Challenges and pitfalls of data collection

Kerrie Westaway

Introduction

Data collection is a vital phase of the research process. The purpose of data collection is to obtain accurate, high quality meaningful data that are maximally effective in answering the research questions. Without this, the robustness and correctness of the study will be questioned. Subsequently, effective management and execution of the data collection process is essential.

Even with considerable time and effort devoted to the process it can be a challenging and unpredictable time. It frequently takes longer than anticipated, is often more demanding and difficult than first thought and can take the researcher through a kaleidoscope of emotions from feeling intense excitement and a sense of achievement to feelings of self-doubt and anguish.

This research project was no exception and several issues arose during the process of data collection that both confronted and challenged the researcher. This paper will discuss some of those challenges and pitfalls faced when collecting quantitative data for a descriptive survey that investigated the management of the International Normalised Ratio (INR) in warfarin therapy during the initial five months of treatment. It will also elucidate to the strategies that were employed to overcome those challenges in order to obtain the intended quality and quantity of data within the limits of the study.

Background

Warfarin has long been proven to be clinically effective in the primary and secondary prevention of venous thromboembolism, for the prevention of systemic embolism with prosthetic heart valves, for the prevention of acute myocardial infarct in patients with peripheral arterial disease and for the prevention of stroke, recurrent infarction or death in patients with acute myocardial infarction (The Medical Research Council's General Practice Research Framework 1998). Most recently, random trials have shown that the incidence of stroke in patients with atrial fibrillation is reduced by 60-70% when treated with warfarin (Samsa et al. 2000).

The realisation of the clinical and cost effectiveness of warfarin in the treatment of thromboembolic disease, especially for the reduction in the risk of stroke in patients with atrial fibrillation, has given rise to a worldwide

increase in the number of patients receiving warfarin (Murray 2003). Warfarin use in Australia has increased between 6%-9% per annum in the last four years (Australian Pharmaceutical Index 2004).

Conversely, warfarin is a potentially hazardous drug and can cause life-threatening hemorrhagic complications, with research indicating that patients are at the greatest risk of bleeding in the first few months of treatment (Campbell et al. 2001). Warfarin remains a difficult drug to dose due to its narrow therapeutic index (Ho & Brighton 2002) and its interaction with numerous drugs (Baker et al. 2004). Additionally, poor patient compliance, the presence of co-morbid conditions, patients of an older age (>75 years) and therapeutic activity that deviates from current best practice impact on the quality of anticoagulant control achieved (Campbell et al. 2001).

Several studies have shown that excessive anticoagulation with a prolonged INR above 4.0 increases the risk of bleeding complications exponentially (Gallus et al. 2000; Levine et al. 2001; Panneerselvam et al. 1998). Approximately one in six INR values will exceed the desired range, particularly during the first few months of therapy and the resultant risk of a major bleed is estimated to be approximately 5% in the two weeks following an INR above 6.0 if the patient is treated with warfarin withdrawal alone (Oden & Fahlen 2002).

Consequently, poor control of warfarin therapy may lead to increased morbidity and mortality, frequent doctor visits, longer and more frequent hospital stays and resultant increased health and hospital costs (Rigby et al. 1999). Thus, it is becoming more important to clearly define the safety issues surrounding oral anticoagulant therapy. As a descriptive survey design, this study was the first step in identifying and quantifying the gaps that may exist in the current system.

Data Collection Procedure

The research project was conducted in a large metropolitan teaching hospital in Adelaide, South Australia and in GP surgeries within the surrounding districts. As this study was a hypothesis generating study, the sample size was based on accuracy required for prevalence estimates. Basing the sample size requirements on a prevalence of 50% and requiring an accuracy of ±5% with 90% confidence, a sample of approximately 300 patients was required.

Patients initiated on warfarin who were thereafter followed up by their general practitioners (GPs), other health institutions or hospital clinics were included in the study. Because the phenomena under investigation were complex, data collection, which required 20 months to complete, was multifaceted. First, the study incorporated data from the patients' medical records and included basic patient demographics, relevant

medical information pertaining to their warfarin therapy and treatment and outcomes of any episodes of excessive anticoagulation with a prolonged INR above 4.0.

Second, on discharge from the hospital, the researcher contacted each patient once a month for five months or less if the warfarin was ceased prior to this time to determine if they had incurred an INR above 4.0. If they had, they were asked a series of questions pertaining to their health status and compliance to therapy during the week leading up to the episode. At the time of the initial INR above 4.0 only, they were asked about their understanding of warfarin therapy in general. The patient was also asked to provide their warfarin doses, INR results and the corresponding dates if able to.

Third, prescribing doctors both in the hospital setting and the community were asked to complete two questionnaires. The first questionnaire sought to obtain information regarding the medical management of patients receiving warfarin therapy and the second questionnaire sought to obtain information pertaining to the causes and treatment of an INR above 4.0 when it occurred. Lastly, on completion of the five months, after consent was gained from the respective GP or health institution to access the patient's medical records, accuracy of data was confirmed and any missing data was collected.

Challenges and Pitfalls of Data Collection

Some of the challenges and pitfalls of data collection of this study related to conducting a pilot study to evaluate the questionnaires and the feasibility of obtaining data in such a way that was practical and effective, time constraints, cost factors, and ethical issues. A separate challenge was dealing with the emotional ups and downs of the day-to-day circumstances of working alone as a novice researcher, overcoming the abovementioned challenges and keeping abreast of the large amount of data generated each day.

The Pilot Study

Prior to commencement of data collection, two major challenges confronted the researcher. The first challenge was to pretest and evaluate the questionnaires that had been designed and constructed for the study. According to Polit and Hungler (1999) after questionnaires have been drafted and critically discussed with various individuals who are knowledgeable about the construction and identification of technical difficulties associated with questionnaires, they can be pre-tested. The second challenge was to evaluate whether the data that was required was going to be accessible and if it was, how to go about collecting it in such a way as to cause minimal disruption in the busy routine of participants and in a way that would support participation in the study. Burns and Grove (1997) caution that when research involves people and all of their

complexities, the process can be unpredictable and challenging and during the course of data collection this proved to be very true.

Roberts and Taylor (1998) recommend that a pilot study should be a smaller version of the main study and should be conducted as far as possible in the same type of setting involving the same type of participants. The purpose of a pilot study is to evaluate the feasibility of the study and refine instruments, determine the length of time it takes to administer a complete package of instruments and to determine whether the time burden is too great or not for participants (Polit & Hungler 1999).

Thus, the first step was to conduct a pilot study to assess the feasibility of the study and to pre-test the questionnaires, evaluate the methodology, make adjustments if necessary and document and analyse the data prior to commencing the main study. After the Human Research Ethics Committee (HREC) at the main study institution granted approval to conduct a pre-test of the patient questionnaires, four patients already receiving warfarin therapy in the hospital were approached. The interviews were conducted in such a way as to mimic the process expected in the main study.

The results of the patient pre-test indicated that all patients approached to participate in the study agreed to do so. The questions asked were clearly understood, none were found to be objectionable or threatening in any way and the questions flowed on from one to the next. Although Polit and Hungler (1999) recommend 10-20 pre-tested questionnaires prior to commencing a study, the researcher declined from interviewing any further patients in the pre-test phase because the comments made by all four of the participants were congruent with each other and it was highly unlikely that any new comments would be made. The researcher was also reluctant to conduct a larger pre-test given the ready availability of patients fitting the criteria, the possible small attrition rate, accessibility to the quantitative data from patients' medical records and similar times taken to complete each questionnaire.

In order to pre-test the questionnaires for the prescribing doctors, a metropolitan teaching hospital was identified away from where the main study was to be conducted. This was done to avoid alerting potential participants at the main study hospital as to the nature of the study. On receiving ethics approval from the HREC, the researcher approached nine doctors to participate. The interviews were conducted in such a way as to mimic the process expected in the main study. The results of the interviews indicated that no questions appeared to cause offense but some did require re-wording to be better understood. The length of time it took to complete each questionnaire was also noted. Other vital information with regard to how, when and where to contact the doctors, whether they would participate or not, whether conducting an interview in their busy schedules was a realistic expectation or not and what the attrition rate might be was also evaluated and

documented. Additionally, the process enabled the researcher to become familiar with interviewing in the real setting and to develop a degree of confidence towards the process.

While all doctors approached, agreed to participate in the pre-test, three of the nine doctors requested to complete the questionnaire in their own time. This alerted the researcher to the possibility that conducting an interview with doctors may not be possible to accomplish. This in time proved to be the case when the main study was commenced. While all hospital-based doctors agreed to participate, finding a time to do so proved very difficult. After several failed attempts at organizing a time for the interviews, it was decided that the doctors would complete the questionnaire in their own time and return to it to the researcher on completion. However, this did not solve all of the problems. While all doctors were agreeable to completing the questionnaire, many failed to do so. The researcher kindly prompted them several times over a space of days to several weeks, however, some doctors failed to ever produce the completed questionnaire.

Burns and Grove (1997), advise that there is a loss of participants to some extent from all studies for various reasons. In this study, a number of hospital-based doctors who originally agreed to complete the questionnaire failed to follow through. The researcher proposed that the major contributing factors were busy workloads and time constraints placed on the doctors. Burns and Grove (1997) suggest ways to overcome this problem include increasing the sample size or to continue data collection for a longer period of time to achieve the set sample size. However, none of these options may be feasible and thus a smaller than expected sample size may result, which in turn may affect the power of the planned statistical analyses and ultimately prove inadequate to test the hypotheses (Burns & Grove 1997). In this study, the researcher was unable to increase the number of planned participants as all treating doctors associated with each patient that could be contacted were already asked to participate. Additionally, the sample size of the study was based solely on the primary outcome, that is, the patients and not the doctors.

A similar challenge arose when trying to contract the hospital-based doctors to complete the questionnaire pertaining to causes and treatment of patients who had incurred an INR above 4.0. In several instances, a prolonged period of time of four to eight weeks had passed from when the INR above 4.0 had occurred to the time when the doctor could be contacted or was able to complete the questionnaire and the doctor's memory of that specific patient and incident was not clear, resulting in a less than accurate record. Additionally, while doctors were agreeable to participate in completing one questionnaire, they hesitated at having to complete more than one. This same challenge was faced with the GPs and it became immediately obvious to the researcher that, due to the constraints on GPs in the current health system, it would be first, extremely difficult to contact them in a timely fashion and second, impossible for them to complete a questionnaire for each patient as often as would be required. As difficult and disappointing as it was to make the decision to delete the second questionnaire from the study, it was decided that the researcher would obtain as much

information from the medical notes and the patients as was possible with regards to cause and treatment of episodes of excessive INRs above 4.0.

In order to pre-test the questionnaires amongst the GPs, questionnaires were mailed to 61 GPs practicing in the northern suburbs, a geographical area not expected to be involved in the main study, with a letter of introduction, an explanation of the nature of the study and a self-addressed envelope for ease of return. A request to answer each question, provide feedback on the time taken to complete each questionnaire and any other comments regarding the content of the questionnaires was also made. According to McMillan and Schumacher (2006) response rate from the initial mailing of questionnaires with a stamped return-address envelope will usually be 40-60%. Of the 61 GPs in this pre-test 9 (15%) responded to the request. Crookes et al. (2004) caution that if the response rate is less than 50% the researcher must be wary when making statements of generalisation to the wider population as the survey respondents may differ from non-responders.

The poor response rate highlighted the second major challenge of the study and that was to engage the participation of GPs and subsequently gain access to the patients' medical records. The poor response rate may have been attributed to the current critical shortage of GPs working in Adelaide and their subsequent demanding workload. According to Lisa Allison (2005:004), a medical writer for the local newspaper, The Advertiser, the chronic shortage of GPs in Adelaide has reached a crisis point in the outer suburbs of the city with the patient ratios being reported to be as high as 1:5521 by the Southern Division of General Practice in the worst affected southern districts of Woodcroft, Reynella and Old Reynella and 1:7596 in the worst affected northern area of Williamstown. According to Allison (2005:004) the Federal Government's ratio of 1:1400, which was set as a benchmark to define doctor shortages, has been exceeded.

The Main Study

In order to overcome the potential problem of a poor response rate of GP participation in the main study, it was decided to offer reimbursement for any time spent providing information or completing the questionnaires which would hopefully encourage more participation and subsequent access to patient's records. The amount of reimbursement was set in accordance with the Southern Division of General Practice's policy for participation in divisional activities. After recruitment of the patient, the GPs were sent a copy of the patients' written consent and a request for permission to access patient data on completion of the five-month study period. At the end of each patient's study period, the questionnaire was mailed to the GP with a request to complete. This method proved successful with an estimated 95% of GPs granting access to patient's records and 66% completing the questionnaire. However, not all GPs who initially agreed to allow access to the patient's medical records did so at data collection time and thus those data for that patient was not able to be checked or was completely unobtainable, resulting in missing data.

Burns and Grove (1997) describe this as passive resistance where health professionals and institutional staff working with the subject have an impact on the data collection. This was again experienced when nursing staff were requested to flag potential patients within the hospital setting. All staff were initially agreeable to flag potential patients but often failed to do so. The researcher was acutely aware that the real-life daily demands of work and time constraints contributed significantly to the lack of potential patients flagged. Roberts and Taylor (1998) suggest that rewarding the involvement of clinicians in a study often encourages further participation. While the researcher decided to provide staff with chocolates every one to two months, they were given in appreciation and acknowledgement only of any participation by staff and it was not inferred that chocolates would motivate staff to flag more potential patients or solve the daily work issues they confronted. The researcher acknowledged that without the help of the staff it would have been impossible to flag the patients and subsequently conduct the study.

Time factors

Although much consideration was given to the time factors involved prior to data collection, many aspects of the study not considered initially required more time than was anticipated. Burns and Grove (1997) advise that data collection requires two to three times longer than initially thought. This is generally due to the fact that events during data collection are often not under the control of the researcher such as a sudden lack of potential participants or unpredictable heavy workloads of staff (Burns & Grove 1997).

This was evident in several areas of this study, particularly when recruiting patients in their homes. Frequently much time was lost in attempting to contact patients by telephone and the journey itself often required one to two hours of travel. While some patients wanted to discuss aspects of their warfarin therapy and relay their grievances with the current health system, some were simply lonely and wanted some company for a while. Collecting data from the GP surgeries was also fraught with long waiting times (once again reflecting current health system issues) as was the process of collecting data from other health institutions, especially from private hospitals. In some instances, it required up to three to four weeks of formal communication, multiple reminder telephone calls, set appointments and time to travel to the various surgeries and institutions to collect data.

Not all of the time factor issues could be solved, however, several strategies were set in place by the researcher to improve time management. In order to minimise time spent recruiting in the patient's home and travel times, the researcher informed the patient on admittance to the home of previously set engagements and an attempt was made to recruit patients in the same area in the one day if possible. If waiting times for appointments were prolonged, the researcher went prepared with paper work to carry out while waiting.

Cost factors

The cost of reimbursing GPs and additional costs that surfaced as data collection progressed were a source of anxiety to the researcher. There were costs associated with the researcher's time, costs for petrol, printing and postage and costs for consultation with a statistician. The decision was made to apply for a number of scholarships. Although this was a very time consuming process and required close communication and much help from the researcher's supervisors, this decision proved to be successful and provided a most fruitful source of finance which was well worth the time and effort required.

Ethical factors

The key ethical issues during data collection of this study surrounded the maintenance of human rights of the participants, that is, their right to privacy, anonymity and confidentiality. All potential patients were flagged by hospital staff prior to the researcher approaching them. Informed written consent was gained prior to access to their medical records and all data collection forms were coded to de-identify participants. Those patients discharged to the hospital-at-home program were verbally informed of the current study, provided with a 'Participant Information Sheet for Patients' and asked if they were agreeable for the researcher to contact them via hospital-at-home staff while making their daily visits. All raw data was collected by the researcher only and kept in a locked cabinet in the researcher's office. Access to data was limited to the researcher only.

Management of data

During the initial stages of data collection, keeping abreast of the large amounts of data that were generated daily was a major concern to the researcher. Because data collection for this study was both long and complex and large quantities of data were generated, it was essential to manage the data in an effective and practical way. Blaxter et al. (1996) suggest that meticulousness is an important skill for the researcher to acquire in recording the progress of the project, as is avoiding missing data and providing complete sets of data for each subject where possible. The most practical method of management in this study was to draw up a flow chart on butcher's paper that was meticulously updated and maintained on a daily basis throughout the duration of data collection. Additionally, the researcher documented ideas, comments and conversations with professionals and any relevant information regarding the study in a large workbook. Consequently, any confusion was avoided and much stress and anxiety was alleviated.

The emotional journey of data collection

Lastly, it is important to acknowledge the emotional journey that data collection invariably takes the researcher on. As Blaxter et al. (1996) suggest, some days will be filled with excitement and immense interest while other days will require all effort simply to accomplish the required workload. The authors propose that although working alone is an expected part of research, feelings of loneliness and obsessiveness are the two most common 'downs' encountered during research and are particularly prevalent

during the data collection phase. This is especially true when there is no one to discuss or share challenges and progress with and every spare minute is spent either thinking about the research project or actually doing it (Blaxter et al. 1996).

While it is easy to identify with these emotional experiences during this study, strategies that were found to be helpful (especially since the data collection phase was both complex and long) were validation that these emotions are common and normal, regular and frequent contact with supervisors, family support and compartmentalizing the research. Blaxter, Hughes and Tight (1996 p.167) state compartmentalizing research is to ‘give it a certain time and space in your life, but no more, making sure that you leave yourself opportunities to maintain and engage in some of your other interests'. The authors also suggest three basic strategies to combat obsessiveness. First, plan, schedule and revise plans from the initial stages of the research to keep the required work feasible. Second, initiate a family member or good friend to identify signs of encroaching obsessive behaviour and third, develop a network of fellow researchers to share experiences and concerns with (Blaxter et al. 1996).

In summary, although considerable time and planning was invested in the data collection phase prior to commencement, it proved far more complex and time consuming than ever anticipated. This, it would appear is a common occurrence. Many days were frustrating and difficult, but many days were exciting and fulfilling especially when problems were overcome. The secret to a smooth and effective data collection phase can be seen in, not only the pre-planning of data collection, but conducting a pre-test or pilot study and the subsequent evaluation of those results prior to the main study. While it is unlikely that all problems will be solved in the pre-test phase, many problems will be and the process will enable the researcher to commence data collection in the main study with a sense of confidence. Additionally, flexibility in dealing with unexpected problems while maintaining the integrity of the study, consistency in data collection, meticulous management of data collected and maintaining positive relationships with participants and other professional staff members involved in the study can be clearly seen as beneficial.

REFERENCES

Allison, L. 2005, 'Crisis point: one doctor to 7596', *The Advertiser*, 28 October, p.004.
Australian Pharmaceutical Index. 2004, IMS Health, Sydney.

Baker, R.I., Coughlin, P.B., Gallus, A.S., Harper, P.L., Salem, H.H. & Wood, E.M.; The Warfarin Reversal Consensus Group, 2004, 'Warfarin reversal: consensus guidelines, on behalf of the australasian society of thrombosis and haemostasis', *Medical Journal of Australia*, vol.181, no.9, pp.492-497.

Blaxter, L., Hughes, C. & Tight, M. 1996, *How to Research*, Open University Press, Philadelphia.

Burns, N. & Grove, S.K. 1997, The Practice of Nursing Research: Conduct, Critique, & Utilization, 3rd edn, Saunders, Philadelphia.

Campbell, P., Roberts, G., Eaton, V., Coghlan, D. & Gallus, A. 2001, 'Managing warfarin therapy in the community', *Australian Prescriber*, vol.24, no.4, pp.86-89.

Crookes, P.A., Davies, S. & Chiarelli, M. 2004, Research into Practice: Essential Skills for Reading and Applying Research in Nursing and Health Care, 2nd edn, Bailliere Tindall, London.

Gallus, A.S., Baker, R.I., Chong, B.H., Ockelford, P.A. & Street, A.M. 2000, 'Consensus guidelines for warfarin therapy. Recommendations from the Australasian society of thrombosis and haemostasis', *Medical Journal of Australia*, vol.172, pp.600-605.

Ho, L.L. & Brighton, T. 2002, 'Warfarin, antiplatelet drugs and their interactions', *Australian Prescriber*, vol.25, no.4, pp.81-85.

Levine, M.N., Raskob, G., Landefeld, S. & Kearon, C. 2001, 'Haemorrhagic complications of anticoagulant treatment', *Chest*, vol.119, pp.108s-121s.

McMillan, J.H. & Schumacher, S. 2006, *Research in Education, Evidence-Based Inquiry* 6th edn, Pearson Education, Boston, United States of America.

Murray, E. 2003, 'Anticoagulation management in primary care', *Practice Nurse*, vol.26 i8, pp.34-40, retrieved 16 August 2005 from Expanded Academic ASAP database.

Oden, A. & Fahlen, M. 2002, 'Oral anticoagulation and risk of death: a medical record linkage study', *British Medical Journal*, vol.325, pp.1073-1075.

Panneerselvam, S., Baglin, C., Lefort, W. & Baglin, T. 1998, 'Analysis of risk factors for over-coagulation in patients receiving long-term warfarin', *British Journal of Haematology*, vol.103, pp.422-424.

Polit, D.F. & Hungler, B.P.1999, *Nursing Research. Principles and Methods*, 6th edn, Lippincott, Philadelphia.

Rigby, K., Clark, R.B. & Runciman, W.B. 1999, 'Adverse events in health care: setting priorities based on economic evaluation', *Journal of Quality Clinical Practice*, vol.19, no.1, pp.7-12.

Roberts, K. & Taylor, B. 1998, Nursing Research Processes: An Australian Perspective, Nelson, Melbourne.
Samsa, G.P., Matchar, D.B., Goldstein, L.B., Bonito, A.J., Lux, L.J., Witter, D.M. & Bian, J. 2000, 'Quality of anticoagulation management among patients with atrial fibrillation', *Archives of Internal Medicine*, vol.160, pp.967-973.

The Medical Research Council's General Pracice research Framework, 1998, 'Thrombosis prevention trial: randomized trial of low-intensity oral anticoagulation with warfarin and low-dose aspirin in the primary prevention is ischameic heart disease in men at increased risk', *Lancet*, vol.351, pp.233-241.