Correlational Research: Importance and Use in Nursing and Health Research

Authors: Elizabeth A. Curtis¹, Catherine Comiskey² and Orla Dempsey³

Abstract

Purpose: Discuss key issues and considerations about correlational research and put forward suggestions for averting potential problems during preparation and application of the design.

Background and Discussion: The importance of correlational research has been reported in the literature yet many research texts including nursing books offer little discussion on the design; with several providing approximately one to three pages in chapters entitled, for example, quantitative designs. This article targets this gap by focusing key considerations for planning a correlational study. Specifically, it discusses the importance and purpose of correlational research, its application, analysis and interpretation with contextualisations to nursing and health research.

Review Methods and Data Sources: Materials were sourced primarily from books on research and statistics as few relevant research papers were located. Several databases were used and the date range for searches was 1980-2015.

Conclusion: Findings from correlational research can be used to determine prevalence, relationships among variables and to forecast events from current data and knowledge. In spite of its many uses however, prudence is required when using the design and analysing data. To assist researchers in reducing mistakes key issues were singled out for discussion and several options put forward for analysing data.

Implications for Healthcare Practice: Correlational research is used widely by nurses and other health professionals therefore this article should be particularly useful for novice researchers from these disciplines. Furthermore, findings generated from such research have utility for healthcare practice: they can be used for example, to inform decision-making in nursing/healthcare and improve or initiate health related activities/changes.

Key words: nurses, nursing research, correlational research, variables, sampling, measurement tools, analysis techniques, educational needs, ethical issues, evidence-based practice

¹Dr. Elizabeth A Curtis, Assistant Professor, School of Nursing and Midwifery, Trinity College Dublin

²Dr. Catherine Comiskey, Professor of Healthcare Statistics, Head of School, School of Nursing and Midwifery, Trinity College Dublin

³Dr. Orla Dempsey, Primary Market Research Specialist, Novartis Ireland Ltd.
INTRODUCTION

Correlational research is concerned with establishing relationships between two or more variables in the same population or between the same variables in two populations (Leedy & Ormrod 2010). In nursing (Prematunga 2012) and indeed other disciplines (Fitzgerald et al. 2004) exploring the relationships among various variables is a significant part of healthcare research. Understanding the associations/relationships that exist among “human phenomena is an abiding impetus for scientific enquiry in all of the social science disciplines, and that impetus transcends even the most polarised paradigmatic distinctions between various research methods ……” (Fitzgerald et al. 2004:143).

The origins of correlation and linear regression techniques can be traced back to the work of Sir Francis Galton (Miller & Millar 1996). Galton, while known for developing correlation the majority of his work was in the area of inheritance which resulted in the development of regression, from which correlation was a logical conclusion. The mathematical development of correlation was however attributed to Karl Pearson. Some authors have tried to establish whether correlation and regression did indeed exist before the work of Galton but as Denis (2001) reported drawing a conclusion on this is not straightforward. It would seem that the mathematical tools used in regression and correlation had been developed before Galton, however, Galton used the discovery of regression to promote his theory of inheritance. This, according to Denis (Denis 2001) is a critical point: Galton had a specific reason for discovering correlation while others before him did not.

The purpose of this article is to discuss key issues and considerations about correlational research and put forward suggestions for averting potential problems during preparation and application of the design. It does not provide a comprehensive resource on the topic. Rather, it discusses issues that we consider to be important and provides examples from actual research studies where appropriate. The article concludes by exploring the utility of correlational research for evidence-based practice and defends its continued use within healthcare. Readers across disciplines as well as internationally should find the article relevant given that the principles presented are standard within the discipline of research. Furthermore, once correlation has been established through research, findings can be used by researchers globally to make predictions and construct hypotheses for future research.

Background – Importance and Purpose of Correlational Research

The importance of correlation research has been emphasised by authors such as Woodworth who published a book entitled Experimental Psychology in 1938 (Woodworth 1938) and Lee Cronbach who published an article entitled The Two Disciplines of Scientific Psychology (Cronbach 1957). Woodworth established two major distinctions in quantitative research methods; the distinction between (a) independent and dependent variables and (b) experimental and correlational methods. According to Woodworth, experimental designs/methods are characterised by manipulation of variables while a correlational design/method measures two or more characteristics from the same person and then calculates the correlation between the characteristics. In Woodworth’s view these two research approaches had equal value – correlational research must “be distinguished from the experimental method, but standing on a par with it in value, rather than above or below…..” (Woodworth 1938: 3). Cronbach was also concerned about the fact that correlational research had second-rate status in scientific
psychology. He believed that a synthesis should take place with advocates of each design valuing them equally, and utilising both strategies. As Cronbach puts it “It is not enough for each discipline to borrow from the other. Correlational psychology studies only variance among organisms; experimental psychology studies only variance among treatments. A united discipline will study both of these, but it will also be concerned with the otherwise neglected interactions between organismic and treatment variables.” (Cronbach 1957: 681). Whether this synthesis happened remains debatable but is issue/topic is certainly worthy of further discussion elsewhere. Interestingly many text books including nursing books provide little discussion on correlational research, with several providing one to three pages in chapters entitled for example non-experimental designs or quantitative research designs. Books devoted entirely to correlation designs are available for those who wish to engage in further reading (Bobko 2001, Miles & Shevlin 2001, Vogt & Johnson 2012).

Correlational research can be described in different ways. Cohen (1968) for example, suggests that because parametric analyses (e.g. t-tests, ANOVA, ANCOVA) explore relationships among variables then quantitative studies would therefore produce correlational evidence. But, as Cohen (1968) further explains describing research by the analysis techniques is not useful because under such a broad classification all research findings would fall within a single category. The purpose of correlational research is to investigate “the extent to which differences in one characteristic or variable are related to differences in one or more other characteristics or variables.” (Leedy & Ormrod 2010:183). A correlation occurs if one variable (X) increases and another variable (Y) increases or decreases. A study that produces a correlation coefficient of 0.00 signifies that there is no association between the variables investigated.

In preparing to write this article, it became apparent that there is a paucity of papers on the topic. Therefore, this article should make a contribution to the existing literature since it brings together key issues that researchers planning to use a correlational design need to consider.

Data Sources

Databases searched from 1980 to 2015 include Embase, CINAHL, PubMed, PsycINFO, Science Direct and Web of Science. Key search terms included correlational research or correlation research and sub themes were nursing or health. Some of the searches produced in excess of 100 papers but many of these were not relevant due to the inability of some databases to do phrase searching. Others resulted in fewer papers (e.g. PubMed = 18, PsycInfo = 33).

DISCUSSION

This focuses on three main themes: using correlational research, analysing and interpreting data from correlational research and implications for nursing and healthcare provision.

Using Correlational Research

Many nursing and healthcare research studies arise initially from the need to quantify the numbers of clients attending a particular service within a defined time period or to measure the absence or presence of a particular health characteristic within a defined population or client group. In such cases, the primary research objective is to ascertain how many clients are using or in need of a service. The study design that can readily address this objective is a cross sectional one. However while the primary objective may be to estimate prevalence, in order to
provide an appropriate and targeted service, planners need to know if there are any additional characteristics within the client group that may be related to prevalence. When the importance of additional characteristics is suspected then the secondary objective of a research study is to measure the strength and direction of any possible relationships between prevalence and the additional characteristic. Correlational research is used to address this objective.

**Key Considerations**

Correlational research is used frequently in social and healthcare research because it can be used in any research study that does not wish (or is unable) to manipulate the independent variable(s) under investigation. But, before a research study can commence several factors must be decided, including the selection of suitable variables, identification and selection of the sample and the use of reliable tools.

**Variables**

A variable is used to describe something that varies. Weight, height, blood pressure, happiness, and innovative organisational culture are examples because each can vary among individuals and organisations (Polit & Beck 2012). Two variables may coexist but it is only through research that a relationship can be demonstrated between them or the direction and strength of that relationship be established. Quantitative researchers try to establish how and why things vary and to determine how differences in one variable are associated with differences in another. All correlational studies require a conceptual framework or a description of why the variables might be related to one another and it is important that the researcher determines accurately the variables that exist in the area of research that is of interest to them (Maltby et al. 2010). All variables relevant to the research questions formulated for a research study must be clearly defined, and to determine correlation between variables the level of measurement of all variables must be ordinal, interval or ratio.

**Sampling**

Quantitative research designs including correlational studies usually use large samples that have been attained by a precise process. This is important because the purpose of sampling in quantitative studies is to produce statistically representative data that permits generalisation of findings to the target population (Norwood 2010). Therefore, before selecting a sampling method it is necessary to review the original aim/intention of the study.

**Measurement Tools**

Nursing research requires good quality data, which are reliable. Reliable data are objective, accurate, valid (measures what it is supposed to), free from error and usable (Norwood 2010). Good quality data result from good quality research, which is partially due to meticulous decisions about data collection and measurement (Polit & Beck 2012). Research tools can be used to either measure variables directly or indirectly. Direct measures are used when the variable under investigation is tangible or accessible (e.g. blood glucose levels, level of consciousness, temperature). Questions that obtain information about personal characteristics such as age, marital status, level of education are also considered direct measures. Indirect measures generate information “about indicators of an attribute …of a phenomenon of interest
rather than about the phenomenon itself.” (Norwood 2010: 251). That is to say, the variable being measured is presumed to reflect the variable of interest and permit inferences to be drawn about the phenomenon. In view of this discussion, it is important that appropriate measurement tools are used in research studies (Lehane & Savage 2013). Furthermore, a good description of all measurement tools should be given and the reliability and validity for the tools reported (Watson 2013).

**Analysing and Interpreting Data from Correlational Research**

Before discussing techniques for analysing data from correlational studies it is important to distinguish between correlational statistics and correlational research. If a Pearson’s $r$ is found in a research paper it is easy to assume that a correlational design was used. However, correlation as a statistical tool and a research design are different. While Pearson’s $r$ is a statistical test frequently used in correlational studies to calculate the relationship between variables, it can also be employed in an experimental study to determine the relationship between independent and dependent variables. Similarly, the use of a t-test or an ANOVA in a research paper does not indicate that an experimental design was used in the study (Goodwin 2002). For example, Cox & Comiskey’s (2007) study of the baseline characteristics of clients entering opiate substitution treatment for heroin use studied three treatment groups. For ethical reasons, an experimental design could not be utilised and clients were not randomly assigned to a treatment group.

**Analysis Techniques**

Before conducting a statistical test to determine if a relationship exists between two continuous variables (Comiskey & Dempsey 2013) it is useful to plot the data on a scatterplot to give an indication if the variables are related; and if so, the direction of the relationship (Pallant, 2010). If a linear relationship is indicated on the scatter plot a statistical test can be run to determine the strength of the relationship.

For two continuous normally distributed variables, Pearson’s $r$. is the statistical test to determine if a relationship exists between the two variables and the strength of that relationship. While Spearman’s $\rho$ is the non-parametric alternative to Pearson’s $r$ for non-normal continuous random variables (Comiskey & Dempsey 2013), it is also appropriate for assessing the relationship between two ordinal variables (e.g. two variables measured on Likert scales). The correlation coefficient is calculated based on ranking participants’ responses with the highest score on the Likert scale assigned the highest ranking. Pearson’s $r$ and Spearman $\rho$ correlation coefficients are interpreted as explained above. While correlations cannot be computed for purely categorical data (Comiskey & Dempsey, 2013), statistical tests can be conducted to determine if associations exist between categorical variables.

Initially in correlational research it is useful to summarise the data obtained using frequencies and descriptive statistics. Frequencies are counts of data and are particularly useful for examining the number and percentage of participants in each category. O’Brien et al. (2014a) used this to ascertain the number of patients with different tumour types when examining the cost of febrile neutropenic events in oncology. Descriptive statistics, such as mean and median are used to describe sample characteristics in correlational research. The standard deviation is useful to describe how much the data varies from the mean. In a national survey, Curtis & Glacken (2012) used mean and standard deviation when analysing job satisfaction among public
Confidence intervals (Crispino 2013) can be particularly useful in prevalence studies to provide a range of values within the true population mean or proportion lies (Watson & Coombes 2009).

Statistical tests to determine relationships between dependent and independent variables are decided based on the type of data collected i.e. categorical data; nominal or ordinal or continuous data. For correlation analysis using two nominal variables, chi-square tests for independence can determine if differences/associations exist between independent and dependent variables. This test is appropriate for large sample sizes and when no more than one of the expected cell counts in a table is less than five. The chi-square test for independence was used by Ballard et al. (2013) to examine the prevalence of frailty related risk factors in older adults, specifically associations between independent variables; age group, living status, referral source and dependent variables; frailty-related risk factors. Fisher’s exact test is a suitable alternative if the assumptions on sample size and expected cell count for the chi-square test are violated. Begley et al. (2010) used Fisher’s exact tests when evaluating clinical nurse specialist (CNS), clinical midwifery specialist (CMS), and advanced nurse practitioner (ANP) roles in Ireland.

In correlational research when a researcher wishes to compare two groups on some continuous variable, e.g. the health outcomes of opiate users in two different treatment modalities, an independent samples t-test should be used. The independent samples t-test is suitable for small samples which follow a normal distribution. The Mann Whitney U-test, the non-parametric alternative, compares the median of two independent groups; e.g. comparing Gait speed among younger and older adults (Benzinger et al. 2014). It is suitable for comparing two independent groups if the sample size is very small or the assumptions of the independent samples t-test are violated (Comiskey & Dempsey, 2013).

In correlational studies comparisons between three or more independent groups on a continuous variable may be required. Begley et al. (2013) used a one-way ANOVA to compare average time waiting for (CNS; CMS; ANP) services. Kruskal-Wallis is the non-parametric alternative to one-way ANOVA.

Multiple correlational analysis measures how well a dependent variable can be predicted using a linear function of a set of independent variables. The multiple correlation coefficient $R$, measures the strength of the relationship between the independent and dependent variables. Similar to $r$ and $\rho$, the coefficient $R$, takes a value between zero and one; however, unlike $r$ and $\rho$, it only indicates the strength not the direction of the association.

Regression can be used to model the association of independent variables with dependent variables. Multiple regression is used to model the association between two or more independent variables and one continuous dependent variable. In multiple regression the independent variables can be dichotomous, e.g. smoker (yes/no), or continuous. Multiple regression analysis was used by McKee et al. (2013) to examine predictors of pre-hospital delay time in patients with acute coronary syndrome. Logistic regression can be used to model the association between two or more independent variables and one dichotomous dependent variable (e.g. male/female).

There are a number of common mistakes in the analysis and interpretation of findings from correlational analyses. One of the most common mistakes is to assume that correlation
Correlation examines if a linear relationship exists between an independent and dependent variable. However, causality is the relationship between one event being the cause of another event. While correlation may indicate a relationship between two variables it does not account for other underlying variables which may be causing the effect. Another common error is to make inferences about a group. Correlational analysis may indicate there is a relationship between two variables; however, it is not always possible to make inferences about the general population based on a correlation. Frequently researchers assume a correlation of zero implies independence however it just indicates no linear relationship exists between the two variables. While correlation implies that an independent variable and a dependent variable may be related, it does not imply that a change in one variable leads to a change in the other variable (Plichta and Kelvin, 2013).

Implications for Healthcare Practice

Given that this is a discussion rather than a research paper the implications for nursing and healthcare practice are discussed in relation to two key themes:

(a) the contribution findings from correlational research can make to evidence-based practice

(b) whether nurses and other health researchers should continue to use correlational research to explore clinical, educational and theoretical problems.

Correlational Research and Evidence-based Practice

True experimental designs potentially provide the highest evidence for causal relationships between variables (Talbot 1995). Experiments test hypotheses to establish causality and are in a position to achieve this because they require manipulation, control, and randomisation. It is for this reason that researchers are often encouraged to conduct true experiments. For reasons such as (a) some research questions are more suitable for qualitative designs (b) ethical approval may not be granted because of denial of the treatment for a control group, it may not be possible to carry out experimental research. This however, does not mean that findings from research studies using other designs including correlational research are not relevant to evidence-based practice (Thompson et al. 2005, Sousa et al. 2007). Correlational research as pointed out by Thompson et al. (2005: 182) “…. can be used to inform causal inferences and thus evidence-based practice.” The first method involves testing rival causal models statistically even when using a correlational design and involves the use of structural equation modelling (SEM) or covariance structure analysis. SEM can be used to analyse causal models involving latent variables and is gaining popularity among nurse researchers, but, it is a complex procedure. The second method is based on logic and uses logic and theory to exclude other possible explanations “in support of making a …. plausible causal inference” (Thomson et al. 2005: 182).

Continued use of Correlational Research

The answer to this question is based on both practical and ethical issues. With regard to practical issues there are instances when it would not be realistic or possible to use an experimental design. For example, if researchers wanted to study differences between age groups, personality types or ethnic behaviours it might not be possible to randomly assign to groups. Furthermore, a study that specifically sets out to explore the correlation between job satisfaction and
absenteeism has its own value even if subjects cannot be assigned to high job satisfaction and low job satisfaction conditions (Goodwin 2002). Further support for correlative research is reported by Wong et al. (2013) who carried out a systematic review to examine the relationship between nursing leadership and patient outcomes. Findings suggest relationships between leadership and patient outcomes such as higher patient satisfaction, lower patient mortality, lower medication errors and lower hospital-acquired infections. Such findings are particularly important for nursing and can be used together with findings from similar studies to augment nursing care practices/procedures. Similarly Hanrahan et al. (2010) used correlational analysis techniques to determine the relationship between work environments and burnout among psychiatric nurses. The findings demonstrated lower levels of burnout was significantly associated with better work environments, more effective managers, strong nurse-physician relationships and higher nurse patient staffing ratios. Once again these findings can be used to support changes in the work environment.

Further support for correlative research is attributed to its major advantages. The design is straightforward to use, can be relatively inexpensive unless using large surveys, can be completed fairly quickly and data provides a useful starting point for researchers exploring a phenomenon for the first time. Researchers can establish the direction and strength of a relationship between variables thus allowing further research to be carried out. In other words, before causality can be explored an empirical relationship between relevant variables must be established (Polit & Beck 2012). Furthermore, the results from correlative research can be used to generate hypotheses to be tested in quasi-experimental and experimental designs. The fact that correlative research does not determine causation is regarded as a disadvantage (Talbot 1995; Sousa et al. 2007). However, acknowledging this disadvantage does not in any way detract from the significance and contribution correlative studies make to the field of research. So, based on this discussion we wish to reaffirm support and encouragement for the use of correlative research by nurses and other healthcare professionals.

CONCLUSION

This article provides a succinct summary not only of the historical background to correlative research but also a summary of its importance and when and how it could be used within nursing and health research.

The first question asked within many research topics is, what is the scale or the size of the problem? This question arises from the need to plan and design services that are appropriate to the often changing needs within healthcare. To address these needs in a targeted way, researchers not only need to know the size or scale of the problem but also which client groups are in greatest need. Correlational research allows us to ascertain what client factors, be they demographic, socio-economic or health and wellbeing are correlated most strongly with the primary outcome of measure. Through the use of correlative research these questions can be addressed and services developed appropriately for the age group or gender or other client group most strongly correlated with the primary outcome or health attribute of interest. Furthermore, the results of these preliminary correlative research studies can provide invaluable information on the direction of future research that may be required on the specific variables that have been shown to be correlated with the outcome or attribute of the original study.
Acknowledgements:

We wish to thank Ms. Jessica Eustace-Cook, Subject Librarian for Nursing and Midwifery, SFI Research Support Librarian, Trinity College Dublin for her assistance in conducting a literature search on the topic. We appreciate her expert knowledge and are delighted she was able to assist us in this important task.
References


