Data Collection:

There are broadly three ways of finalizing the survey tool (i.e., questionnaire) for data collection:

Use of Already Available Validated Questionnaire

Keeping in view the planned research question/ objective and the required data to answer this research question, while trying to finalize a tool for data collection, we might come across various related tools already published in the literature. If any of the existing tools is able to serve the purpose under the planned study, one can use the same without the need to design and validate the tool. However, if there is copyright, permission from the corresponding author has to be obtained.

Use of Modified Version of Already Available Validated Questionnaire:

In contrast to section 6.1, if any of the existing tools may be used under the planned study after due modification/ changes, one can use the modified version of the available tool without the need to design and validate the tool. However, in view of the need for the involved local environment under the planned study, required modifications in the existing tool might vary, such as a change of only language, making the existing questions more lucid and clear through cosmetic changes, and addition/deletion of questions. In this case also, if there is copyright, one has to take due permission from the corresponding author. For completeness, if major modifications are involved in the existing tool, one also has to validate the modified version of the tool before its use in data collection.

Use of a Self-Designed Questionnaire:

Often, the required validated questionnaire may not be readily available in the literature. Also, if available, they may not be appropriate to be used even after due modification in a given setup. Hence, there are occasions/situations when the researcher has to design a new questionnaire and validate it before its use in data collection. The design of a questionnaire requires a number of drafts. Further, as is obvious, it consists of various steps. The major steps in designing a questionnaire are:

Step I: Contents

Keeping in view the planned research questions/ objectives and required data to answer these research questions, while designing a tool for data collection, one has to consider objective-specific variables in their original scales of measurement. To complete the planned objectives under the study, one has to further decide on optimal questions that will be needed to measure those variables. The collected data on all the required

variables need to maintain their raw forms. In other words, no change in their original scales of measurement is required at the time of data collection. This will allow understanding the distributions of all variables and relationships with each other fully, and deciding the final scales of measurements of the considered variables optimally for further data analysis. As is obvious, while developing the questionnaire, some of the variables may get added, dropped, or changed, if and when necessary, at various stages. Sometimes, due to intuitive realization during questionnaire development, the objectives of the study may also be modified.

Step II: Formulating Questions

Regarding the formulation of possible questions in a questionnaire that may provide the required information for each variable, one has to choose questions that are specific and precise enough. For instance, one may avoid asking: "Where do you usually seek treatment when you are sick?" If necessary, questions may be broken into different parts. For example, instead of the above question, one may focus on illnesses that have occurred in the family over the fortnight/month and ask what was done to get them treated, each of them from their onset/diagnosis. Further, each question needs to measure one aspect at a time. For instance, one needs to avoid asking like: "How many children would you and your wife prefer to have?" Also, leading questions that are suggestive of a certain answer need to be avoided. For example, do you agree that the state health officers should make surprise visits to each district health center monthly? A complex question also needs to be avoided; better to use simpler/ neutral questions. Also, better to avoid words with double or vaguely defined meanings.

Step III: Sequencing of Questions

The interview schedule or questionnaire needs to be designed to be consumer-friendly. Further, to ensure quality data, prefer to develop the questionnaire as short as possible. If necessary, one may plan to conduct the interview in two or more parts. For instance: National Family Health Survey consisting of various sections including Household Questionnaire, Women's Questionnaire (e.g., background characteristics; antenatal care & birth preparedness; contraception; knowledge, opinion and attitude on sexually transmitted infections/ reproductive tract infections and others); and Youth Questionnaire (male/ female).

Guide for Interview

For this, a detailed instruction manual for each item within every section of the questionnaire is an integral part of any developed questionnaire. There is a need to accommodate some of these instructions in the questionnaire itself. One needs to specify about reading/not read the alternative answer in the manual. Also, special mention regarding the required probing should be made. Criteria to decide not availability of the respondent should be explained. Further, the need and procedure for the consideration of substitutes for such respondents should be explained. It is better to elaborate on the required strategies to deal with various anticipated problems during the interview.

Data Collection Techniques

Various techniques may be applied in data collection, such as

- One can make use of available information. It might comparatively be inexpensive, but it often has problems with its availability, completeness, and methodology.
- One can observe oneself, which can provide additional and often accurate information (e.g., respondent's behavior) but often time-consuming.
- One can interview the eligible respondent face-to-face. For this, as indicated earlier, trained researchers are required. Often, it is preferred in the case of the involvement of a large number of respondents.
- One can administer the developed questionnaire for self-reporting by the respondents. For this, questionnaires may be delivered by hand to respondents and collected later. Also, questionnaires may be sent to the respondents by mail with clear guidelines to fill out and requesting mailed responses. Sometimes, if feasible, sub-groups of the respondents may be requested to assemble in specific places at a particular time. They may be given oral and/ or written instructions to fill out the questionnaires on their own and return them.

Data Collection Tools

Various tools may be used in data collection, like pen and paper, eyes and other senses, watch, scales, microscopes/ other instruments, checklist, data-compilation form, Interview schedule, questionnaire, and tape recorders/ tablets.

Systems for Collecting Data

There are broadly two systems for data collection:

Regular Systems of Data Collection

This is nothing but a routine system of data collection, which usually consists of a registration system for collecting data on an aspect (e.g., event) as and when they become available. For example:

- A registration system for patients attending hospitals/ health care facilities/ clinics
- A vital statistics (e.g., marriage, migration, births & deaths) registration system
- A disease notification system (e.g., leprosy, tuberculosis)
- A reporting system for cancer cases (e.g., population/ hospital-based cancer registries)

For the regular systems of data collection, rules and regulations instituting a particular system are established, giving legal backing to it (e.g., national/ state level systems). They consist of approval of items on whom information has to be collected; designs of forms and registers to be used for recording information; physical establishment of office facilities; recruitment of personnel and their training; the recording procedure- who supplies the information, when the information has to be registered; and specification and design of registration receipts. Our country has established many such regular systems of data collection, like the census, the civil registration system, the sample registration system, and others.

Ad hoc Systems of Data Collection

In contrast to regular systems of data collection, ad hoc systems of data collection[10] usually consist of a survey to collect information regarding a specific research purpose, which are not available completely under an existing regular system of data collection. As is obvious, ad hoc systems may involve special investigative studies or an elaboration of some aspects of data that have been collected under a regular system. The analytical results of collected data may be able to fulfill the planned research objectives along with their administrative implications. For Example, a survey to estimate the proportion of pregnant mothers with malnutrition in a defined population, and an investigation of breastfeeding practices among mothers who delivered a live birth in the previous two years in a region. To maintain the quality of data, avoiding various types of biases, the ad hoc surveys are usually planned and carried out following various steps under related research methods.

Summary

The data collection is an integral part of any research study[11,12] hence related tools and methods play an important role in strengthening the reliability and validity of the collected data. Keeping in view of feasibility, one needs to prioritize considering optimized tools and methods of data collection on each variable in raw form, ensuring opportunities to understand their distributions and associations with each other fully before finalizing their optimal scales of measurements for final analysis of data. It may be worthwhile to mention here that going through study-design-specific available guidelines[13] for preparing study reports may help at each step of research methodology[11], including data collection.

References

1. WHO. Health Research Methodology (A Guide for Training in Research Methods). 1992. Oxford University Press, Delhi.
2. KR Sundaram, SN Dwivedi, V Sreenivas. Medical Statistics: Principles & Methods. Wolters and Kluwer (Health), New Delhi. 2015 (Second Edition).
3. Altman Douglas G. Practical Statistics for Medical Research. Chapman & Hall/ CRC. 1991. ISBN 0-412- 27630-5.
4. Dwivedi SN. Understanding of scales of measurement and their consideration in clinical research. Central India Journal of Medical Research. 2023; 2(1):13-16.
5. Dwivedi SN. Biases, including confounding and effect size modification in clinical research and their mitigations. Central India Journal of Medical Research. 2024; 3(2):6-12.
6. Armitage P, Berry G, Matthews JNS. Statistics in Medical Research. Fourth Edition. Wiley Inter Science. 2008. ISBN: 9780632052578.
7. Dwivedi SN. Understanding of study designs in clinical research: Major prompting points. Central India Journal of Medical Research. 2024; 3(3):12-15.
8. Gomes J, Pandey S, Pandey R. Reliability and Validity in a Research Study. Research Methodology (Methods & Techniques). 2nd Edition, New Age International Publishers. 2004.
9. Dwivedi SN. Understanding of scales of measurement and their consideration in clinical research. Central India Journal of Medical Research. 2023; 2(1):13-16.
10. NFHS-5 (2022). National Family Health Survey (NFHS-5), 2019-21. Final Report. International Institute for Population Sciences, Mumbai.
11. Armitage P, Berry G, Matthews JNS. Statistics in Medical Research. Fourth Edition. Wiley Inter Science. 2008. ISBN: 9780632052578.
12. Dwivedi SN. What are the major research methodology steps under a clinical study? Central India Journal of Medical Research. 2023; 2(2):4-9.
13. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Bull World Health Organ. 2007;85(11):867-72.