[abstract]

This article describes the basics tenets of quantitative research. It addresses the concepts of dependent and independent variables and explores the concept of measurement and its associated issues such as error, reliability and validity. Experiments and surveys are the principal research designs in quantitative research. These are described and their key features explained. The importance of the double-blind randomised controlled trial is emphasised, alongside the importance of longitudinal, as opposed to cross-sectional, surveys. Essentials features of data storage are covered, with an emphasis on safe, anonymous storage. Finally the article explores the analysis of quantitative data, considering what may be analysed and the main uses of statistics in analysis.

[author] Roger Watson Editor-in-chief, Journal of Advanced Nursing and Professor of nursing, Faculty of Health and Social Care, University of Hull, Hull, England. Correspondence to: R.Watson@hull.ac.uk

[text]

QUANTITATIVE RESEARCH encompasses a range of methods concerned with the systematic investigation of social phenomena, using statistical or numerical data. Therefore, quantitative research involves measurement and assumes that the phenomenon under study can be measured. Quantitative research sets out to gather data using measurement, to analyse this data for trends and relationships and to verify the measurements made.

Some items are easy to measure, such as height and weight; other items, such as what people think or how people feel, are difficult to measure. Quantitative research encompasses this entire spectrum. Similar criteria are applied to verify, calculate and analyse data for all types of measurement. Quantitative research may be considered a way of thinking about the world. It is essentially deductive: measurements are made; analysis is applied; and conclusions are drawn. It is pointless to dispute whether quantitative research is superior to qualitative research. The researcher may even choose to use both quantitative and qualitative methods in his or her research design, in a combined or mixed-methods approach (Andrew and Halcomb 2009. The mixed methods approach will be addressed in a later article in the series. A unique feature of quantitative research is its ability to formally test theories by formulating hypotheses and applying statistical analyses (Figure 1).

[Start figure]

[Figure head] Figure 1
Quantitative research designs

Theory

Hypothesis

Operationalisation of concepts*
Selection of participants

Survey design   Experimental design
(Interviews/ questionnaire)   (Treatment and control groups)

Collect data

Analysis

Results
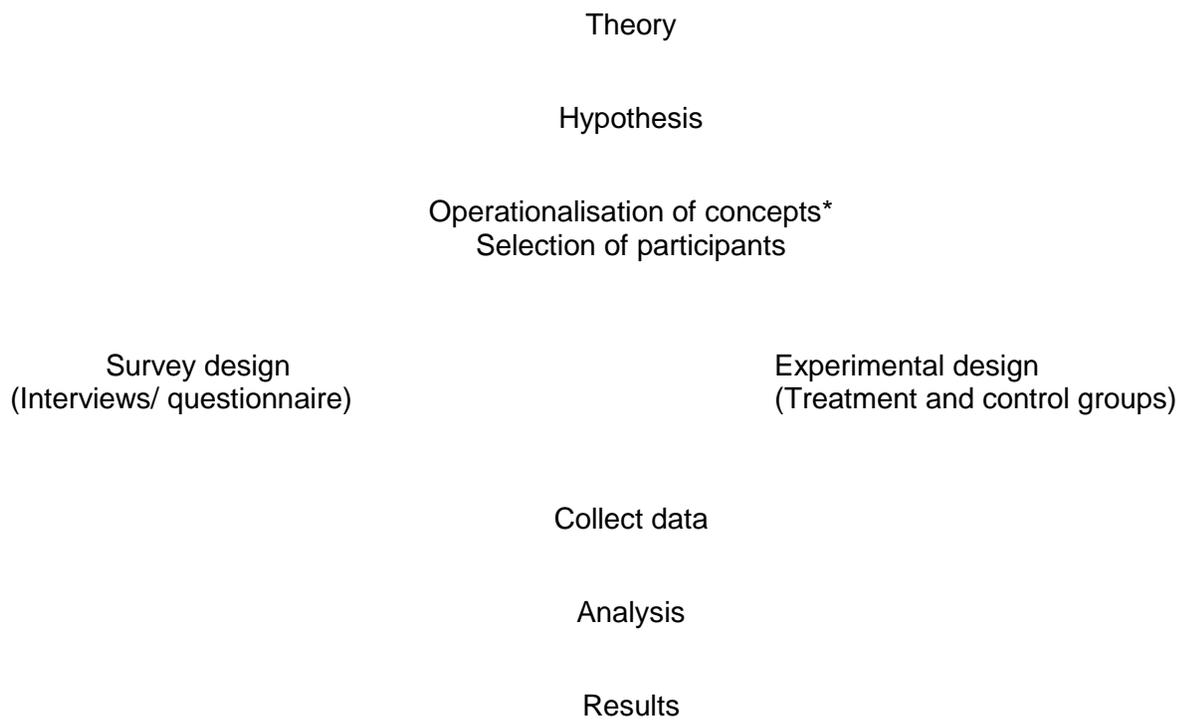
(Adapted from Bryman and Cramer 2005) * this means deciding which are the independent and dependent variables and how to measure the latter

[arrows to be added]

[end figure]

[A head] Variables

A variable is anything that may be measured in quantitative research, for example height, weight, attitude or wellbeing. There are two types of variable, independent and dependent (Pierce 2013). An independent variable is one that may influence the measurement of the dependent variable. For example, if you were studying the relationship between the frequency of positional change and the development of pressure ulcers, then positional change would be the independent variable and pressure ulcer development would be the dependent variable.

[A head] Measurement

There are different kinds of measurement, which can be placed in a hierarchy, using a theory of measurement (Stevens 1946). The different levels of measurement and their properties are shown in Table 1. Nominal measurement is the lowest level on the hierarchy because it is essentially a system of classification, rather than measurement. Ordinal measurement begins to order phenomena, but this measurement is limited and imprecise. Interval and ratio level measurements provide precise and accurate measurements. However, it is rarely possible to make ratio level measurements in quantitative research, which involves the study of social phenomena. Generally, measurement in quantitative research is made at the ordinal and interval levels of measurement.

[start table]

[Table head] Table 1 Levels of measurement

Measurement level	Attributes	Examples
Ratio	A zero value is meaningful, permitting direct comparisons between measurements. (For example, twice as many patients were seen in one month compared to the previous month.)	Height, weight, length.
Interval	Distance between measured variables is meaningful.	Temperature scales where the zero point is arbitrary, but set intervals are meaningful (for example, Centigrade or Fahrenheit).
Ordinal	Attributes can be ordered.	Opinion measured by asking if you: 'strongly agree', 'agree'; 'don't know'; disagree'; 'strongly disagree'.
Nominal	Attributes are only named (weakest level).	Hair colour, gender, nationality.

(Adapted from http://www.socialresearchmethods.net/kb/measlevl.php)

[end table]

[B head] Error in measurement

There is always error associated with measurement, by whatever means measurements are made. These sources of error apply to physical measurements, such as height and weight, and also apply to other types of measurements in the social sciences. Error may come from several sources in measurement (Shields & Watson 1999): instrument error, human error and random error. It is possible to minimise instrument error and human error, but it is not possible to control for random error. Random error should be allowed for in the design and use of any instrument. In the social sciences, an instrument may be a questionnaire or observational checklist. Instrument and human error may be of two kinds:

- Within instrument (or within human)

- Between instrument (or between human).

Within instrument errors mean that the same instrument can give different measurements on different occasions. Between instrument errors mean that two seemingly identical instruments can give different measurements. Similarly, within human errors mean that an individual using the same instrument can obtain different measurements on different occasions, while between human errors mean that two people using the same instrument can obtain different measurements on the same occasion.

Error cannot be entirely eliminated but steps should be taken to minimise it (Bowling 1997). Good instruments should be designed to minimise instrument error. In social research, this means ensuring that questionnaires and observational checklists are clear and easy to understand, and that the questions asked only address the phenomena that are being studied. For example, if you are interested in measuring difficulty with feeding in older people with dementia, then you should ask questions which address that problem alone and omit any questions that address other aspects of behavior such as agitation or wandering. In designing instruments, a balance should be struck between 'authenticity' and 'directness' (Messick 1994). An authentic instrument measures as much as possible about a phenomenon, at the risk of becoming indirect, while a direct instrument focuses on only the items directly concerned with the phenomenon, at the risk of losing some authenticity.

[B head] Reliability and validity

Reliability and validity involve estimating - and minimising - the level of error associated with measurements made using a given instrument (Streiner and Norman 2008). Reliability is the extent to which an instrument makes the same measurement each time it is used. Validity is the extent to which the measurement made by an instrument measures what we are interested in.

It is useful to consider physical measurement to explore these concepts. For example, if we measure a patient's blood pressure several times with a blood pressure monitoring device, we should get approximately the same measurement each time we use it, provided that the patient's blood pressure has not changed. Allowing for human error, the measurements would be reliable. Now imagine that the blood pressure monitoring device is faulty, so that it measures blood pressure a few millimeters of mercury below the true value. If we take successive measurements of the patient's blood pressure, we shall still get the same measurement each time we use the device. However, the measurements will be wrong, because they are not valid. This illustrates an important point, which is that measurements can be reliable but not valid. However, for measurements to be considered valid, they should be reliable.

Reliability and validity can be tested and improved by making adjustments to instruments, if the levels of reliability and validity are too low. With questionnaires this usually involves revising the items in the questionnaire, removing or clarifying ambiguous questions. The principal features of the different types of reliability and validity are provided in Table 2.

[start table]

[Table head] Table 2

Definitions and types of reliability and validity

Reliability	Definition
Internal consistency	The extent to which all the items in a questionnaire measure the same thing.
Test: re-test reliability	The extent to which an instrument (such as a test) gives the same result on two occasions.
Intra-rater reliability	The extent to which the same person obtains the same measurement on two occasions.
Inter-rater reliability	The extent to which two people obtain the same measurement.
Validity	Definition

Criterion validity	The extent of correlation with another validated measure.
Concurrent validity	The extent of correlation with another measure of the same phenomenon at the same time.
Predictive validity	The extent of correlation with another measure at a later time.
Convergent (divergent) validity	The extent of correlation (or lack of correlation) with measures of another phenomena predicted to correlate (or not to correlate) with the new scale.
Discriminant validity	The ability to discriminate between cases, such as severe and mild, between cases and non-cases.

[end table]

[A head] Quantitative research designs

There are two broad categories of research design in quantitative research, experimental designs and survey designs (Figure 1).

[B head] Experimental designs

An experiment is a study where the researcher can manipulate one variable, the independent variable, and study its effect on a dependent variable. For example, if you wished to study the effect of the dose of an analgesic on pain levels, you could vary the dose of the analgesic (the independent variable) and measure the effect on the pain level (the dependent variable). There are many types of experiment. For the purposes of this article we shall focus on the randomised controlled trial experimental design, used to test the effect of treatments on people.

 [C head] The randomised controlled trial

The randomised controlled trial is considered to be the best method for testing the link between cause and effect in clinical interventions. Its essential features are randomisation and use of a control group. The randomised controlled trial is rated near the top of the hierarchy of evidence, at level II, as a method of providing evidence for clinical practice (Centre for Reviews and

Dissemination 2009). It is superseded in the hierarchy, only by systematic review with meta-analysis, a method of combining the results of randomised controlled trials and evaluating the combined evidence.

Randomised controlled trials should preferably be 'blind': either those taking part do not know if they are in the intervention group or the control group, or the person who is collecting the data does not know this. The optimum design is the double-blind randomised controlled trial where neither the participants, nor the person collecting the data, know who is in which group (Smith 2008).

The simplest form of randomised controlled trial requires at least two groups of participants: a treatment (also referred to as experimental or intervention) group and a control group. The treatment group receives the treatment being tested and the control group does not. However, the control group should be treated in exactly the same way as the treatment group, or as closely to this as is possible, except that they do not receive the treatment. When testing drugs, for example, this is achieved by administering a 'placebo', which looks identical to the treatment drug, except that it contains no active ingredient. This matching of treatment in the two groups, as far as is possible, is to take into account the 'placebo effect', whereby anyone involved in a randomised controlled trial - whether receiving the treatment or not - may respond as if they were being treated. The placebo effect must be the same in both groups for the effect of the active drug to be measured correctly. With nursing interventions, for example testing a support surface for pressure ulcer prevention, it may be difficult to provide a placebo in quite the same way. In such cases it is customary to administer the usual care that a person may receive for pressure ulcer prevention to the control group, and to compare this usual standard of care with the new support surface being tested.

There are many possible methods of allocating individuals who have agreed to participate in the randomised controlled trial to either the treatment group or the control group. Randomisation is used to minimise bias in allocating individuals to the two groups. For example, we could be accused, sub-consciously or even deliberately, of allocating all the people who are more likely to respond to treatment to the treatment group and the remainder to the control group. This might introduce bias into the experiment, which could exaggerate the effects of our treatment.

Blinding, as explained above, is a process of concealment and can be either single blind or double blind. The purpose of blinding is to minimise bias in either the researcher or the participant, or both, by concealing to them that they are receiving the treatment or the control (in the case of the participant) and/ or which participants are receiving the treatment or the control (in the case of the researcher). Double blinding is preferred, but this is difficult to achieve with nursing interventions.

[B head] Surveys

Surveys are often used in nursing research. These frequently involve distributing questionnaires; but they may also be conducted by interview or by observation. In contrast to experiments, surveys cannot easily distinguish between cause and effect, but they are useful for gathering large amounts of data to describe samples and populations (Hallberg 2008). Surveys may be either cross-sectional or longitudinal. Cross-sectional studies are relatively easy to conduct, as they only have to be done once. Longitudinal studies are more complex, especially ones conducted over several years, as they require repeated surveys. Attrition is a significant problem with longitudinal studies (Aldridge and Levine 2001).

There are three different types of longitudinal survey design: trend studies, cohort studies and panel studies (Watson 2008). They each have their advantages and disadvantages. Trend studies are concerned with trends in a population. A classic example is the study of voting

intentions in the run-up to general elections. The population is sampled on one day and again, at intervals. The sample surveyed is always part of the same population but does not, necessarily, comprise the same people. Therefore, this is a relatively simple type of survey to perform, but does not provide information about how specific individuals change over time.

Cohort studies and panel studies are similar in that they use the same sample group at each stage—as opposed to different people at each stage—but they differ slightly in how they use the groups. In a cohort study, the study uses a defined cohort (a group of people with a shared characteristic). The people surveyed at each stage, for example, could all belong to the nursing class of 2013, but the same individuals may not be surveyed each time; each group surveyed will be a sub-sample of the defined cohort, ie the nursing class of 2013. In contrast, in a panel study, exactly the same people are surveyed at each stage. Therefore, cohort and panel studies are more informative about how individuals change over time than trend studies, but are more difficult to conduct and are susceptible to attrition.

[A head] Handling data from quantitative studies

Quantitative studies produce numbers which should be interpreted before conclusions may be drawn. These numbers (the data) may be entered, stored and analysed using some form of electronic database. Data entry may be into a Word© document or an Excel© Spreadsheet, for example. Some initial data analysis is possible in Excel©, but data may be imported into a statistical package, such as SPSS© (Statistical Package for the Social Sciences), to permit more sophisticated analysis (Pallant 2007). Data entry often requires transcription from hard copies of questionnaires or observational schedules. This has to be done carefully and double-checked. Increasingly, surveys are distributed via the internet and data can be imported directly into an analytical package such as SPSS©. It is essential to store data carefully, once it has been entered into any package, as loss of data may jeopardise the study. It is good practice to create

a master copy of the data, which should not be altered but may be copied in the event that subsequent files are lost or inadvertently altered. It is essential to create and store safely a backup copy of the master copy in case it is inadvertently deleted or irretrievably altered. Security is important if data are sensitive or confidential. Files should not contain any information that could identify individual participants. Data should be stored safely and password protected.

[B head] Statistical analysis

Quantitative data may be analysed statistically (Watson *et al* 2006). Data may be described in terms of percentages, central tendency (mode, median, mean) and spread (range and standard deviation). These terms are explained in Table 3. Analysis of the data in the sample may be used to draw inferences about the population as a whole. Analysis is usually performed using a set of analyses known as inferential statistics. These allow you to investigate, for example, the differences between the mean values in the treatment and control groups in a randomised controlled trial and to investigate the associations between variables such as pain and analgesic dose. The important criteria in inferential statistics is whether something is statistically significant. Statistical significance is usually expressed as a probability, which measures 'How likely was this to happen anyway?' If the probability is very low, conventionally below 0.05 (less than a 1 in 20 chance), then we are justified in stating that our observation is statistically significant at this probability. Statistically significance implies that the observed benefits are likely to have happened as a result of the treatment being tested, or that the observed relationship between variables is real.

[Start table]

[Table head] Table 3 Definition of statistical terms

Term	Definition

Central tendency	A description of where the central point in a dataset is.
Mode	The most frequently occurring number.
Median	The number in the middle of a set of data.
Mean	The number obtained by adding up all the numbers in a dataset and dividing by the number of numbers.
Spread	A description of how widely the data diverge from the central tendency.
Range	The difference between the largest and smallest values in the dataset.
Standard deviation	A measure that describes the 68% of the data either side of the mean in a normal distribution.

 [end table]

[A head] Conclusion

This article described the main principles of quantitative research, such as variables, measurement, error, reliability and validity and explores the two principal research designs - experiments and surveys. Instruments should be designed to ensure that they have good reliability and validity. Random error cannot be eliminated in quantitative research. Instrument and human error can be eliminated or reduced. Experiments and surveys are used to study the relationship between variables. Experimental designs are best for relating cause and effect. The most effective experimental design is the double-blind randomised controlled trial. Surveys are

most useful for studying people and populations. The vast majority of surveys are cross-sectional, but the best survey designs are longitudinal as these can be used to study changes in people and populations. In quantitative research, it is important to correctly collect data and store it securely on electronic databases and to analyse quantitative data, using appropriate statistical methods.

[Reference head] References

Aldridge A, Levine K (2001) *Surveying the Social World: Principles and Practice in Survey Research.* Open University Press, Buckingham.

Andrew S, Halcomb L (2009) *Mixed methods research for nursing and the health sciences* Wiley-Blackwell, Oxford

Bowling A (1997) *Research Methods in Health.* Open University Press, Buckingham. [Q: this is now available in a fourth edition: http://www.amazon.co.uk/Research-Methods-Health-Investigating-services/dp/0335262740/ref=dp_ob_title_bk Bryman A, Cramer D (2005) *Quantitative Data Analysis with SPSS 12 And 13: A Guide for Social Scientists.* Taylor and Francis, London.

Centre for Reviews and Dissemination (2009) *Systematic Reviews: CRD's guidance for undertaking systematic reviews in health care* CRD, University of York

Hallberg IR (2008) Surveys. In Watson R, McKenna H, Cowman S, Keady J (Eds) *Nursing Research: Designs and Methods.* Churchill Livingstone, Edinburgh, 179-188.

Messick S (1994) The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher.* 23, 2,13-23.

Pallant J (2007) *SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using SPSS for Windows Version 15.* Third edition. Open University Press, Maidenhead.

Pierce E (2013) Designing and conducting quantitative research studies. In Curtis EA, Drennan J (Eds) *Quantitative Health Research Methods: Issues and Methods.* Open University Press, Maidenhead,131-150.

Schneider Z, Elliott D, LoBiondo-Wood G, Haber J (1999) *Nursing Research: Methods, Critical Appraisal and Utilization.* Mosby, Sydney.

Should this reference be updated to the 3e? I was unable to check the contents of the 3e but the proposed 4e contents still has something on measurement error, https://www.elsevierhealth.com.au/media/anz/samplechapters/9780729541374/Nursing%26MidwiferyResearch4e_9780729541374_Schneider_Whitehead_SampleChapter_lo_res.pdf]

Shields L, Watson R (2013) Common quantitative methods In Schneider Z Whitehead D, LoBiondo-Wood G, Haber J (eds) *Nursing and midwifery research: methods and appraisal for evidence-based practice* 4th edition Mosby, Sydney pp 161-183. Smith G (2008) Experiments. In Watson R, McKenna H, Cowman S, Keady J (Eds) *Nursing Research: Designs and Methods.* Churchill Livingstone, Edinburgh, 189-198.

Stevens SS (1946) On the theory of scales of measurement. *Science.* 103, 2684, 677–680.

Streiner DL, Norman GR (2008) *Health Measurement Scales: A Practical Guide to Their Development and Use.* Fourth edition Oxford University Press, Oxford.

Watson R, Atkinson I, Egerton P (2006) *Successful Statistics for Nursing and Healthcare.* Palgrave Macmillan, Basingstoke.

Watson R (2008) Longitudinal studies. In Watson R, McKenna H, Cowman S, Keady J (Eds) *Nursing Research: Designs and Methods.* Churchill Livingstone, Edinburgh, 261-268.