

Statistical terms Part 2: Principles of research study design: Understanding the options, indications and limitations.

SADJ August 2016, Vol 71 no 7 p319 - p323

LM Sykes¹, F Gani², HD Dullabh³

INTRODUCTION

Consider this proposed study.

Aim: *This cohort case study will use a causal, cross sectional design, based on historical records, to explore and describe the long-term, sequential effects of eating.*

H₀: *This study will provide evidence to disprove the null hypothesis that "Of those who acquire the habit of eating, very few survive."*

By the end of this paper, it is hoped that you will understand each of the terms mentioned above, as well as how and when to use them depending on the study design. You may also then realize how ludicrous it is to use repetitive, redundant, superfluous, verbose, illogical jargon – not to mention that this study is impossible to carry out! The only hypothesis that could be proven with a hundred percent confidence is the alternative (H_1), namely, the well-known cliché that "A little knowledge is dangerous."

Research begins with the formulation of a clear question, collection of evidence focused around that specific problem, analysis of the results, followed by a critical appraisal of their validity (closeness to the truth) and relevance (importance and usefulness).¹ In order to do this, the study design needs to be appropriate for the specific problem, bearing in mind the levels of achievable evidence.

This paper will explain the different research study designs, highlighting **Uses** and **Limitations** of each, to help researchers select the one most appropriate for their needs. However, it is good to keep in mind where each fits into the "The Evidence Ladder" (Table 1), as that will affect the strength of the results which are to be published.²

Study design refers to the plan used to address a research problem.¹ The nature of the problem should guide

the selection, and will determine which model will most effectively obtain the evidence needed to answer the designated question.¹ It comprises the blueprint that will be used for data collection, measurement and analysis. In general, clinical research can be experimental or observational.

In *experimental* enquiry, the researcher has control over some form of intervention. In *observational* enquiry, the research participants / patients are observed at a specific time (*cross-sectional*) or over a time period (*longitudinal*). Where the observation looks forwards to gather new data it is a *prospective* study, while those that use existing data (e.g. old records) are *retrospective* studies.²

Designs can vary considerably, but regardless of which is chosen, it should achieve the following objectives:

1. Identify the problem clearly.
2. Review, synthesize and critically analyze relevant published literature.
3. Specify a clear research question related to the problem (the hypothesis).
4. Describe which data is needed to test the hypothesis, and how the data will be obtained.
5. Justify which methods of analysis will be used to test the hypothesis.¹
6. Critically appraise the evidence, assessing its validity (closeness to the truth) and relevance (importance and usefulness).² Draw meaningful conclusions and recommendations, based on sound judgment as well as statistical calculations. "Let us not use statistics like a drunk uses a lamp-post, more for support than illumination" (Romano Prodi).³

ACTION RESEARCH

Once a problem has been identified, some form of intervention (action) is carried out during which time observations are made pertaining to the outcomes. The intervention may be repeated in cycles over time until there is sufficient understanding of the problem.

Uses: Projects based on this type of research have no controls and are most suited to community situations, where the focus is on finding implementable solutions rather than testing theories. As Heinrich Heine observed "You cannot feed the hungry on statistics".³

1. **Leanne M Sykes:** BSc, BDS, MDent (Pros). Department of Prosthodontics, University of Pretoria.
2. **Fatima Gani:** BDS, MSc. Department of Prosthodontics, University of Pretoria.
3. **Hemant D Dullabh:** BChD, MSc (Dent), MDent (Pros). Department of Prosthodontics, University of Pretoria.

Corresponding author

Leanne M Sykes:

Department of Prosthodontics, University of Pretoria.
E-mail: leanne.sykes@up.ac.za

Limitations: The projects are harder to conduct and to write up than conventional research, are subject to personal involvement and bias, and are time consuming as the studies occur in cyclical stages.

OBSERVATIONAL STUDIES

These compare subjects against a control group, the researcher having no control over the experiment. There are two types, *direct* observations where the subjects know they are being watched, and *unobtrusive* methods where they are unaware of being observed. Such studies are used for questions of diagnosis, prognosis and causation (and may then be called *epidemiological* studies).

Uses: Observational studies are good in situations where it is unethical or impractical to carry out large research projects. Their structure is more flexible because there is no intervention, merely observation, and data is emergent over time. They are good for studying interactions amongst group participants. Results may reflect real life situations.

Limitations: Reliability is low and behavioral studies are almost impossible to replicate. The findings reflect only that study population and as such may not be generalized to other populations. They are susceptible to observer bias where the researchers see “what they want to see”, while the group under scrutiny may also behave differently, knowing they are being watched. The outcomes cannot be used to deduce any form of cause and effect relationship. A drawback in the reporting is that “*Statistics do not convey emotion. They merely shock us for a while and then we move on*” (Madeleine M. Kunin).³

CASE STUDIES

These are usually used to describe a rare condition or to explain a novel innovation, and involve an in-depth study of a particular situation. The detailed description may alert others to an important new problem who may try to narrow down a broad field into smaller, easier to test theories. However, along with expert opinions, case studies are considered the lowest levels of evidence.²

Uses: These studies are used to examine real-life situations about which very little may be known, and to determine whether the observations can be applied to a broader population. They help provide detailed descriptions of rare conditions.

Limitations: Due to the single / small number of cases, the studies are uncontrolled, unreliable, and the results cannot be generalized to wider groups. They can also not be used to assess cause and effect relationships. Remember, “*Society and medicine tend to treat us all as members of populations, whereas individuals are all unique, and population statistics do not always apply*” (Craig Venter).³

CASE-CONTROLLED STUDIES

In these studies, people with a condition (the cases) are matched with a group of people who don't have the condition (the controls). The researchers investigate historical data to try to identify whether the cases had been exposed to any common factor that may have led to the condition.

Uses: They are quick and inexpensive to carry out and are useful for rare disorders, or conditions where there is a long delay between exposure and outcome.

Limitations: the main disadvantage is they rely on memory, in which case they may be prone to “recall bias”, or depend on medical records which may be inconsistent, incomplete or inaccurate.²

COHORT DESIGNS

In these studies, it is already known that a group of people have been exposed to some treatment or causative agent (e.g. a vaccine, drug, environmental toxin). These will form the study group, the treatment /exposed group. A separate cohort is drawn from the same population to which the study subjects belong and form the control group, the non- exposed sample. The two groups must be linked by some common feature that is relevant to the problem being investigated. This is easier than trying to look at an entire population. The subjects are then followed forward in time to see how many of each group develop a disease or outcome. The outcomes may be quantitative, in which case statistical occurrences within that subgroup will be analyzed, or qualitative, in which instance data is gathered by observation. Cohorts can be “open” or “closed”. Open cohort studies are dynamic and the study population varies with time, thus the involvement of each subject is relevant only for the duration of time in which they are being studied. Cohort studies are used to calculate rate-based data. Closed cohort studies use static populations. There is a set number of participants, which may remain constant or decrease due to dropout.

Uses: They are less expensive and easier to carry out than randomized control trials (RCTs). In addition they can be used where a RCT would be unethical (e.g. you cannot expose a healthy population to a toxin to study its effects, or withhold a potentially beneficial drug. Thus the test cohort would be a group of people who had previously been exposed to the toxin). Such a study would help to confirm the cause and effect relationship, and make it possible to monitor effects over time.

Limitations: The researcher can never guarantee that cohorts are properly matched or that there are no other confounding variables between the two populations. Not satisfying these conditions could affect the results. The tests can compare the two groups only in terms of the one similar variable, and may take years to complete, as the researcher has to wait to observe whether the condition of interest develops. There is no random selection and thus low external validity. Consider yourself as part of a cohort of friends. “*Statistics show that one out of every four South Africans is suffering from some form of mental illness. Think of your three best friends. If they're okay, then it's you!*” (with apologies to Rita Mae Brown).³

LONGITUDINAL DESIGNS

These are a variation of cohort studies having only one group who have been exposed to some toxin or have been diagnosed with early stages of a disease. They are then followed with repeat observations and evaluation at set time intervals.² This allows the researcher to track changes over time and to see how changing variables may influence the status. They help establish the direction and magnitude of causal relationships.

Uses: They allow an analysis of the duration of a phenomenon, and measure changes in variables over differing time periods.

Limitations: In these studies there is a presumption that trends will remain unchanged, which is often not the case. In addition, the data collection methods and technology may change, both of which necessitate additional qualitative analysis to explain the fluctuations. A large sample is required, assembled under accurate sampling methods and the study takes time to complete. There is the added difficulty of maintaining the same sample over an extended time (sample attrition), and as a result the investigation is usually limited to studying just one variable at a time. For example tobacco usage where *“Smoking has been shown to be a leading cause of statistics”* (Fletcher Knebel).³

CROSS SECTIONAL DESIGNS

These studies try to establish an association between a causal factor and a condition.² They have three key features: there is no time dimension (they happen at one specific point in time), they rely on existing differences (no changes due to an intervention), and groups are selected based on these existing differences (no random allocation). They can measure differences only, can establish an association, but changes are not considered nor can the data be used to make causal inferences.

Uses: They provide an overview of a specific event at a set point in time. There is no intervention on the part of the researcher, and all data is collected at that one time. The sample populations are purposefully selected based on existing differences, and the method can be used to study large populations. Survey techniques are generally used, which are quick and inexpensive to conduct.

Limitations: Identification of the sample population is not easy. Results are time bound and static, and cannot be used to predict sequences or progressions, nor to establish cause and effect relationships. There is no follow up, and thus the possibility exists that a different result could have been achieved if the same populations were to be studied at a different time.

EXPERIMENTAL DESIGNS

These follow a strict blueprint where all factors that may affect the results are under the control of the investigator, and the studies provide the highest levels of evidence. They are used in cause and effect situations, where cause precedes effect, where there is consistency between cause and effect, and where the two are closely correlated. They can be controlled, where there is an experimental and control group, randomization, and an intervention (the independent variable) administered to the former and not to the latter. Both groups are measured and compared on the same dependent variable. Uncontrolled experimental studies have no control or use historical data for the control and as such are very weak because circumstances may have changed, they are prone to bias and there is no guarantee of reliability and standardization of data collection.²

Uses: The researcher controls the situation in order to find out what causes an event, to identify cause and effect, and to distinguish placebo from treatment effects. Observational studies include case studies, cohort designs, case-controlled studies and cross-sectional studies.

Limitations: The intervention is artificial, and as such, the results may not be generalizable to the whole population. In addition, this setting could alter the research subject's

natural behaviour. They can be costly. For ethical or technical reasons, many problems cannot be studied with experiments, where *“Consumers are statistics, patients are people”* (Stanley Marcus).³

DESCRIPTIVE RESEARCH

These studies provide answers to questions of Who, What, When, Where and How, but cannot establish Why! They describe the current status of events, situations and conditions.

Uses: They allow for observation of subjects in a completely natural environment, and are often used prior to conducting a quantitative study to give the investigator a general overview of the situation, and help develop a more focused study.

Limitations: Although a large amount of data may be collected, the results cannot be used to provide definitive answers or disprove a hypothesis. Most studies are based on observations and thus cannot be replicated, and heavy reliance is made on observer related factors. In science one may tend to forget that *“Life is not just a series of calculations and a sum of statistics. It's about experiences, and participation, and is more complex and so much more interesting than what is obvious”* (Daniel Libeskind).³

CAUSAL DESIGNS

These test the effect that a specific intervention or change has on an existing condition. They can be seen in terms of a condition statement: *“If X, then Y, where X is the phenomenon that changes (the independent variable), and Y is the resulting situation (the dependent variable)”*. For any causal relationship to be valid, there needs to be an empirical association between the two variables, they must occur in the appropriate time sequence (cause must come before the effect), and they must not be confounded by some other variable(s).

Uses: These more searching designs help the understanding of how things work, based on linking variables and eliminating other influences. As such, they should be replicable.

Limitations: Not all relationships are causal, and it is difficult to prove that events may not be the result of other confounding variables. Thus, causality can only ever be inferred but never proven. There is a common statistician's warning that *“Correlation is not causation”* (Thomas Sowell).³

EXPLORATORY DESIGNS

These are used to gain insight into situations where there have been very few previous studies, and form a basis for later investigations. They are very flexible and help address issues of all types (What, Why and How?).

Uses: They are good for gaining background information. They allow the researcher to form a clear picture of the details, settings and concerns, to generate new ideas, to formulate a tentative hypothesis, to determine whether a study will be feasible, and may help direct future research.

Limitations: They use small samples, and are exploratory in nature, thus findings cannot be applied in general, and definitive conclusions cannot be drawn. While the approach may be flexible, it is often unstructured and thus

lacks the usual rigorous standards of data collection and interpretation needed to draw conclusions. Mark Twain may have been suspicious of how investigators define the term “flexible” in exploratory studies when he mused “*Facts are stubborn, but statistics are pliable.*”³

HISTORICAL DESIGNS

These entail collecting, verifying and evaluating past evidence in order to support or refute a hypothesis. The major limitation is that they rely on many types of documentary evidence like records, logbook, reports, archives and collections. It is difficult to ensure that the results are authentic, reliable and valid.

Uses: They are unobtrusive and have no actual interventions that may affect their results, or be biased by researcher- participant interactions. They are good for studying trends, providing background information, and can be used repeatedly for different studies or to replicate previous findings.

Limitations: They are totally reliant on the amount and quality of historical data available, which cannot be manipulated in any way to suit current conditions. They can be time consuming. Other major limitations are missing data, inconsistent reporting style or persons, personal biases, lack of control and internal validity, or gaps, which need to be identified and acknowledged. One hopes that these archived records are not akin to the ancient historical documents that Stephen Leacock mentioned when he stated that “*In ancient times they had no statistics so they had to fall back on lies.*”³

PHILOSOPHICAL DESIGNS

Based on a broad approach, the philosophical approach sets out to challenge deeply embedded assumptions. Rational arguments are applied to challenge the relevance, logic and evidence about fundamental issues. The study can take on one of three forms:

- Ontology - describes the nature of reality (What is real and what is not?)
- Epistemology – explores the nature of knowledge and on what it depends. (How can we be certain of what we know?)
- Axiology – studies values and how these relate to interest, desire, will, experience and means-to-an-end. (What is the difference between a matter of fact and a matter of value?)

Uses: They provide a basis for ethical decision making, for understanding the purpose of research, and to help refine concepts and theories. Philosophy informs the methodology and critical thinking, and offers clarity to the practical and theoretical use of terms, concepts and ideas.

Limitations: The analysis is very abstract, answering “*So what?*” types of questions. Writing is often dense, replete with jargon and excessive quotations. It has limited practical use, as it is difficult to move from the philosophical thoughts to application in real-life issues. Philosophically, “*If we knew about the real facts and statistics of mortality, we’d be terrified*” (V.S Ramachandran).³

SEQUENTIAL DESIGNS

These are carried out in a deliberate staggered approach, where one stage is completed and then followed by the

next and so forth, with each stage building on the previous, until enough data is gathered. As such there is no predetermined sample size, as the researcher will analyze data after each sample set, decide whether or not to accept the null hypothesis, the alternate hypothesis or to repeat the study on a new group of subjects. Thus, a limitless number of subjects can be studied before a decision is taken by the researcher. Sequential designs can be quantitative in which case a sampling technique will be used to gather data and statistical methods will be used for analyzing the results. If a qualitative framework is used, then methods such as interviews and observations are used for data collection.

Uses: There is a limitless sample size and sampling schedule. Due to the sequential nature, there is scope to make minor changes in the study design with each new study population based on findings from the previous results. As such, it is useful for exploratory research, as it requires little effort, expense, time and workforce.

Limitations: The sampling is not randomized, each sample is usually small, and samples are not representative of the entire population, meaning that the findings cannot be generalized. It is also difficult to account for and interpret variations between sample groups over time.¹

META-ANALYSES

These advanced investigations systematically evaluate and summarize results from a number of previous studies. This serves to increase the overall sample size, allowing the researcher to develop a new understanding of a problem by applying critical reasoning to all of the combined results. They are good for analyzing differences in results between studies, which increases the precision of estimating effects. They must adhere to strict criteria of study selection, and depend on the accuracy of the results and analysis of each study. They become difficult to interpret when there are major differences in findings between studies. In order for a meta-analysis to be considered valid, the researcher must:

- Clearly define the objectives
- Formulate precise definitions of the variables and outcomes being evaluated
- Have good justification for identification and selection of the included and excluded studies
- Be able to assess and acknowledge research bias
- Evaluate the degree of heterogeneity among the sample sizes in each study, and
- Justify the techniques used to evaluate the studies.

Uses: Meta-analyses help identify gaps in the literature, allow for review of one topic over an extended time period and from a variety of sources, and help clarify which policies can be scientifically justified. They also overcome the problem of small sample sizes and highlight research problems for future studies.

Limitations: They are very time consuming, and data may be meaningless if the criteria used for analysis are not clear and strictly adhered to. The lack of uniformity within studies can make it difficult to synthesize the results. Ben Bernanke was correct in warning that “*Aggregate statistics can sometimes mask important information.*”³

RANDOMIZED CONTROL TRIALS (RCTS).

These represent one of the highest levels of evidence (Table 1). In RCTs, participants are randomly allocated into one of two groups. The first group (the experimental group) receives the experimental treatment, while the other (the control) receives conventional treatment, a placebo or nothing.

Uses: They are used in many types of research, and are the best for questions related to therapy, such as testing of new drugs, devices or surgical procedures. These studies may be further divided into:

- a) Single blind, in which participants do not know in which group they are allocated so as not to influence their behaviour. However, the researcher DOES know who is in each group, which could jeopardize the study as the researcher can subconsciously influence the participants or the results.
- b) Double blind, in which neither the researcher nor the participants know the assignments to the groups. A third party will be privy to this information and can make it available at the completion of the study for analysis of the results. It is most useful when the control group receives an identical placebo drug. However this cannot be used in a number of studies for ethical and acceptability reasons, such as a patient receiving "sham surgery".² These would then be considered "open" trials, as the investigator and patient know the intervention. One way to overcome this problem is to have three other persons blinded. Firstly someone other than the original investigator should evaluate the outcomes, secondly the statistician doing the analysis should not know details, and finally the investigators who write up the results should also be independent (this seldom actually happens).² RCTs are so highly valued because the randomization keeps both groups as similar as possible, which along with blinding, sample size justification, and appropriate outcome measures and statistical analysis, help minimize bias.²
- c) Two special types of RCTs used often in Dentistry are the cross-over studies and split-mouth designs. They require smaller sample sizes to detect an effect, but they need to adhere to stringent criteria and may be associated with ethical and technical issues.

Limitations: They may be difficult to carry out due to ethical, legal or technical factors.

INTEGRATIVE STUDIES

It is not advisable to base major clinical decisions on results of a single trial – no matter how large or well executed. Thus integrative studies take all relevant information from previous research that address the same issue and collate this into a systematic review. When it is possible to analyze the combined results statistically, it is called a quantitative systematic review or a meta-analysis.²

Uses: Although these studies are retrospective observational research, they adhere to stringent inclusion and exclusion criteria and use scientific methods to control for bias. As such they are considered to provide the highest levels of evidence in the hierarchy.

Limitations: The researcher is required to carry out a thorough in-depth literature search, and to adhere strictly to the pre-established inclusion and exclusion criteria.

Table 1: The Evidence Ladder.²

The Evidence Ladder (From highest to lowest)
High-quality systematic reviews
Large randomized trials with clear-cut results
Small randomized trials with uncertain results (positive trends but no statistical significance)
Non-randomized trials with contemporary controls
Non-randomized trial with historical controls
Cohort studies
Case-control studies
Dramatic results from uncontrolled studies
Case series / descriptive studies
Reports of expert committees and opinions of respected authorities, based on clinical experience

Differing study designs, samples and analysis methods may make it impossible to compare results, despite the topics being the same.

GRADING EVIDENCE

Judging evidence in Dentistry is difficult due to the clinical nature of the discipline and variations in operator technique and skills. Guyatt (1992) proposed an evidence-based approach to tackle these challenges, stipulating that there should be formal rules for evaluating the trustworthiness of evidence.⁴ This led to the development of "The Evidence Ladder / Pyramid" (Table 1) which ranks evidence from the highest (top), to the lowest (bottom).^{4,2} Case reports and expert opinions are considered the lowest, while meta analyses and RCTs are rated topmost in terms of reliability and biological plausibility.⁵ (Note that a low level does not imply poor quality or low value, but is used as a basis for making clinical decisions for humans.) Low level study designs often lead on to the formulation of more in-depth hypotheses and studies.⁵

CONCLUSIONS

In conclusion, while one needs to acknowledge that research, science and advancement have beneficial potential, we must always remain cognizant that patients are people, and NOT as Horace put it "We are all just statistics, born to consume resources."³ In addition, all the evidence in the world means nothing unless findings are published and recommendations are implemented. After all, "Quoting statistics won't stop the globe from warming if the globe is actually, you know, warming!" (Clive Thompson).³

References

1. De Vaus, D. Research design in social research. in <http://libguides.usc.edu/writingguide>. 2001, SAGE: University of Southern California.
2. Sutherland, S. Evidence-based dentistry: Part IV. Research design and levels of evidence. J Can Dent Assoc, 2001. 67: 375-8.
3. Various. Statistics Quotes- Brainy quotes. 2016 [cited Accessed: June 2016].
4. Guyatt, G. Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA, 1992;268(17): 2420-5.
5. Hujuel, P. Statistical and Methodological Aspects of Oral Health Research. ed. J.F. Emmanuel Lessafre, Brain Leroux. 2009, West Sussex, UK: John Wiley & Sons.