Deakin Research Online

This is the published version:

Gardner, Anne, Millar, Lynne and Driscoll, Andrea 2007, Dealing with data: optimising data consistency to demonstrate the value of nurse practitioner service*, Nurse practitioner series*, vol. 2, no. 1, pp. 19-25.

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30025544

Dealing with data: optimising data consistency to demonstrate the value of nurse practitioner service

Authors: Anne Gardner, Professor of Nursing, James Cook University, Townsville
PhD RN Critical Care Certificate BA MPH MRCNA

Ms Lynne Millar, Research Fellow, Deakin University, Burwood
BA (Hons) PhD Candidate

Ms Andrea Driscoll, Research Fellow, Deakin University, Burwood
PhD (Submitted) RN Critical Care Certificate MNursing Medu

Key words:

Nurse practitioner, data management, research, survey construction, sample.

Introduction

Other authors in this volume have argued for the need to undertake audit and research related to nurse practitioner practice. Audit and research require skills in competently reviewing literature on a topic, developing an answerable question and devising an appropriate design and methodology. However, these are not the only aspects that need consideration before commencing the data collection phase. Too often clinicians successfully review the literature, identify the question, design and methods and then launch into the 'doing' of the project without also planning how data will be managed and analysed. The organising and processing of data are just as important as any other phase of the project, but these components of a successful project are often overlooked in the planning stages.

Audit and research related to clinical outcomes often require regular and repeated collection of information over a long period. A researcher may want to compare patient outcomes after the establishment of a nurse practitioner service with patient outcomes before that service was made available. They may need to demonstrate that additional investigations or medications are required to manage a patient group more successfully. The researcher will certainly have to provide some demonstration of the efficacy and clinical effectiveness of the role on a regular basis. It is important that findings are comparable from year to year and that they can be compared rigorously with services provided by other health professionals or nurse practitioners in other regions. How can a researcher ensure that the information collected will be comparable with other information sets? This article will discuss data management strategies to maximise the reliability and comparability of project findings over time and across services. A list of useful websites is also included (Appendix 1). This information is directed primarily at Australian nurse practitioners and aims to provide guidance about managing data and optimising data consistency.

Ethical issues and data protection

Clinicians will be familiar with handling sensitive information but there are specific rules that must be understood if that clinical information is to be used for purposes other than direct client management. All health services are required, as part of good clinical governance, to undertake quality improvement activities. These activities often include clinical audit. It is not always easy to differentiate between clinical audit for quality improvement purposes and the use of clinical records for clinical research. The National Health and Medical Research Council (NHMRC) has developed guidelines to assist clinicians and researchers to distinguish between research and quality improvement (see Appendix 1). In particular, a clinician may

intend to publish from audit findings (for example, pressure ulcer point prevalence findings). Many journals now require prior review of the project by a human research ethics committee (HREC). Many committees use the NHMRC guidelines as a checklist and will grant exemption or consider the project proposal as a research project if there are significant research ethics issues. Furthermore, HRECs are also bound by the federal national privacy principles, and the state and territory privacy regulations. These principles and regulations can be sourced from the relevant government websites (see Appendix 1). Researchers are strongly advised to contact their HREC if they hope to present or publish their project findings.

Planning for data collection

One of the most important rules of data collection is to collect only as much information as is needed to answer the question. Researchers should ask themselves exactly what they want to know. The following questions may help:

- Who does the study want information about?

- Where is the client population situated?

- What timeframe is relevant to the study?

The first question relates to the importance of having a clearly defined aim, objective or question to guide the project. An emergency department nurse practitioner who provides services to manage patients with a wide range of conditions may decide to concentrate on the evaluation of the three most frequently occurring patient groups rather than cover all groups of patients that make up his or her client profile. Collecting data on a homogeneous group will more easily provide a critical mass of information (an adequate sample size) that can be analysed statistically.

Regarding the second question, a nurse practitioner may provide both clinic and outreach services. He or she may decide that information about the clinic services are already collected at some level and so data collection can be limited to outreach services.

The timeframe is also important. For example, is it sufficient to collect information for two months each year and should it be the same two months each year? Alternatively, should the information be collected continuously? There will be many other questions that are relevant for planning of a particular project so this section is not exhaustive. However, these reflections will enable the researcher to decide the extent and the limits of data collection. It is a waste of valuable time and, more importantly, it is often unethical to collect information that is not required.

All of these questions will inform the calculation of sample size. Sample size is critical as it is the major influence on statistical power (Shaughnessy, Zechmeister, & Zechmeister, 2003). Greater statistical power will enable the researcher to correctly conclude whether there has been any treatment effect or not. A sure way to increase statistical power is to increase sample size. With a large sample, findings will be more robust. A good rule of thumb for calculating required sample size is 50 + k8 where k equals the number of variables to be measured. Therefore, if there are three variables of interest, the sample size should be at least 50 + (3 x 8); 74 participants. Remember, however, that statistical significance does not necessarily equate to clinical significance.

To explore the possibility of causal relationships between the therapy provided and subsequent patient outcomes, information must be collected about alternative explanations, that is, other factors that might influence the outcome. A sample of sufficient size is also necessary, as is a composition that reflects the population to which the findings will be generalised.

Data analysis

The questions to be answered will guide the type of data analysis undertaken. Data analysis should be planned early, as the type of analysis used has an impact on the number of participants required. The more complex the statistical analysis, the more power required and, hence, the larger the sample size required. For example, for simple descriptive statistics only, the sample size required would be quite small. However, to undertake statistical modelling, a much larger sample size would be required.

At this stage, it is useful to try to imagine what results are to be presented. Describe the tables to be included? Analysis of an audit of patient outcomes after treatment with a particular therapy might require a table presenting demographic characteristics (for example age, postcode). These two variables would seem to be very straightforward, but are they? Age can be collected as exact age in years at time of consultation, date of birth or age group. Date of birth is the most useful of the three options because the other two can be calculated from date of birth. A researcher cannot analyse below the data level observed. This is known as the atomicity principle. If a question asks school age participants to nominate which age group best reflects their chronological age, this age variable is measured at the ordinal level (for example, 5-9 years, 10-14 years, 15-19 years) and data cannot be further divided (for example, into 5-6 years, 7-9 years) or the information used as a continuous variable.

The second example of a seemingly straightforward variable is postcode. In a tertiary referral hospital the postcode recorded on patient admission may refer to the patient's current place of residence (local hostel or relative's house) or the person's primary place of residence which may be several hundred kilometres away but which is much more reflective of the real catchment area for the hospital.

Design of data collection tools

The next step in data management is to design the data collection tools. A researcher may use tools that have already been developed and rigorously validated in well designed studies, or they may decide to design their own tool because there is no relevant tool available. Validity and reliability are important principles in date collection. Table 1 provides definitions for validity and reliability related to questionnaires. Many good textbooks provide a guide for the process of designing the content of data collection tools.

Table 1: Validity and reliability of questionnaires

Term	Description
Face validity	On reading the questionnaire, ensure that the survey items appear to answer the research question
Construct validity	There should be strong relationships between the items in the questionnaire and the intent of the questionnaire
Convergent validity	The questionnaire should share strong relationships with other questionnaires developed to measure similar constructs
Divergent validity	There should not be relationships with questionnaires designed to measure different constructs
Reliability	The questionnaire should repeatedly measure the same constructs

This paper is more concerned with the practical aspects of tool design. Practical aspects include:

- Layout of the questions on the page of a paper-based survey.

- Options provided to the participant if a question is not applicable.

- Numbering sequence of a chart abstraction tool.

These aspects of tool design assist the researcher to record information reliably, reduce the potential for missing data and assist in transcription from a paper record to a computer database or from one database to another.

One of the most important stages of a project is the pilot study (Gardner, Gardner, MacLellan & Osborne, 2003). At the very least, the research team should pilot the use of the questionnaire or chart abstraction tool and refine the tool layout on the basis of pilot study findings. For example, chart abstraction tools should be piloted on each of the documents that are to be audited. An audit of reported medication errors across six Victorian hospitals was recently undertaken (Gardner et al., 2007). In some hospitals the incident reporting system was paper based and in other hospitals the system was electronic. The data abstraction tool was piloted in three of the six hospitals before starting the study, which enabled the streamlining of data collection considerably. With more than 1200 reports to audit, this streamlining was very important.

Consistency with data collected nationally and internationally

Nurse practitioner services are new to the Australian health system and have the potential to greatly improve access and timely health care for many underserviced Australian groups (Gardner & Gardner, 2005). Health and demographic information is recorded at the local, national and international level and some nurse practitioner services will be included in this health service information, although a specific service may not currently be attributed to a nurse practitioner. Systematic analysis of nurse practitioner services has not been undertaken on a national level to date. A federally-funded project is currently underway and will address many aspects of this deficiency (Reforming healthcare: nurse practitioner and workforce redesign: AUSPRAC). A proposed national minimum dataset to inform workforce planning (N3ET) (Australian Health Ministers' Advisory Council, 2006) will guide aspects of the data collection for AUSPRAC (Appendix 1). The components of the proposed national minimum dataset can also be used to assist nurse practitioners in designing audit and research projects at the local level. Nurse practitioners can include data already collected as part of health service reporting and can use data item specifications for other relevant information that is not currently collected about their services or specifically attributed to them. For example, many clinical interventions currently carried out by nurse practitioners can already be coded using existing categorisations such as ICPC-2 PLUS codes. More information about this example can be found on the Australian Institute of Health and Welfare (AIHW) website (Appendix 1). This categorisation can inform the way that nurse practitioners collect data and indeed may mean that existing data collection processes can be used, thus reducing the workload associated with audit and research. The other important advantage of using standardised items is that information is collected in a format that will allow comparison with other services. Appendix 1 provides a list of websites and other resources that will help novice researchers to understand and use nationally and internationally accepted data elements for data collection in audit and research. Standardised quantification of clinical practice is a key component in rigorous audit and research if nurse practitioners are to demonstrate their clinical effectiveness. This quantification will not be possible in all aspects of service delivery but should be a key aspiration of clinicians and researchers who are committed to data consistency, rigorous evaluation of health services and to optimising the quality of clinical practice.

Data management

How are data managed after collection? Again this must be planned before the collection commences in order to control the structure and flow of the data (Munro, 2005). Senior researchers or project leaders, while often not directly entering data into the computer program, should be involved in developing and monitoring the procedures. This section explains some of these responsibilities. Table 2 provides definitions of some of the terms that we will be using.

Table 2: Useful definitions

	Definition
Variable	Any quantity that varies. Any attribute, phenomenon or event that can have different values*
Value	Magnitude of measurements*
Coding	Translation of information e.g. questionnaire responses into numbered categories for entry into a data processing system*
Descriptive statistics	Describe or characterise data by summarising them into more understandable terms without losing or distorting much of the information**
Inferential statistics	Set of statistical techniques that provide prediction about population characteristics based on information in a sample from that population**

* Last (2001).
** Munro (2005).

An important step is to plan for entry of the data into an electronic database. Time should be taken to plan the structure of the data processing system, as it will be necessary to both store the data and to analyse them. The first decision that needs to be made is what data processing system will be used. There are several levels of complexity of software and different products may be used to store and to analyse data. Excel™ is a readily available product that provides spreadsheets, can perform many numerical operations on the data and is one of the best packages for production of charts and graphs. Access™ is a much more sophisticated product, known as a relational database. This means many spreadsheets can be set up and connected in different ways. The relationships may be 'one-to-one', 'one-to-many' or 'many-to-many'. For example, 'one-to-many' means that demographic information can be entered on each patient once on a spreadsheet dedicated to data at the patient level and then these demographic fields can be linked to several spreadsheets dedicated to data at the level of separate consultations. This linkage has several advantages. It will reduce the need to enter the same information several times and it will maximise the reliability of the data because the same piece of information, used in several worksheets, is only entered once. Another advantage to Access(tm) and other relational databases is that they usually have an option to design an electronic form. This can be printed and used for the data collection and can also be used as the screen for entering data. If the online screen has the same appearance as the paper collection tool then the quality and consistency of data entry is maximised. The main disadvantage to Access(tm) is that the database designer and operator require some initial training. In addition, its analytical and data presentation capabilities are limited so data will probably need to be exported to another product for analysis. Other general considerations are the cost of software and compatibility between software applications.

An alternative database option is to enter the data directly into a statistical package such as SPSS(tm) (Statistical Package for the Social Sciences). The screen appearance of the spreadsheet may not be so user friendly but this option eliminates the need to export and import data from one software product to another. Transporting data between products can be very simple or it may be time consuming and fraught with difficulties of compatibility.

Once the software is chosen, then a database needs to be set up. Each spreadsheet has columns, which are used for the variables, and rows which are used for individual patients or participants. The field in a database is the column ascribed to a particular data item. There may be several fields that relate to a single variable. For example, Table 3 presents a small section of a spreadsheet set up to record responses from a survey. 'Item 5' relates to a response to a survey question. This item was initially recorded on the paper record in a text format and was transcribed into the database in this format. The variable was then recoded using numerical codes into 'Item 5R'. Item 6 has been managed in the same way. In column eight, the two variables

'Item 5R' and 'Item 6R' have been computed (in this case added) to form a new variable 'subscore', which is a summary score.

Table 3: An example of a spreadsheet

Participant ID	Date of birth	Postcode	Item 5	Item 6	Item 5R	Item 5R	Sub-score
6675	24/07/1996	2144	Rarely	Often	1	3	4
6676	13/05/2004	3300	Sometimes	Rarely	2	1	3
6677	08/08/08	9999	Often	Sometimes	3	2	5

Finally, if a researcher is using data already recorded, they will be able to import data from the health service records to the project database. The compatibility of the datasets must be tested and it may be wise to design a database around the format of the existing records. Information technology staff at the health service will be able to offer advice in this regard. Once the database has been set up, it should be tested using pilot data entered on the data collection sheets or transferred from existing data sets. Necessary modifications should be made before the project commences.

Summary

Nurse practitioner services are a new addition to the Australian healthcare system and have the potential to optimise the quality of care for many sectors of the population. Rigorous evaluation of nurse practitioner services is necessary to provide evidence to demonstrate the exciting potential of these new services. Managing data to optimise consistency and accuracy is an important component of any clinical audit or research project. Planning for data consistency is often not well executed. Important aspects of planning include determining the optimal level of data collection; how many participants are needed for an adequate sample size; the format and coding of data items and what data analysis is needed. Rigorous and comparable data management processes are an essential part of good clinical audit and research. It is therefore important that nurse practitioners are well-trained in robust research methodology in order to demonstrate both efficacy and clinical effectiveness.

References

Nurse Practitioner Workforce Planning Minimum Dataset. (2006). Retrieved 1 August, 2007, from http://www.nnnet.gov.au/publications.htm

Bowling, A. (2002). *Research methods in health: Investigating health and health services (2nd ed.)*. Philadelphia: Open University Press.

Chew, D., Allan, R., Aroney, C., & Sheerin, N. (2005). National data elements for the clinical management of acute coronary syndromes. *Medical Journal of Australia, 182* (9), S5-S14.

Gardner, A., & Gardner, G. (2005). A trial of nurse practitioner scope of practice. *Journal of Advanced Nursing*. 49(2), 135-145.

Gardner, A., O'Connell, B., Takase, M., Condon, C., Ski, C., Baker, L., Ostaszkiewicz, J., & Hanna, B. (6-8 August 2007). *A structured analysis of medication errors: a multi site study*. Poster session presented at the 5th Australasian Conference on Safety and Quality in Health Care, Brisbane.

Gardner, G., Gardner, A., MacLellan, L., & Osborne, S. (2003) Reconceptualising the objectives of a pilot study for clinical research. *International Journal of Nursing Studies, 40*, 719-724.

Last, J. M. (2001). *A Dictionary of Epidemiology* (4th ed.). Oxford: Oxford University Press.

Munro, B. H. (2005). *Statistical Methods for Health Care Research (5th ed.)*. Philadelphia: Lippincott Williams & Wilkins.

Pallant, J. (2005). *SPSS Survival manual (2nd ed.)*. Crows Nest: Allen & Unwin.

Shaughnessy, J., Zechmeister, E., & Zechmeister, J. (2003). *Research methods in psychology*. New York: McGraw Hill.

Rapley, P. (1994). Managing quantitative data. In Jeanette Robertson (Ed.), *Handbook of clinical nursing research* (pp. 119-134). Edinburgh: Churchill Livingstone.

Watts, R. (1994). Managing qualitative data. In Jeanette Robertson (Ed.), *Handbook of clinical nursing research* (pp. 135-150). Edinburgh: Churchill Livingstone.

Appendix 1: Relevant websites

Australian Bureau of Statistics 1285.0 - Demographic Variables (1999)
http://abs.gov.au/ausstats/abs@.nsf/Latestproducts/

Australian Institute of Health and Welfare. Data Online
http://www.aihw.gov.au/dataonline.cfm

Australian Institute of Health and Welfare. National Data Dictionary (2003)
http://www.aihw.gov.au/publications/hwi/nhdd12/nhdd12-v1.pdf

N3ET - National Nursing & Nursing Education Taskforce (2004-2006)
http://www.nnnet.gov.au/

National Health & Medical Research Council. National Statement on Ethical Conduct in Human Research (2007)
http://www.nhmrc.gov.au/publications/synopses/_files/e72.pdf

National Health & Medical Research Council. When does quality assurance in heath care require independent ethical review?
http://www.nhmrc.gov.au/publications/synopses/e46syn.htm

The Office of the Privacy Commissioner .Health Guidelines
http://www.privacy.gov.au/health/guidelines/index.html#1

The Office of the Privacy Commissioner. State and Territory Privacy Laws
http://www.privacy.gov.au/privacy_rights/laws/