Work in progress – not to be quoted without authors' permission

Paper for discussion at Health Economists' Study Group Meeting
University of Oxford, January 2005

Appropriate estimation of staffing costs in economic evaluations

T Nicholson, K Gerard and P Roderick

Health Care Research Unit, University of Southampton
Mailpoint 805, South Academic Block, Southampton General Hospital, Southampton SO16 6YD
Email : T Nicholson: apn@soton.ac.uk; Tel 02380 794775

Abstract

A key cost driver in many economic evaluations is staff, though the precision required of estimates depends on the research question. However, a current challenge is the appropriate allocation of staff time to the costs for particular patient groups or interventions, e.g. in situations that are complex due to casemix, staff roles, location or nature of service delivery.

The paper describes work in progress. Preliminary work considered the characteristics that may help predict the type of economic evaluation where measurement of staff time is important. The paper describes methods available to measure staff time within and beyond health economics, and presents a few examples of economic evaluations as case studies. In addition, the paper briefly reviews literature on the use of bar code scanners for time measurement in health care.

The latter part outlines the proposed research study to measure time nurses spend with different types of patients using bar code scanners. The study aim is to examine what difference, if any, the measurements make to the estimates and precision of staffing costs compared with a priori assumptions. Additional issues addressed include the feasibility of such data collection based on experience to date and acceptability to staff. These data would feed into a value of information analysis to help inform the future economic evaluations.

Acknowledgements

The National Co-ordinating Centre for Research Capacity Development (Researcher Development Award) is funding this work. The views expressed in this paper and any errors are those of the authors alone.

1. Appropriate estimation of staffing costs in economic evaluations – what is the problem?
Cost analyses are central to economic evaluations and programme budgeting and marginal analysis. When planning a cost analysis, it is important to ensure that adequate research effort goes into valuing the utilisation of resources that are most likely to affect the choice between alternative courses of action. This will enable decision makers to make the 'correct' decision and decrease their decision uncertainty.

Broad advice on costing is available to researchers, for example by Brouwer et al (2001). Thus, the research question should drive the level of detail required for the cost analysis, and together with the perspective adopted, guide the resource use categories identified, measured and valued. Similarly, the study's purpose should direct whether predominantly gross costing ('top-down')[a] or micro costing ('bottom-up')[b] is appropriate. Potentially the former has the advantage of greater generalisability, whilst the latter provides greater precision and insight into contributory cost drivers.

However, one current and pressing challenge facing health economics is the appropriate costing of staff time. Staffing is often a key cost driver since salaries account for 62% of NHS expenditure. The challenge comprises two issues. Firstly, there is a need for greater understanding of the circumstances under which it is appropriate to apply 'quick and dirty' rather than in depth methods of allocating and costing staff time. Examples that need in depth methods might include where there are differences between patients within a heterogeneous group due to casemix. Alternatively staff roles or service delivery may differ across patients and not be reflected accurately in costs averaged across the patient group.

Secondly, when more precise cost estimates are necessary, what method of time measurement is suitable? There are little routine NHS data available about the amount of time staff spend on different patient-specific tasks. Often, isolation of the costs of interest is impossible from the aggregated data available at Trust level. For example, staffing expenditure usually covers relatively broad patient groups through procedure codes or Health Resource Groups (HRGs). Similarly, average length of stay often drives the HRG based NHS Reference Costs (National Tariff), but this conceals differences between patients in staff time required.

The value of information intertwines the two issues above, since logically staff costs that potentially have a large impact warrant more research effort. However, in depth staff costing may be expensive or unfeasible in some situations.

[a] Apportioning the total cost (including all overheads etc.) to cost centres (e.g. wards, consultants, etc.) or specialities, i.e. disaggregating total cost to lower levels. The method can derive costs at any level, i.e. per ward, bed-day, procedure, patient etc.
[b] Calculating the overall cost, e.g. of a procedure, bed-day etc., by summing the costs components calculated by multiplying each resource use element by its unit cost. For example, costing staff for a procedure involves multiplying the time taken to perform the procedure by the staff's hourly salary.

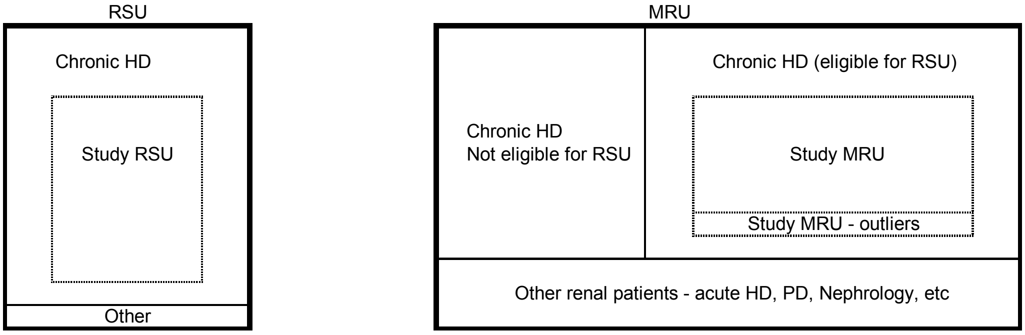
In summary, given that most health economists have limited 'medical' knowledge, the following questions seem important. Is it possible to predict evaluative situations where staff costs are important? How should researchers measure and value staff time to derive estimates of sufficient precision (accuracy and exactness) that are fit for purpose? Linked to answering these, the following question should be self-evident. Does the value of the research effort match the costs of undertaking the research?

This paper presents work in progress that forms one of the author's (TN) PhD thesis. The paper is in nine sections. The next section provides a background to the current paper and proposed research. Section 3 considers the characteristics that may help predict the type of economic evaluation where measurement of staff time is important. Sections 4 and 5 briefly review alternative methods of allocating staff time within and beyond health economics, and present some research papers as case studies. Section 6 is a brief literature review of the use of bar code scanners for time measurement. Sections 7 and 8 outline the proposed research study to use bar code scanners to measure staff time and the feasibility of this approach based on experience to date. The final section raises some possible issues for discussion.

2. Background to this paper

Concern over this issue arose in a study of chronic haemodialysis in main and their linked nurse-led satellite units (Roderick et al, in press). On face value, the costing task seemed straightforward, but Figure 1 illustrates the problem. In the renal satellite units (RSUs), only a minority of patients attended for reasons other than chronic haemodialysis, represented by the small 'other' section in the left part of the figure. However, at the main renal units (MRUs) patient mix was more diverse and complex, as shown in the right of the figure. Attendance for 'other' reasons affected a larger proportion of patient who attended for treatment of acute renal failure, complications of peritoneal dialysis and other nephrology problems. Furthermore, many chronic haemodialysis patients ineligible for nurse-led satellite care potentially utilised resources, particularly nursing staff, more intensely, for example due to treatment instability. In addition, the chronic haemodialysis workload sometimes spilled over with outlier patients on wards. Thus, RSUs comprised a more homogeneous chronic haemodialysis workload and the MRUs a more heterogeneous and less clearly demarcated one.

Figure 1: Study population



Accurately differentiating shared clinical staff resources between study and non-study patients to ensure like-for-like comparisons proved unworkable. Data on patient dependency or nursing requirements were unavailable to adjust staffing costs at the MRUs to reflect differences between patients eligible and ineligible for satellite care. The alternative was to assume that non-study patients required the same amount of nursing time as those eligible for RSU care. However, this potentially overestimates costs, as ineligible patients were more likely to be a sicker population and thus require a greater intensity of nursing input.

The research project engendered the wish to examine the issue of how to undertake the costing of staff time, although it is clear that such detail is not necessary in all cost analyses. Whilst an obvious choice for further research would be to return to look at renal dialysis, additional settings would better inform whether the technique chosen has wide applicability. The next section looks at possible characteristics of evaluations that might predict when in depth measurement of staff time would be fitting. Here, primarily the purpose was to inform the study design (so that it would reflect realistic situations), but additionally it was an attempt to help other researchers.

3. Characteristics of economic evaluations that might predict the importance of measuring staff time

An examination of the economic evaluation abstracts in the NHS EED database[c] was a prompt to identify a checklist of characteristics for further development. These items included country and currency; type of intervention (e.g. device, drug, service delivery); whether primarily modelling; setting; multi-centre study (for effectiveness and/or resource use); adequacy of reporting resource use and/or unit costs and a judgement of whether the described intervention's impact on staff was likely to be important. Initially the intention was to develop the checklist of characteristics through an iterative process, by using the checklist to identify full papers to pursue.

However, this strategy proved unhelpful after using the checklist on a small convenience sample of UK based-evaluation papers (45 of the 191 abstracts published 2003-2004, summarised in Appendix 1). In 14 of these studies, staffing costs seemed likely to be important, but this proved difficult to judge. Whilst staffing appeared important in all five psychological interventions and more than half the service delivery interventions (4/6), for others the proportions were smaller, i.e. drug (1/19), drug delivery modes (1/2), surgery/procedure (2/4) and other (1/2).

Without further information about the intervention, only briefly described in NHS EED abstracts, it was impossible to make a judgement about the potential impact on staffing. Even having clinical experience to draw on did not resolve the uncertainty about the impact on staffing in almost one-third of cases (13/45). To continue the process would have involved making arbitrary guesses. This was

[c] Using 2001 or 2002 or 2003 or 2004 (all fields) AND "This record was compiled by CRD commissioned reviewers according to a set of guidelines developed in collaboration with a group of leading health economists". Using "partial economic evaluation" (all fields) would have retrieved additional papers.

unhelpful in terms of prospective identification of economic evaluations where staffing issues would be important, especially as most health economists come from non-clinical disciplines.

Using NHS EED abstracts had the advantage that, in the absence of the actual article, more information was available than in the journal abstract, although it also proved a challenging data source. For example, “United Kingdom” in the country code field was sometimes misleading and there were inconsistencies between reviewers in reporting. Some reviewers stated that resource use and unit costs were not reported separately (not denoting which was absent), but others stated explicitly whether the lack of reporting applied to one or both. Furthermore, a preliminary examination of some of the full papers revealed the usual problems of data extracting from journal articles – description of costing methods was brief and uninformative. Therefore, it was impossible to gather the necessary data in a short timeframe.

Finally, the checklist appeared to muddle the following three issues. Firstly, identifying economic evaluations where comparison on a like-for-like basis might show a differential impact of staffing on results (both in resource use and cost terms). Secondly, determining how large such differences in staff costs might be. Thirdly, identification of potential barriers to detailed staff costing. For example, for pragmatic reasons, measurement is difficult where staff are geographically dispersed, or where information is unavailable on grades of staff.

Therefore, i.e. to tackle this a priori, it seemed necessary to go ‘back to basics’ and simply rethink the costing process and scenarios. From this, it appears that general characteristics that demarcate economic evaluations where there are likely to be differences in staff time and hence costs are those with:

1. differences in staff time to administer the intervention;
2. differences in staff time as a result of the intervention, i.e. where patient outcomes have an impact on staff contacts or length of stay;
3. the need for training patients and/or carers;
4. psychological interventions, including advice or counselling for patients and/or carers (i.e. staff time is the intervention);
5. role substitution – potentially with differences in times to deliver care, salary scales, and also training, specialisation, professional background or clinical skills. Of note, there may be interactions between different staff. For example, in haemodialysis, doctors and nurses deliver combined care at main units whilst care is usually only from nurses at satellite units;
6. differences between the comparison groups in terms of heterogeneity of patient casemix due to different disease/illness, procedures required or severity within a disease/illness group.

If the durations of patient-staff contacts are similar, data collection on the frequency of contacts may be adequate if they can be coupled with reliable ‘routine’ unit costs data (e.g. for admissions, outpatient visits, and appointments).

However, there is another challenge to choosing items that warrant measurement. Planning a cost analysis requires extensive knowledge of the resources to perform the intervention and treat any side effects. In addition, the researcher must consider how the epidemiology of the disease process before, during and after the intervention affects resource use and the duration of any changes. To access this information potentially requires discussion with the relevant health service stakeholders, given that most health economists do not have such a background. Whilst the clinical staff may be appropriate to identify which resource items differ, they may not be able to judge their magnitude.

Most researchers consider the selection of a meaningful cost difference the domain of decision makers. This is probably the biggest challenge to planning. Traditionally, decision makers who will use the research outputs, particularly at a local level, are outside the commissioning and design stages of research projects. Furthermore, differences in individual items may appear inconsequential until aggregated, so choice remains difficult. It is possible to set up sensitivity analyses to inform the data collection, but this too is costly and if populated by 'wrong' estimates still may lead to inadequate data collection.

These issues about identification and choice of resource items to measure highlight the fact that research involves or affects many stakeholders. They all view the same scenario, but from slightly different viewpoints. Therefore, whilst a checklist may be a useful prompt, it cannot replace good communication, in language understandable by all involved, to capture the scene and avoid missing issues vital to individual stakeholders.

There are also feasibility issues about measuring staff time. Section 4 considers the anticipated research expense and expertise factors affecting the choice of time study methods. For pragmatic reasons, there are additional aspects to consider that require considerable knowledge about the intervention and its delivery. These include the challenge of tracking relevant staff and patients due to geographical dispersal. The research question will help determine whether the main driver to the data collection is patients or staff. Researchers need to consider the involvement of staff (both who and where), in relation to intervention monitoring (prescription/planning and oversight) and delivery. Conversely, researchers could tackle the issue from the patient perspective, but either way may differ from the manner of the research project's data collection. Workload arrangements will affect the ease of data capture from staff. For example, it will be easier where study patients form the major workload (e.g. specialised clinics etc.); patients have predefined appointments; or patients are relatively homogenous; but these may differ between the comparators. In addition, the nature of the staff-patient contact affects the method of time measurement (i.e. whether face-to-face, by telephone or other e.g. video conferencing, postal). Interventions delivered by input from multiple members of staff, and rapid changes of task or patient will also influence the feasibility of data collection.

Measurement of staff time may appear necessary to address appropriately the research question. Only if such research is truly unfeasible or too costly in relation to the value of the information is its omission acceptable. The next section considers methods to measure staff time.

4. Methods available to measure staff time

It seemed important to examine the management literature since the origins of time study lie in manufacturing and production. At this stage, it was simply a brief review based on books to learn from another discipline and apply the knowledge to health economics, rather than a comprehensive literature review.

Time study techniques complement those of motion study to measure and improve productivity. The five main techniques are stopwatch time study; work sampling; predetermined time standards; standard data; and expert opinion time standards/historical data. Traditionally, time measurements are in decimal minutes (or hours) using in order of increasing accuracy mechanical or digital stopwatches, and computers.

Stopwatch time study is the best-known method amongst employees and there are two principal approaches. Continuous time measurement records from the start of the first activity until the end of the study, but it requires subtraction of each time from the preceding one to derive the individual elements. In contrast, the snapback method stops the watch at the end of each activity. The observer notes the time and then zeros the watch again before starting to record the next activity. The continuous technique has integrity as it accounts for the whole period and therefore cannot conceal any elements and is the preferred technique of unions according to Meyers and Stewart (2002).

In order to avoid missing consecutive activities, a three-watch study uses three continuous watches. Pulling a side lever simultaneously stops one watch to allow reading of the time, restarts one watch for the next task, and zeros the third watch ready for the following task. However, whilst this provides elemental time, it does not account for the whole period.

In Methods Time Measurement (MTM), the time unit is one-thousandths of an hour (time-measured unit, TMU) and therefore it uses a specific stopwatch that measures time in TMU. Digital stopwatches often combine the continuous and snapback facilities and some have a memory to store readings. A variety of hand-held computers are available (e.g. Datamyte and RateSetter) that automate calculations and improve accuracy. Furthermore, videotapes may act as stopwatches especially if there is an in-built time recorder. An additional time measurement tool is the tachometer. This determines the speed of machines and conveyors from the number of revolutions per minute, or in lorries gives the number of driving hours.

Work sampling applies the principle used in Gallup polls and other surveys of random sampling based on the theory of probability. At random times, observation of staff determines how they spend their time on predefined tasks. Ratios (as percentages) for each task are the number of observations of the task divided by the total observations. Application of this ratio to the total staff working hours covered by the study period gives the time per task. The technique is especially suitable to large production volumes or where the duration of observation required is long. Here it is more efficient than a stopwatch study as the observer does not have to watch the entire process each time. An alternative name for work sampling appears to be Multi-Moment Measurement (de Keizer et al, 1998); although in this case, the observations were at 10-minute intervals.

Predetermined Time Standards (PTS) is actually a methods and time study that involves breaking down a product manufacturing plan into workstations, each of which has a motion pattern. The time standard comprises the aggregated times assigned to each motion available from a table of previously measured reference values. Its main use is before production starts to help plan the equipment and staff required and associated costs. Varieties of specific Predetermined Time Standards Systems (PTSS) exist, for example Methods Time Measurement (MTM) and Modular Arrangement of Predetermined Time Standards (MODAPTS). Similarly, as the basis for working out why jobs vary, standard data uses previously set time standards, sometimes calculated from formulae, e.g. walking time as a function of the distance covered.

Expert opinion time-standards are simply prospective estimates of the time required. Historical data can be useful for estimates where there are previous recordings of task durations, although they conceal past inefficiencies if the task took longer than it should have.

In health care, predetermined time standards and standard data rarely exist. These methods are also less suitable as usually the basis of health care working is not the workstation concept. Some historical data are available, for example in theatre records detailing the start and finish times of operations.

Both stopwatch time study time and work sampling rely on an external observer. This poses a particular challenge in service industries, as simple observation cannot easily identify all tasks linked to a particular product. For example, in health care, patient-specific indirect care (i.e. not face-to-face) would be difficult to capture without questioning the staff as to whom the activity concerned. In addition, both stopwatch and work sampling studies potentially are costly as they require numerous or knowledgeable researchers.

Interestingly, self-recording does not appear to be an accepted time study technique in time-study books. This is presumably because the logging process interferes with the task and is therefore unacceptable during production processes. However, where the focus is to capture indirect patient care rather than activities per se, self-recording seems the only practical solution. Potentially newer technology can minimise the disruption compared with traditional hand-written paper logs. Hand-held bar code scanners that date and time stamp each scanned activity offer a simple solution to hand entries. They have the advantages of allowing simultaneous recording by multiple staff, only require uploading to computer not data entry, and when combined with other software can automate the data processing.

Stopwatch study, work sampling and self-recording are all subject to bias if the observer/recorder consciously or accidentally omits activities. Most staff are likely to be suspicious of examination of their working practices. Therefore, good communication about the nature of and involvement in the project are likely to be vital to the success of data collection and credibility of the results.

The next section looks at how researchers have handled the issue of costing staff time in economic evaluations. Section 6 returns to the use of scanners and reviews literature specifically on the use of bar code scanners in measurement of staff time in health care.

5. Examples of handling of costing staff time from economic evaluation literature

Five papers were selected from the journal Health Technology Assessment to use as case studies where costing staff time was likely to be important. This journal was useful, as it has virtually unrestricted reporting. For convenience, the articles (chosen by their titles) were about role substitution.

The papers illustrated four main points. Firstly, there was often a lack of transparency in reporting so that readers had to accept the results at face value. Costing methods for staff time (or the rationale) rarely received much attention in research papers. Though this may be explainable due reporting restrictions, it could simply reflect researchers' lack of concern or insight into the data they used. On other occasions, there were disturbing reporting inconsistencies. For example, questions purported to be in the data collection instruments were missing from the paperwork shown in the appendices. Researchers may have simply included the wrong version, but it is a worrying omission for items so crucial to the economic evaluation.

Secondly, there seemed to be little attempt to check the validity of the time data, although this is often true of the general study data. Thirdly, the presentation of time data was not always appropriate or adequate for economic evaluations. Examples include presentation of only median rather than mean times and staff costs not disaggregated into resource use and cost data. Fourthly, self-reporting appeared a popular choice, but whilst time measurement is not easy, there seems little justification for recording activity times after the event as these are unlikely to be accurate. In addition, it is not clear whether self-recording was chosen because the situations lent themselves to this technique, or simply whether it seemed a cheap option.

Kinley et al (2001) examined the use of appropriately trained nurses (ATN) and pre-registration house officers (PRHO) in pre-operative assessment before elective general surgery. They collected data alongside a multicentre randomised (equivalence) trial including time spent on assessments by the

PRHO, ATN and anaesthetist. The method of time measurement was not stated although it appeared to be through self-recording on the assessment forms. It seems likely there is a publishing error as the data presented suggested that anaesthetists and ATNs took exactly the same time (mean and SD identical). Good aspects about the study were its reporting of times as means and standard deviations, deterministic one-way sensitivity analyses and probabilistic modelling using parameter distributions derived from the trial data. However, it was unclear whether the selected ranges used in this Monte Carlo simulation were the limits of the actual data or the choice of the researchers.

Morrell et al (2000) undertook a randomised controlled trial comparing a non-intervention control group and community support workers (SW) providing a maximum of ten three-hour weekday visits in the first 28 days after delivery. The SW assistance included talking to the mother, helping with the baby or siblings, other practical help (e.g. housework).

There were several key issues about the paper. It did not describe where or how the SWs recorded the start and finish times and there was no mention of attempts to check data validity. The data available demonstrated a wide range in visit duration from 10 to 375 minutes (mean 143, SD 37 minutes), with short times for initial visits or when there was no access. Resource use data on the frequency of NHS and other contacts came from self-completed questionnaires administered to the mothers. The researchers cross-validated the GP contact data by examination of a sample of GP notes. However, it is unclear whether the SWs intervention influenced the duration of such visits; the researchers gave no rationale for not measuring visit duration with the exception of midwifery visits (described below).

Community midwives completed a separate activity log survey to derive the duration of antenatal and postnatal contacts and travel. The researchers estimated costs per visits by combining the time data with the midwife's actual salary and randomly assigning a cost from this distribution to each mother's midwifery contacts. Thus, the data did not relate specifically to the mothers in the study.

It would be difficult to assess the full impact of SWs a priori, as it was a new intervention. The actual intervention was the principal cost driver for the SW group's significantly higher incremental cost at both 6-weeks and 6-months. Sensitivity analysis found that even reducing SW visits to a 2-hour maximum and increasing SW visits to three per day (instead of 2.5) only produced a relatively small reduction in the mean incremental cost. No disaggregated costs were significantly different; in this instance, it may be that other contacts durations are unimportant, but the study did not investigate this in the sensitivity analysis.

The paper by Townsend et al (2004) reported a study that compared routine examination of newborn babies by midwives (an extended role) and paediatric senior house officers (SHOs). The researchers stated they estimated the duration of assessment from the examination sheet that included entry of start and end times. However, the sheet in the appendix only included the 'time now' and no end time.

Furthermore, the paper presented staff time results as medians not means. In addition, it would have been possible to validate partially the time data from 39-videotaped examinations recorded for quality checking the assessments.

Caine et al (2002) stated that patients attending a bronchiectasis clinic recorded the duration of consultations in a randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care. However, there are no details of how or where patients documented this, none of the patient-related paperwork shows start and end times and there is no mention of any attempt to validate the data. Similarly, the data presented do not indicate the data variability (SD, IQR etc).

Boland et al (2003) conducted a randomised controlled trial of clinical and cost-effectiveness of blind and image-guided insertions of Hickman lines in adult cancer patients by nurses. Key outcomes included the procedure time, time in the interventional X-ray suite to reposition misplaced catheters, and waiting time between initial insertion and repositioning.

The authors argued that "given the confidential nature of the service provided by nurses" (p22), it was not possible for researchers to observe and time nurses' activities. Whilst this may reflect a local issue, it is not true in general. Recording of times was after the event. For the Hickman line procedures, the nurses' start and finish times were items buried on the fourth (last) page of the case report forms. Retrospective time recording is unlikely to be very accurate. Indeed, the minimum-maximum times presented suggest rounding by staff to the nearest 5 minutes and all the times had an air of precision. For example, the mean image-guided procedure time was 40 minutes (95% CI 38 to 42 minutes) with minimum-maximum times 20-150 minutes respectively.

The authors also stated that it was difficult for nurses to record individual patient times as nurses "tended to treat more than one patient at a time" (p22-23). Instead, they asked nurses to estimate the amount of time they spent with each patient. Based on the researchers' rationale, this decision seems illogical and appears to have little validity. Additionally, the method used to determine waiting times between basic insertion and repositioning is unclear. It is difficult to assess the impact of staff time, as sensitivity analyses did not appear to cover these data. Even given the data limitations, it would have been interesting to see the effect of using these 'rough' estimates of staff time in a probabilistic sensitivity analysis.

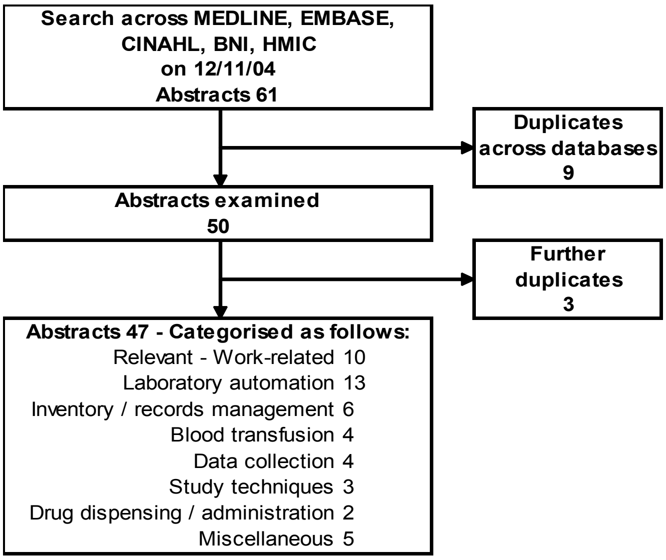
Despite the availability of times for X-ray suite activities from a logbook, these data were unused in the economic evaluation. The researchers justified this on the basis that the suite had to be block-booked for one hour. They argued that staff could not predict in advance for which patients this time would be an over or under-estimate and that it would average out. Yet this issue warranted investigation as sensitivity analyses showed the cost of interventional X-ray might alter the preferred insertion method.

Overall, the impact of 'inaccurate' time estimates is difficult to assess outside sensitivity analyses for individual evaluations. However, the issue of staff time does not seem to receive sufficient coverage to be confident that decision makers would always arrive at the same conclusion. The next section examines bar code scanners as one possible method of measuring staff time in health care.

6. Literature review of the use of bar code scanners to measure staff time in health care

A brief literature search across MEDLINE, EMBASE, CINAHL, The British Nursing Index (BNI) and Health Management Information Consortium (HMIC) for bar code papers related to time/work measurement produced the results shown in Figure 2[d].

Figure 2: Results of literature search



Excluded papers

There were 37 papers that covered issues not directly related to staff time measurement. The following explains the categories used in Figure 2. Laboratory automation covered laboratory procedures and techniques where bar codes aided sample identification or stored the actual test data. Those on blood transfusion and drug dispensing/administration related how bar codes could aid product administration and minimise patient identification errors. Data collection included papers that used bar code scanners to collect data from a variety of sources, including for hospital management (clinic waiting times), pharmacy information and nutritional data. Study techniques described the use of bar codes to facilitate study (e.g. to access self-teaching packages).

Included papers

There were 10 time/work related papers. The following summary used abstracts whilst awaiting the

[d] Simultaneous search across OVID MEDLINE (1966 – November Week 1 2004), EMBASE (1980 to 2004 Week 45), CINAHL (1982 – Nov Week 1 2004), The British Nursing Index (BNI, 1985 – October 2004) and Health Management Information Consortium (HMIC, September 2004) on 12/11/04. Search terms used: bar?code.mp AND (time or staff or skill?mix or case?mix or work measure$ or workload or resource management or manpower plan$ or grade mix or staff$ level$).mp

arrival of the papers. Currently, this means that information is very limited and only allows an overall impression.

Many papers only described the use of bar code scanners to record activities (especially for nurses) and either presented minimal or no results. This applied to six papers. Two of these covered a study by Holmes et al (1997)[A,B] that compared the amount of care provided in special dementia and traditional nursing home units in New York State. A bar code system recorded the time and care input. The only results available in the abstracts were the finding of differences for three of the ten staff groups studied. The four papers by Ammentorp et al (1988), Percival and Westcott (1990), Allen (1992), and Macfarlane and Lees (1997) simply appeared to describe the use of bar code scanners.

An abstract by Blount (1999) showed the most promise. It is for a PhD thesis (in the US) that examined the correlation between actual and derived standard nursing care hours. The nurses used hand-held bar code scanners to record the actual time spent with individuals and activity involved. The standard time was the required number of hours designated to six Medicus Patient Classification System (PCS) categories. This apportioned hours between shifts and types of nursing staff (i.e. registered, licensed and attendant). Data collection was from 33 hospital units at a large medical centre. The study found a high correlation between actual and standard hours (0.742 $p<0.001$). However, assessment of agreement is more appropriate than correlation since the latter says nothing about whether derived time accurately reflects actual time. There are further concerns about the study methods. For example, the abstract states that the study used category standard hours rather than raw PCS scores. Furthermore, the nurses' did not categorise the patients on the care received but prospectively, based on the care they would need. The impact of these limitations is difficult to assess without knowing more about the health system.

The study by Stewart and Short (1999) is also promising. It evaluated hand-written and bar coded logging of events observed on video tapes of simulated resuscitations. Compared with the gold standard actual videotape time, the bar code method was more accurate than hand-written logging as assessed by the mean absolute errors and their standard deviations ($p<0.01$). Omission of events was not significantly different.

The remaining two abstracts present studies that may be useful background information. One is a PhD thesis (Martin, 1990) that aimed to design and test the validity of a patient classification system based on resource consumption for home care. Fifteen nurses used bar code scanners to collect time data in a convenience sample of 404 home visits over three weeks. Regression analysis found that the classification system developed by clinical staff, and clinical and demographic factors accounted for 42% of the variance in direct care. Discriminant analysis found the model and other variables could correctly separate 71% of visits into different length groups. However, the authors concluded that resource use was highly variable and that the model needed further refinement.

The other abstract describes work by Rotondi et al (1997) on the use of a real-time patient tracking system using bar codes to assess variations and bottlenecks in an operating theatre in the US. Multidisciplinary teams worked to improve the efficiency of theatre utilisation from less than 50% based on information collected for 17 events during patients' surgical stay.

Unpublished/local experience of bar coding projects

In addition, to the brief literature review above, some information is available from five bar coding projects that have been co-ordinated successfully by Southampton University Hospitals NHS Trust. Predominantly these examined working practices. Two of the studies resulted in journal publications. Macfarlane and Lees (1997) (as mentioned above), described two projects and the frequencies of activities. The second paper, by Walsh et al (2003), presented staff time, but not costs. The amount of time data recorded was poor and the researchers considered it only sufficient "to provide little more than contextual information" (p316). From discussion with the main researcher, the poor data quality was particularly due to difficulties in one of the comparison settings for a variety of 'political' and pragmatic reasons. However, the person who oversaw all the bar code scanner projects did not find such problems apparent in the other projects. The other projects were unpublished or not designed to cost staff time. In addition, there is a question about the method of analysing time data from repeated measurements on the same individuals - an issue revisited in the next section.

7. Proposed study using bar code scanners to measure nursing time

The primary aim of the proposed observational study is to assess whether measurement of nurse-patient time using bar code scanners improves the costing of nursing staff time in economic evaluations. To answer this, the study will address a number of subsidiary questions.

1. What difference, if any, does this method of allocating staff time make to the estimates and precision of nursing costs compared with a priori assumptions, e.g. that all patients take the same time?
2. What is the value of such information for economic evaluations? In this context:
    - What is the research effort required to collect such data?
    - Is it feasible to use bar coding to collect staff time/cost data?
    - Is the data collection method acceptable to staff?

Initial piloting will test the establishment and methods of data collection, building up from a few hours to a week's data collection. The plan is to collect data at more than one location to gain an understanding of implementing the technique in different settings. These are likely to be a ward at Southampton University Hospitals NHS Trust and a chronic haemodialysis unit. Patient group selection, by negotiation with senior nursing staff, will reflect situations where staffing time is important. For example, at the haemodialysis unit the senior nurse could identify patients potentially eligible or ineligible for nurse-led satellite care. In ward-based data collection, the patients and staff of interest will need to be at a small number of locations.

The data collection has the following components:

1. Nurse time: Each nurse and patient will have a unique bar code identifier. The nurses will use bar code scanners to record the start times of patient contacts (direct and indirect) or non-patient specific activity. Alternatively, it may be necessary to record the start and end times of patient-specific tasks and allocate the remaining non-specific time for each nurse's shift across the patients. These methods with be investigated at the piloting stage.

   For each patient, nurse time will be aggregated (in total and by grade) across the following time categories i) direct, ii) indirect, iii) non-specific, and iv) total patient time (the sum of i to iii). The main data collection will be in one-week blocks sampled every 3 months over a year in order to investigate variation in patient mix and staffing levels. Nurses who do not record their time will have their shift hours attributed across the patients they cared for. This will reduce the effect of the time measurement, bringing the mean value closer to the 'usual' assumption that all patients take the same amount of nursing time. However, it ensures that all the ward/units nursing costs are attributable to patients. TN will conduct 'spot checks' to validate the recordings of face-to-face patient-staff contacts.

2. Patient dependency data: The intention is to obtain this data by asking the nurses or a nurse manager to rate each patient's dependency once a day using their local rating systems or Karnofsky Performance Scale (Karnofsky and Burcherval, 1949), a scoring system that measures functional activity in renal patients.

3. Data to derive the cost of nursing time: Data required for this includes collection of information on staff whole-time-equivalents (in post and vacancies), grade, pay scales, unsocial hours arrangements, financial accounts data on staff expenditure (including agency staff), and capacity utilisation in terms of actual and potential patients. The plan is to combine each individual's total patient time with the nursing salary costs to calculate the nursing cost per patient using accepted costing techniques.

4. Predominantly assessment will be descriptive for the feasibility of using bar code scanners to measure nursing time. The aim is to describe the challenges to undertaking such research and quantify as far as possible the research effort and costs required to collect such data. However, the assessment will also examine the extent of 'missing' data where the bar code scanners were not used and the validation data from spot checks.

5. Acceptability of the data collection method to staff will involve administration of a questionnaire used in a previous study. From informal talks with staff, it should be possible to obtain additional feedback about the feasibility of bar coding as a general data collection tool.

Data analysis

Key quantitative analyses are comparisons of a) the nursing cost per patient derived with and without the adjustment for nurse-patient time; and b) nurse-patient time across the four data collection periods and overall to determine the robustness of the estimates. Secondary analyses include the comparison of patient dependency and total patient time and the impact on costs of capacity utilisation in terms of actual and potential patients. The data are complex due to repeated measurements on the same patient and by multiple nurses (i.e. clustered data). A discussion of whether multilevel modelling is necessary in these circumstances would be helpful.

Expected benefits

The expected benefits are that bar code measurement of staff time will produce stochastic estimates for use in probabilistic sensitivity analyses, rather than reliance on best guess estimates of upper and lower limits. However, since this project will also assess the overall research costs involved, potentially it will inform value of information analyses for future economic evaluations that might consider using bar code time measurement. The following section describes some of the feasibility issues to date.

8. Feasibility issues of setting up a bar coding project – experience to date

This section details some of the lessons learned from setting up the bar code scanners. This took an estimated 12 working days, although this is a relatively crude estimate as much of the work preceded the actual PhD start date. The overall conclusion, so far, is to allow additional time for what, on face value, should be a simple task. The last part of the section details the associated costs for the equipment.

Sourcing bar code scanners

The scanners for this research (short-term at any one location) required the following specifications: easy to use (to facilitate use and minimise staff training needs); small and lightweight (to fit easily into a pocket); portable and cheap (maximum £100 per scanner).

There were many challenges to obtaining the scanners. It was a steep learning curve to understand technical aspects to facilitate discussions with suppliers. Few manufacturers produce scanners to the required unsophisticated specifications and most suppliers seemed disinterested in a 'small' and/or Public Sector order. This hampered both accessing information and sourcing the scanners. Frequent technology changes meant that acceptable models rapidly became obsolete with insufficient stock levels to fill the order. The newer 'cheaper' consumer style scanners had more restrictive facilities than previous models, for example, with potentially inadequate storage capacity for time research purposes and lack of programmability. The model purchased was the Symbol CS 1504 Consumer Memory Scanner (see Figure 3). However, long-term investment for time-related research seems unlikely with a short warranty (90-days) and time-limited product support.

Figure 3: CS 1504 scanner

Setting up scanners for downloading scanned data to computer

Setting up the scanners and establishing procedures to download data is not easy for those without excellent IT skills. Consultancy from the scanner supplier (1/2 day) proved worthwhile expenditure. For example, communication between the scanner and computer is via a special USB/9-pin serial connection cable and required searching for additional drivers on the scanner manufacturer's website.

Software is required to set the scanner time/date, save files of bar code data and export data. However, such software is not included automatically in the scanner purchase and required separate negotiation with a supplier. Frustratingly, it transpired that the Sales Manager was not very knowledgeable about the functioning of 'cheaper' scanners. With hindsight, confirming the expectations of the order in writing would have avoided some difficulties. The matter did not resolve swiftly even on following advice from the supplier's consultant about third party software (MiniPro Version 1.0 2004, CS1504 Bar Code Extraction software from MMR Software, http://www.cs1504.com). The supplier's consultants could have written specific software, but this would have been pricey and exceeded the budget. Instead, after many hours of trial and error attempts, the codes exported as text into Excel and this got around the cause of the problem – US date formats.

Using various functions in Excel, it is possible to validate the bar codes, separate them into their component data (time/date, centre, staff ID, task, and patient ID), and calculate the task duration. Furthermore, it is possible to download the scanner's device identifier as a data check by cutting and pasting and subsequently manipulating data from the MiniPro download screen. Establishing this process was not straightforward due to the truncated and compressed way the scanners stored their serial numbers.

Labelling software

Production of printed bar codes (referred to as labels) required special software bought from the scanner supplier (Loftware Label Manager Version 7.2.2.11), along with a further 1/2 day consultancy time for installation and training. This was helpful as it included advice tailored to the project's data collection. However, subsequent installation of labelling software on a laptop computer was unsuccessful. There were several weeks' delay whilst awaiting a reply from the consultant and eventually it took technical advice direct from the software manufacturer to resolve the problem. This necessitated swapping the serial port key for USB licence key.

Safety issues

The rather cryptic product information meant that sorting out the safety issues with the University's Laser Safety Adviser took additional time, although necessary to comply with regulations for PhD projects. Hopefully clarifying the position 'now' will also smooth the passage through the Trust's research governance procedures. However, the issue of possible, though unlikely, interference to radio communications still needs investigation.

Table 1: Costs of bar code scanners

Total cost (£)	Comment	Supplier
4,300.50	Symbol CS1504 USB Memory Scanner Kit (50 x £86.01)*	Zetes Ltd
412.83	Loftware Design and Print Module (V7.2) - Labelling software*	Zetes Ltd
125.46	Loftware Maintenance 12 Month Support*	Zetes Ltd
750.00	Consultancy (1 day): On-site installation and setup of Symbol CS1504 devices and software and creating suitable bar code*	Zetes Ltd
54.02	MiniPro (bar code extraction) software*	MMR Software
31.80	36" beaded link nickel chains – for scanners (for 50)	Niceday
£ 5,674.61	**Total (* no VAT)**	

9. Agenda for discussion

This paper presents work in progress that forms the start of TN's PhD. Comment and advice are welcome from health economist colleagues. Particular issues are:

- Is the proposed research on the right tack?
- Is multilevel modelling necessary to analyse the time data?
- Should the unit cost applied to the staff time reflect the actual or theoretical grade mix (i.e. used or average across all the relevant staff grades)?
- What methods have others used to cost staff time? How successful have these methods been in terms of data quality and the feasibility of data collection?
- Are there ways to identify a priori those economic evaluations for which measurement of staff time is necessary? Would crude modelling to identify important cost drivers before designing data collection work? Is it the overall budget impact or statistical significance that decides whether a particular cost driver is important? How would one accommodate pre-study modelling – at the grant application stage, or before starting the main study (and cover its costs)?

References

Allen T. (1992) Hospital bar code breakthrough. Health Direct (25):10

Ammentorp W, Bock W, Irrgang S et al. (1988) Assuring quality care: measuring and monitoring the caring process. Journal of Quality Assurance 10(2):10-2

Blount MA. (1999) Correlation of actual nursing care hours with patient classification system standard hours. Virginia Commonwealth University

Boland A, Haycox A, Bagust A et al. (2003) A randomised controlled trial to evaluate the clinical and cost-effectiveness of Hickman line insertions in adult cancer patients by nurses. Health Technology Assessment 7(36)

Brouwer W, Rutten F, Koopmanschap M. (2001) Costing in economic evaluations. In: Drummond M, McGuire A, editors. Economic evaluation in health care. Merging theory with practice. Oxford: Oxford University Press Chapter 4:68-93.

Caine N, Sharples LD, Hollingworth W et al. (2002) A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic. Health Technology Assessment 6(27)

de Keizer NF, Bonsel GJ, Al MJ. (1998) The relation between TISS and real paediatric ICU costs: a case study with generalizable methodology. Intensive Care Medicine 24(10):1062-1069

Holmes D, Teresi J, Lindeman DA et al. (1997)[A] Measurement of personal care inputs in chronic care settings. Journal of Mental Health and Aging 3(1):119-27

Holmes D, Teresi J, Ramirez M et al. (1997)[B] The measurement and comparison of staff service inputs in special dementia care units and in traditional nursing home units using a bar code methodology. Journal of Mental Health and Aging 3(2):195-207

Karnofsky D and Burcherval J. (1949) The Clinical Evaluation of Chemotherapeutic Agents in Cancer. New York: Columbia University Press

Kinley H, Crozoski-Murray C, George S et al. (2001) Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and pre-registration house officers in pre-operative assessment in elective general surgery. Health Technology Assessment 5(20)

Macfarlane L and Lees P. (1997) Hospital activity. Barred facts. Health Service Journal 107:26-9

Martin DL. (1990) Classifying home health patients based on nursing resource consumption: model development and validation. Rush University, College Of Nursing

Percival M and Westcott J. (1990) Data in the districts. Health Service Journal 100 (5188): 255-56

Macfarlane L and Lees P. (1997) Hospital activity. Barred facts. HSJ 107:26-29

Meyers FE and Stewart JR. (2002) Motion and time study for lean manufacturing. 3rd ed. Prentice Hall: Upper Saddle River, New Jersey

Morrell CJ, Spiby H, Stewart P et al. (2000) Costs and benefits of community postnatal support workers: a randomised controlled trial. Health Technology Assessment 4(6).

Roderick P, Nicholson T, Gerard K et al. (in press). An evaluation of the costs & effectiveness of & quality of care of renal replacement therapy provision in renal satellite units in England & Wales. Health Technology Assessment

Rotondi AJ, Brindis C, Cantees KK et al. (1997) Benchmarking the perioperative process. I. Patient routing systems: a method for continual improvement of patient flow and resource utilization. Journal of Clinical Anesthesia 9(2):159-69

Stewart J A and Short FA. (1999) Time accuracy of a bar code system for recording resuscitation events: laboratory trials. Resuscitation 42(3):235-40

Townsend J, Wolke D, Hayes J et al. (2004) Routine examination of the newborn: the EMREN study. Evaluation of an extension of the midwife role including a randomised controlled trial of appropriately trained midwives and paediatric senior house officers. Health Technology Assessment 8(14)

Macfarlane L and Lees P. (1997) Hospital activity. Barred facts. HSJ 107:26-29

Walsh B, Steiner A, et al. (2003) Nurse-led inpatient care: Opening the 'black box'. Int J Nurs Stud40:307-319

Appendix 1: Summary of the 45 UK economic evaluation abstracts (2003-2004) from NHS EED

Technology	N	%
Unknown	1	2
Device	1	2
Diagnostic	0	0
Dressing	1	2
Drug	19	42
Drug - delivery modes	2	4
Immunisation	2	4
Other	2	4
Psychological	5	11
Screening	0	0
Screening - Diagnosis	1	2
Service delivery	6	13
Service delivery - Device	1	2
Surgery/procedure	4	9

Setting	N
Unknown	2
Community (probably)	3 (2)
Primary & Secondary Care	2
Primary Care (probably)	8 (1)
Secondary Care (probably)	21 (4)
Tertiary Care (probably)	1 (1)

Sensitivity analysis	N
No	11
Results but no info or scenarios only	2
Yes	32

Currency	N
$ US	3
£ UK	39
£ UK & $ (US ?)	1
£ UK/$ US	1
Euros	1

	Multi-centre assessment			**Based on UK – assessment of resource use at a centre**			**Adequate reporting**	
	Effectiveness	**Resource use**	**Primarily modelling**	**Staff at one/ limited locations**	**Patients at one/ limited locations**	**Likely staffing costs important**	**Resource use**	**Unit costs**
Unknown	3	4	2	8	16	13	1	1
No (probably*)	10 (1)	11 (2)	32 (1)	3 (0)	6 (2)	11 (2)	17 (10)	13 (9)
Yes (probably*)	22	16 (1)	10	32 (2)	18 (2)	16 (3)	13 (4)	17 (5)
Probably (mixed sources)	9	10						
Not applicable					1			

* or partial for unit costs