UNIVERSITY OF SOUTHAMPTON
FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES
School of Medicine

Appropriate estimation of staff costs for economic evaluations: A case study in haemodialysis

by
Ann Patricia Nicholson

Thesis for the degree of Doctor of Philosophy
July 2008

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

SCHOOL OF MEDICINE

Doctor of Philosophy

APPROPRIATE ESTIMATION OF STAFF COSTS FOR ECONOMIC EVALUATIONS: A CASE STUDY IN HAEMODIALYSIS

by Ann Patricia Nicholson

The thesis examines methods to measure or attribute resource use and costing for economic evaluations in health care. The literature review found minimal evidence comparing top-down and bottom-up (micro) costing. To cost nursing inputs for patients, studies rarely measured staff time and poorly reported methods. In chronic haemodialysis (HD), 'case mix' variations in nursing between patients were ignored.

The empirical work evaluated nurses' self-recording using barcode scanners and observer work sampling to measure the nursing time per patient. Initial piloting eliminated patient-level work sampling due to problems linking data to patients. Barcode scanning captured 80% of nurses' hours; data quality was acceptable. It covered 4 weeks for 169 patients. Costs, in 2006, included employers' National Insurance and superannuation.

Relative to the 'top-down' nursing expenditure per HD session (£44.56 to £50.79), the bottom-up cost was underestimated by up to 10%: 4% due to the unit cost using expected rather than actual working hours, and 6% due to missing patient-level resource use data. Multiple linear regression clustered by patient found those ineligible for care at satellite units needed extra nursing input (mean 8 minutes, 95% CI 4-11, or £2.30 to £7.22 per session) compared with those eligible.

Conclusions were that top-down (expenditure based) and bottom-up estimates of staff costs cannot reconcile due to averaging at different points, their attribution of resource use or costs to patients, and valuation of unit costs. More guidance is required on which unit cost of staff time (per hour paid, worked or patient-related) best reflects the opportunity cost of staff time. Barcode scanning successfully captured data, but required considerable research effort, making it impractical for most multicentre studies. Cost differences between patients were 5-14% of the nursing cost per session or 1-5% of the overall cost per session. Hence, they had minimal effect on results of economic evaluations.

Table of Contents

Abstract i

Table of Contents ii

List of Tables vii

List of Figures x

List of Figures x

List of Appendices xii

Declaration of authorship xiii

Acknowledgements xiv

Abbreviations xv

Chapter 1 Introduction **1-1**

1.1 Application of economics in costing health care 1-1

1.2 Background to costing health care 1-4

1.3 The importance of staff costs 1-7

1.4 Measuring staff inputs 1-9

1.5 Aims and objectives of the thesis and empirical work 1-11

1.6 Outline and scope of the thesis 1-11

Chapter 2 Costing health care **2-1**

2.1 Overview of costing 2-1

2.1.1 The production and cost functions 2-1

2.1.2 Key costing terms 2-3

2.1.3 Scope of costing exercise 2-4

2.1.4 Feasibility issues and targeting research effort 2-5

2.1.5 Resource use (inputs) 2-7

2.1.6 Unit costs 2-8

2.1.6.1 Routine costs available in the NHS 2-11

2.2 Literature review of costing guidance and research 2-12

2.2.1 Overview of literature on costing in economic evaluations 2-12

2.2.2 Top-down and bottom-up costing - guidance and research 2-18

2.2.2.1 Comparison of top-down and bottom-up costing approaches 2-18

2.2.2.2 Recommendations on which costing approach to adopt 2-21

2.2.2.3 Review of empirical comparisons of costing approaches 2-23

2.2.3 Classification of patient heterogeneity 2-27

2.2.3.1 'Case mix' 2-27

2.2.3.2 Diagnosis Related Groups (DRGs) 2-28

2.2.3.3 Healthcare Resource Groups (HRGs) and reference costs 2-30
2.2.3.4 Patient heterogeneity in nursing inputs - nursing 'workload' 2-32
2.3 Conclusions 2-36
Chapter 3 Measuring staff time 3-1
3.1 Types of staff time - categorisation of activity 3-1
3.2 Methods to measure staff time 3-3
3.2.1 Time and motion study (stopwatch study) 3-3
3.2.2 Work sampling 3-3
3.2.3 Self-recording 3-5
3.2.4 Opinion-based methods - self-reporting and expert opinion 3-6
3.2.5 Technology to help data collection - barcode scanning 3-7
3.3 Considerations about choice of time measurement technique 3-8
3.4 Time measurement literature 3-13
3.4.1 Measurement of staff time in costing and economic evaluations 3-13
3.4.1.1 Staff costs in the "Unit costs of health and social care" 3-13
3.4.1.2 Use of time measurement techniques in economic evaluations 3-21
3.4.2 Time measurement studies in health care 3-24
3.4.2.1 Barcoding studies at Southampton 3-24
3.4.2.2 Studies of patient-level time measurement in health care 3-28
3.4.2.3 Comparative time measurement studies 3-30
3.5 Conclusions 3-31
Chapter 4 Renal failure and renal services 4-1
4.1 Epidemiology of established renal failure 4-1
4.2 Treatment options for ERF 4-3
4.3 RRT service provision 4-4
4.4 Challenges for haemodialysis provision - policy initiatives 4-6
4.4.1 Policy initiatives 4-6
4.4.2 Payment for haemodialysis 4-7
4.5 Economic evaluations of RRT 4-10
4.5.1 Overview of economic evaluations of RRT 4-10
4.5.2 Renal satellite evaluation (RSU) study and background to thesis 4-11
4.5.3 Handling of patient heterogeneity in renal economic evaluations 4-16
4.5.3.1 NICE appraisal of home and hospital haemodialysis 4-16
4.5.3.2 Economic evaluations of haemodialysis in different settings 4-17
4.6 Conclusions 4-20
Chapter 5 Introduction to empirical work 5-1
5.1 Haemodialysis as a case study 5-1

5.2 Measurement of nursing time per patient 5-1
5.3 Classification of patient heterogeneity 5-2
5.4 Research questions addressed by empirical work 5-3
5.5 Rationale for data collection sites and piloting 5-4
5.6 Summary 5-6
Chapter 6 Methods for piloting 6-1
6.1 Measurement of nursing time inputs for haemodialysis 6-1
6.1.1 Barcode scanning 6-1
6.1.1.1 Recruitment and consent to barcode scanning 6-3
6.1.1.2 Data management and statistical analyses 6-4
6.1.1.3 Data validity 6-4
6.1.1.4 Feasibility of barcode scanning 6-5
6.1.2 Work sampling 6-5
6.2 Classification of patient heterogeneity 6-7
6.2.1 Patient dependency-scoring tool for outpatient HD 6-7
6.2.2 Karnofsky Performance Scale (KPS) 6-8
6.2.3 Co-morbidity indices 6-9
6.3 Summary 6-10
Chapter 7 Results of piloting (SUHT and Totton) 7-1
7.1 Establishing barcode scanning data collection 7-1
7.1.1 Overview of barcode scanning and data validity 7-2
7.1.2 Feasibility of barcode scanning 7-4
7.1.2.1 Acceptability of data collection methods to nurses 7-6
7.2 Assessment of work sampling 7-7
7.3 Testing methods to categorise patient heterogeneity (Totton) 7-8
7.3.1 Patient dependency-scoring tool for outpatient HD 7-8
7.3.2 Karnofsky Performance Scale (KPS) 7-10
7.3.3 Co-morbidity indices 7-11
7.4 Conclusions from piloting 7-11
Chapter 8 Methods for main data collection 8-1
8.1 The Portsmouth MRU 8-1
8.2 Barcode scanning 8-1
8.3 Classification of patient heterogeneity (KPS and dependency-scoring) 8-2
8.4 Data management and statistical analyses 8-3
8.4.1 Dataset structure 8-3
8.4.1.1 Patient attendances 8-3
8.4.1.2 Barcode scanning - raw nursing time data 8-4

8.4.1.3 Barcode scanning - aggregated nursing time data 8-6
8.4.1.4 Overall data considerations 8-6
8.4.2 Statistical analyses 8-7
8.4.3 Economic consequences 8-10
8.5 Summary 8-11
Chapter 9 Results (Portsmouth) 9-1
9.1 Overview of data collection and patient characteristics 9-1
9.2 Barcode scanning data - validity issues 9-3
9.2.1 Recruitment of nurses and completeness of data 9-3
9.2.2 Examination of possible outliers 9-5
9.2.2.1 Single barcode scans 9-5
9.2.2.2 Aggregated time per HD session 9-7
9.2.3 Nurses' feedback about their individual data 9-9
9.2.4 'Spot check' observations 9-11
9.3 Classification of patient heterogeneity 9-12
9.3.1 Heterogeneity of patients on HD (KPS and dependency) 9-12
9.3.2 Validity of the patient dependency-scoring tool 9-15
9.3.2.1 Construct or empirical validity of dependency-scoring tool 9-16
9.3.2.2 Audit of dependency data quality and inter-rater issues 9-17
9.4 Statistical analyses of nursing time 9-18
9.4.1 Analyses using summary measures 9-18
9.4.1.1 Nursing time and eligibility for RSU care 9-18
9.4.1.2 Nursing time and patient dependency 9-19
9.4.2 Multiple linear regression analyses 9-20
9.4.3 General Estimating Equations (GEE) analyses 9-24
9.4.4 Economic consequences 9-26
9.4.4.1 Resource use implications of differences in nursing time 9-26
9.4.4.2 Nursing costs 9-29
9.5 Conclusions 9-36
Chapter 10 Discussion and conclusions 10-1
10.1 Costing methods 10-1
10.2 Measurement of staff time 10-7
10.2.1 Barcode scanning 10-8
10.2.2 Work sampling 10-11
10.2.3 Time measurement in general 10-12
10.3 Patient heterogeneity in chronic haemodialysis 10-17
10.3.1 Eligibility for RSU care 10-18

10.3.1.1 Implications of findings for costing outputs and reimbursement 10-19
10.3.2 Patient dependency 10-20
10.4 Strengths and limitations of the empirical study 10-22
10.5 Implications for researchers and future research 10-24
10.5.1 Suggestions for costing in economic evaluations 10-24
10.5.2 Recommendations for future research 10-28
10.6 Conclusions 10-30
10.6.1 Costing methods 10-30
10.6.2 Measurement of staff inputs 10-31
10.6.3 Costing and patient heterogeneity in HD 10-33
10.7 Outputs from PhD 10-34
Appendices **11-1**
Glossary **11-48**
References **11-49**

List of Tables

Table 1.1 Conditions necessary for a perfect market - applied to health care 1-2
Table 1.2 Measurement of or valuation of outcomes in economic evaluations 1-4
Table 2.1 Outline of chapter sections addressing key aspects in costing 2-1
Table 2.2 Definition of the production and cost functions 2-2
Table 2.3 Definitions of key costing terms 2-3
Table 2.4 Data collection methods for resource use and main sources of bias 2-8
Table 2.5 Direct or indirect influences on unit costs 2-9
Table 2.6 Examples of sources of unit costs and methods of derivation 2-9
Table 2.7 Cost classifications in the NHS costing manual 2-12
Table 2.8 Review of costing guidance for economic evaluations 2-14
Table 2.9 Implications of poor methods and reporting 2-16
Table 2.10 Comparison of top-down and bottom-up costing approaches 2-19
Table 2.11 Recommendations on choice of top-down or bottom-up costing 2-22
Table 2.12 Comparison of top-down and bottom-up estimates by cost category at two centres (Scotland and Estonia) 2-24
Table 2.13 Examples of studies comparing bottom-up costs and tariffs 2-26
Table 2.14 'Case mix' from different perspectives 2-27
Table 2.15 Summary of commonly used workforce planning methods (adapted from Hurst 2002) 2-34
Table 3.1 Example of work sampling to estimate patient times 3-4
Table 3.2 Observations needed for 2% expected time with 95% confidence 3-5
Table 3.3 Comparison of methods to measure staff time 3-9
Table 3.4 Validity of time measurement techniques 3-12
Table 3.5 Potential factors influencing variation in staff time and costs 3-13
Table 3.6 Proportions of working time for hospital nurses by activity 3-15
Table 3.7 Options in calculating the unit cost of nursing time 3-16
Table 3.8 Sources of information on time (use and unit costs) for health/social care staff in "Unit costs of health and social care" 3-19
Table 3.9 Economic evaluations of role substitution in Health Technology Assessment 3-23
Table 4.1 Haemodialysis reference costs and indicative tariff (England, 2006) 4-8
Table 4.2 Medicare reimbursement for haemodialysis 4-10
Table 4.3 Options for handling patient variation in nursing inputs in RSU study 4-13
Table 4.4 Affect of patient heterogeneity in haemodialysis 4-14

Table 4.5 Economic evaluations of haemodialysis across settings 4-19
Table 5.1 Factors influencing choice of data collection sites 5-5
Table 6.1 Work sampling schedule 6-5
Table 6.2 Interpretation of dependency scores 6-8
Table 7.1 Patients' characteristics (SUHT and Toton) 7-1
Table 7.2 Barcode scanning data collection (SUHT and Toton) 7-2
Table 7.3 Recruitment of nurses by grade (SUHT and Toton) 7-2
Table 8.1 Population of patients at the MRU 8-4
Table 8.2 Example of a patient's nursing time and dependency data aggregated by HD session 8-6
Table 8.3 Data considerations for statistical analyses 8-7
Table 8.4 Additional sources of information to estimate costs 8-11
Table 9.1 Barcode scanning data collection 9-1
Table 9.2 Eligibility for RSU care (N=169) 9-2
Table 9.3 Grade mix of nurses 9-4
Table 9.4 Time by type of nurse (HCAs, RGNs and overall) 9-5
Table 9.5 Descriptive statistics for single direct and indirect care scans 9-6
Table 9.6 Descriptive statistics - aggregated times per HD session 9-8
Table 9.7 Missing nursing data at the start or end of patients' HD sessions 9-9
Table 9.8 Nursing time at the start and end of each patient's HD sessions 9-9
Table 9.9 Estimates of indirect care times per patient (from 10 nurses) 9-10
Table 9.10 Possible scanning mistakes identified from nurses' feedback 9-10
Table 9.11 Agreement between nurse-recorded and observer-recorded activity 9-12
Table 9.12 Descriptive statistics for patient summary KPS scores 9-13
Table 9.13 Descriptive statistics for patient dependency scores 9-15
Table 9.14 Descriptive statistics for summary (mean) patient-specific time 9-19
Table 9.15 Estimated coefficients from preliminary regression models 9-22
Table 9.16 Estimated coefficients from full regression model 9-23
Table 9.17 Patient-specific nursing time per HD session 9-24
Table 9.18 Estimated coefficients from preliminary GEE models 9-25
Table 9.19 Estimated coefficients from full GEE model 9-25
Table 9.20 Comparison of patient-specific time per session from nurse to patient ratios and multiple regression 9-26
Table 9.21 Nurses' hours by activity 9-28
Table 9.22 Summary - illustrating estimation of average unit costs per hour of nursing using bottom-up and top-down approaches 9-30
Table 9.23 Average nursing time per HD session 9-31

Table 9.24 Steps to estimating average cost per HD session using bottom-up and top-down approaches 9-32
Table 9.25 Steps to estimating mean extra nursing cost per patient ineligible for RSU care (using bottom-up approach) 9-34
Table 9.26 Summary of key differences between bottom-up and top-down costing in study 9-35
Table 10.1 Implications of methods to attribute nursing time to the unit costs 10-5
Table 10.2 Comparison of direct (face-to-face) care by hospital nurses 10-10
Table 10.3 Key strengths and limitations of the empirical research 10-23
Table 10.4 Illustration of implications of a decision threshold on costing 10-25
Table 10.5 Cost components 10-26

List of Figures

Figure 2.1 Distributions of resource use 2-30
Figure 2.2 Model of nursing workload 2-33
Figure 3.1 Approach to categorising nursing activity used for empirical work 3-2
Figure 4.1 Incidence and prevalence of adult patients on RRT 1993-2005 (England and Wales) 4-2
Figure 4.2 Dialysis modalities of adult patients 1998-2005 (England and Wales) 4-5
Figure 4.3 RSU and MRU patient populations 4-12
Figure 6.1 Example of barcode label 6-1
Figure 6.2 Barcode scanning process 6-3
Figure 7.1 Recruitment of nurses (SUHT and Toton) 7-2
Figure 7.2 Barcode scanning time recorded (SUHT and Toton) 7-3
Figure 7.3 HCAs and RGNs - proportions of time spent on tasks (Totton) 7-4
Figure 7.4 Scatter plots of dependency scores by visit (Totton) 7-9
Figure 7.5 Distribution of patients' mean KPS (n=54, Totton) 7-10
Figure 7.6 Graph of patients' mean dependency score and mean KPS (Totton) 7-10
Figure 8.1 Example of barcode scanning time recorded for one patient 8-5
Figure 9.1 Age distribution of patients (n = 174) 9-1
Figure 9.2 Number of nurses recorded per individual patient's HD session (1624 HD sessions) 9-3
Figure 9.3 Recruitment of nurses 9-4
Figure 9.4 Barcode scanning time recorded 9-4
Figure 9.5 Durations of single direct and indirect care scans 9-6
Figure 9.6 Box plots of direct, indirect and patient-specific care time per HD session 9-7
Figure 9.7 Difference in KPS scores between data collection blocks (n = 126) 9-13
Figure 9.8 Distribution of patients' mean KPS by eligibility for RSU care (n=147) 9-13
Figure 9.9 Number of dependency ratings per patient (1632 HD ratings) 9-14
Figure 9.10 Distribution of dependency scores (all data) by eligibility for RSU care (1579 ratings) 9-14
Figure 9.11 Distribution of dependency scores (all data) by RGN or HCA rater (1426 ratings, 38 nurses including non-study nurses) 9-16
Figure 9.12 Patients' mean dependency and mean KPS (n = 158) 9-17
Figure 9.13 Box plot of summary (mean) patient-specific time by eligibility for RSU care 9-19

Figure 9.14 Scatter plot of summary measures of patient-specific nursing time and dependency by eligibility for RSU care (149 patients, 1514 HD sessions) 9-20
Figure 9.15 Scatter plot of patient-specific nursing time and dependency per HD session (n = 1566) 9-21
Figure 10.1 Issues for consideration in choosing whether and how to measure staff time in a health care setting 10-14
Figure 10.2 Cost components (for measurement or disaggregation) 10-26
Figure 10.3 Suggested approach to choosing costing approach 10-27

List of Appendices

Appendix 1 Barcodes and barcode scanners 11-1
Appendix 2 Literature search for time measurement studies in health care 11-3
Appendix 3 ISHCOF: Renal services in England and Wales 11-7
Appendix 4 RSU study - Nursing time for MRU and RSU type patients 11-8
Appendix 5 Literature review - Data extraction table for costing or economic evaluations comparing of HD across settings 11-10
Appendix 6 Ethical, research governance and data protection issues 11-13
Appendix 7 Barcode scanning feasibility issues (pre-piloting and costs) 11-14
Appendix 8 Coding of tasks at Portsmouth 11-17
Appendix 9 Nurse information sheet 11-18
Appendix 10 Nurse consent sheet 11-20
Appendix 11 Nurses' acceptability questionnaire 11-21
Appendix 12 Work sampling paperwork (Totton) 11-27
Appendix 13 Patient dependency-scoring tool for outpatient HD 11-29
Appendix 14 Karnofsky Performance Scale 11-30
Appendix 15 Co-morbidity indices 11-31
Appendix 16 Results of nurses' acceptability questionnaire (SUHT and Totton) 11-32
Appendix 17 Feasibility of work sampling (SUHT and Totton) 11-35
Appendix 18 Dependency scoring issues (Totton) 11-36
Appendix 19 'Spot checks' of barcode scanning (Portsmouth) 11-37
Appendix 20 Patient dependency-scoring tool - verification of data 11-38
Appendix 21 Estimation of nursing costs - data sources for resource use and unit costs, and assumptions used 11-39
Appendix 22 Scatter plots of total patient-specific time by visit (Portsmouth) 11-44
Appendix 23 Scatter plots of dependency scores by visit (Portsmouth) 11-45
Appendix 24 Plots of patient dependency scores (Portsmouth) 11-46
Appendix 25 Audit of dependency data quality (Portsmouth) 11-47

Acknowledgements

I would like to thank the following people and departments:

My supervisors Professor Paul Roderick, Professor James Raftery and Dr Karen Gerard for their advice and encouragement; and Dr Jane Wolstenholme for her comments and support, particularly on the transfer thesis.

Professor Val Lattimer for examination of the transfer thesis.

Professors Mac MacDonald and Peter Smith for statistical advice on GEE.

My colleagues at both Public Health Sciences and Medical Statistics and the Research Development Support Unit, and especially for advice from the statisticians.

The nurses and other staff at Southampton, Totton and Portsmouth who participated or made the research possible.

Funding by a Nursing and Allied Health Professionals Researcher Development Award from the Department of Health (National Coordinating Centre for Research Capacity Development).

Finally, for their unfailing support, help and encouragement - Dominic, Sam, Hugh, and family and friends.

Abbreviations

95% CI	95% confidence interval
ABC	activity based costing
CKD	chronic kidney disease
DRG	Diagnostic Related Group
EPO	erythropoietin
ERF	established renal failure
ESRD	end-stage renal disease
ESRF	end-stage renal failure
FCE	Finished Consultant Episode
GEE	General Estimating Equations
GP	General Practitioner
HCA	Health Care Assistant
HD	haemodialysis
HRG	Healthcare Resource Group
IQR	inter-quartile range
KPS	Karnofsky Performance Scale
MRU	main renal unit
NI	National Insurance
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NSF	National Service Framework
OLS	ordinary least squares
PbR	Payment by Results
PCS	patient classification system
PD	peritoneal dialysis
pmp	per million population
PSSRU	Personal & Social Services Research Unit (University of Kent)
QALY	quality adjusted life year
RGN	Registered General Nurses (i.e. qualified nurse)
RRT	renal replacement therapy
RSU	renal satellite unit
RSU study	Renal Satellite Evaluation study (Roderick et al 2005)
SATO	Senior Assistant Technical Officer
SD	standard deviation
SE	standard error
SUHT	Southampton University Hospitals NHS Trust
WTE	whole time equivalent (i.e. full time work at 37.5 hours per week)

Chapter 1 Introduction

The thesis examines methods to cost health care and in particular staff inputs. The nature of health care means that costing is complex. It has multiple purposes such as diagnosis, treatment, and monitoring. Final patient-outputs are difficult to define because each patient is unique. Besides, many resources are shared across patients and services, thereby presenting a challenge to track inputs to outputs.

Health care consumes substantial resources and so there are many reasons to estimate costs including resource allocation, budgeting, and service planning. The thesis focuses on costing for economic evaluations - analyses of the costs and benefits of two or more options to inform decision-making. It addresses three main issues. First, it examines the top-down and bottom-up approaches to costing, as the choice between them must balance a number of competing objectives in relation to data quality, feasibility and research costs. Second, it evaluates methods to quantify and cost nursing inputs for different patients, since staff costs are a major part of health care expenditure. Third, it considers the effect of variation between patients (heterogeneity) both on methods to classify health care outputs and to measure resource use. The empirical work brings these three issues together. It assesses the impact of patient heterogeneity on nursing costs in chronic haemodialysis, a costly, life-saving treatment for end-stage renal failure.

This chapter provides the background to the main issues and introduces key principles and concepts. It discusses the application of economics in costing health care, background to costing, importance of staff costs, and measuring staff inputs. Then, it presents the aims and objectives, and gives an outline of the thesis.

1.1 Application of economics in costing health care

Economics involves the study of how resources and incentives affect choices about the production, distribution, and consumption of goods and services. Resources do not refer to money, but the inputs used to produce the goods or service, namely labour, capital (e.g. property, equipment, etc.) and materials. A key principle is 'scarcity', namely that there are never enough resources to meet all potential demands. Consequently, choices must be made, which have 'opportunity costs' because using resources in one way precludes their use for other options.

An important concept in economics is the 'market'

> "a collection of buyers and sellers that, through their actual or potential interactions, determine the price of a product" p7 (Pindyck and Rubinfeld 2005).

Whilst buyers or consumers determine demand for the product, sellers or producers determine supply. Prices incorporate both buyers' value of the product and sellers' production costs. Prices also reflect the opportunity cost, as individual buyers and sellers have traded-off other possible uses for their resources. Through these interactions between individuals pursuing their own self-interest, the market may determine the most efficient use of resources and maximise benefits to society (Mankiw 2004).

Table 1.1 shows the five conditions Donaldson et al (2005) discussed as necessary for a perfect market - one that maximises benefits to society at least cost.

Table 1.1 Conditions necessary for a perfect market - applied to health care

Condition	Description in terms of health care
Genuine competition	Health care providers (suppliers) are numerous and small enough so that individually they cannot affect prices.
Certainty	Patient demand is predictable and individuals know what they want, when they want it and where they can get it.
Perfect knowledge	Patients have perfect knowledge about their health status, available options (including health care), and potential benefits of those options to improve health. They are able to seek the supplier with the lowest cost.
Consumers act freely from suppliers	Patients can act free of self-interested advice from suppliers (health care professionals).
No externalities	Externalities occur when an exchange between two parties results in positive or negative effects to a third party. E.g. individuals benefit from vaccination; however, the unvaccinated also benefit through so-called herd immunity, yet this may lead individuals to rely on others and avoid vaccination.

Source: Information from Donaldson et al (2005)

Donaldson et al (2005) argued that health care fails to match up to these conditions. In particular, uncertainty about the need for health care and associated financial risks means that payment for health care is rarely solely direct from patients. Instead, a third party (government or insurer) pays using funding from taxation, compulsory social or voluntary private health insurance. As a result, the third party such as a Primary Care Trust in England interrupts the link between the consumer and producer (health care provider)[1]. Although patients may pay a contribution, they do not bear the full cost. This distorts 'prices' either because they may no longer reflect the amount consumers would be willing to pay, or because health care professionals lack awareness about the production costs. Furthermore, patients do not have perfect knowledge and health care professionals have dual roles to determine patients' 'need' for health care and to supply health care. Donaldson et al (2005) also argued that a free, unregulated market in health care leads to socially unacceptable outcomes. Some people at high risk will not receive health care because they have insufficient income to pay for it or a risk-adjusted insurance premium. Conversely, there may be an altruistic 'externality' in knowing that those who need treatment will receive it, even if they cannot afford it.

These features of health care have two main consequences. First, almost all health care systems operate with government intervention and regulation that aims to achieve greater benefits to society on efficiency, equity or moral grounds. Second, another mechanism is required to decide how to allocate resources instead of the market. This is the key reason for economic evaluations. These offer an explicit approach to compare the expected costs and benefits associated with different treatments, interventions, technologies such as devices, or service strategies. They provide information to assist decision-makers to use resources efficiently.

Efficiency is significant in two ways. Allocative efficiency addresses how to maximise benefit from available resources by deciding whether to allocate resources to a given objective amongst competing objectives, and if so at what level. For example, if and at what level should NHS resources be allocated to provide cancer care, fertility treatments, cosmetic surgery, etc.? Technical efficiency addresses how to meet a given objective by either maximisation of output from given resources, or minimisation of costs for a given output. For example, given a decision to dialyse a group of patients with renal failure, should they receive haemodialysis at hospital or home?

Cost-benefit, cost-utility and cost-effectiveness are the three types of economic evaluations and Table 1.2 shows how they differ by the outcome measured. Cost-

[1] Provision of health care may be through the public sector or private sector (for-profit and not-for-profit). Although there may be multiple providers, this does not mean there is real competition, not least because of the limited number of providers in any geographical area. However, markets are more obvious for some technologies, although still with some degree of regulation (e.g. drugs, medical devices, and consumables).

benefit analysis assesses allocative efficiency, although there is debate amongst health economists about whether cost-utility analyses can do so too (Phelps and Mushlin 1991, Garber and Phelps 1997, Tsuchiya and Williams 2001). Cost-utility and cost-effectiveness analyses assess technical efficiency.

Table 1.2 Measurement of or valuation of outcomes in economic evaluations

Evaluation	Measurement or valuation of outcomes
Cost-effectiveness	Units such as life years gained, a point change on a pain scale, etc.
Cost-utility	Quality-Adjusted-Life-Year (QALY) i.e. length of life weighted by a value of the health state
Cost-benefit	Money (i.e. for costs and benefits)

Various bodies commission and use economic evaluations, typically cost-per-QALY or cost-effectiveness analyses. The Department of Health commissions evaluations through the Health Technology Assessment Programme. The National Institute for Health and Clinical Excellence (NICE) both commissions assessments and produces guidance to the NHS in England and Wales. In decision-making, results of economic evaluations are balanced against a number of other factors, including equity issues about the fairness of funding and distribution of health care resources.

1.2 Background to costing health care

The thesis focuses on costing in economic evaluations, although there are many other reasons for costing. The substantial resources devoted to health care have to be accounted for. Black (2005) identified seven kinds of users of financial information: investors, lenders, suppliers and trade creditors, employees, customers, government and their agencies, and the public. A further group was the management of the organisation. He noted that information broadly separated into:

- financial accounting: to record day-to-day financial transactions, summarise transactions to satisfy the groups listed above, and comply with external rules and regulations; and
- management accounting, the internal accounting within the organisation for decision-making and planning.

In the NHS, Brown and Green (2006) identified three main purposes for costing:

- to provide detailed financial information for commissioners and providers of services to support the performance and monitoring of service delivery,
- to benchmark services across sectors,
- to support negotiations for revision of funding and wider decision-making.

Since the purposes for costing and types of information vary, it is unsurprising there are numerous terms and different methods to estimate 'costs'. Indeed, economic and accountancy approaches to costing differ. Costing in economics is based on the 'opportunity cost'. This is the value of benefits from the resources used for their best alternative, regardless of whether bought. Hence, the opportunity cost of creating a new nurse specialist role might be the value of either nursing time no longer available, or the other therapies such as drugs that could have been bought. In contrast, accountancy is based on financial costs, the money actually spent on resources. Consequently, although a donated item has no value in accountancy, it does have an 'economic' value for the replacement cost. Likewise, unpaid care has no accountancy value, but the opportunity cost might be the carer's lost earnings. Furthermore, whilst accountancy operates at an organisation level (such as a company, hospital or the NHS), economists favour a broader, societal perspective (Gold et al 1996, Drummond and Jefferson 1996). Nevertheless, in practice, routine costing in the NHS is accountancy-based, and it is not easy to value resources at their opportunity cost, an issue revisited in section 2.1.6.

Regardless of the differences between accountancy and economics, costing in health care is complicated. One challenge is how to define the outputs or 'products'. Unlike many other industries, health care serves multiple purposes and produces multiple outputs. Fetter (1991), suggested that hospitals have two separate production functions:

- to produce standard or intermediate outputs - goods or services such as meals, laboratory tests, x-rays or inpatient days, and
- to diagnose and treat patients, which involves use of the various intermediate products, and is the main overall function and output. For example, a hospital may treat patients across all age groups and according to numerous patient categories (e.g. emergency / elective admissions, medical / surgical).

Final 'patient-outputs' are difficult to define in terms of a 'market' because patients are unique. They vary in many characteristics: physical, mental, social, and clinical diagnosis, procedures or illness severity. Hence, uniform (homogenous) outputs, with exact quality control guidelines, do not exist. Yet, as noted by Fetter (1989), health care needs a classification of 'case mix' - products or patients - to help understand cost differences both between hospitals and between patients within a hospital. Diagnosis Related Groups (DRGs) were an attempt at such a classification for use to examine and control expenditure, and to make reimbursement fairer (Fetter 1989, Fetter 1991). In the US, government (Medicare and Medicaid) and insurers' reimbursement of health care providers is based on DRGs. An equivalent classification, Healthcare Resource Groups (HRGs), was developed in the UK (Sanderson 1989). Indeed, HRGs are central to national reference costs and to the national tariff used for 'Payment by Results' (PbR) to reimburse NHS health care providers (Department of Health 2002b).

A further challenge is how to link resource inputs to different outputs, especially as staff, services and departments are often shared. Decisions are needed about how to attribute resources where they cannot be tracked directly to outputs. In particular, whilst patients vary in their need for staff input, some patients may require consistently more or less staff input. Being able to quantify the effect of this variation between patients (heterogeneity) is central to ensuring like-for-like comparisons between services. For example, a common NHS policy has been to move care into the community (Department of Health 2006d). Costing a new service or setting is relatively straightforward if patients are well defined and have similar care needs. In contrast, the existing service often comprises a heterogeneous group of patients for whom it is more difficult to attribute staff inputs.

Costing entails three steps; i) identification of resource use; ii) measurement of resource use in meaningful units; and iii) valuation by multiplying the measured resource use by a monetary value (the unit cost). The health economics literature describes two broad approaches to costing - top-down and bottom-up costing - (Drummond et al 2005, Brouwer et al 2001, Luce et al 1996). Top-down or gross costing is a process of disaggregation. It uses financial (expenditure) data divided by units of activity. For example, the cost per bed-day could be the total ward costs for staff, equipment, drugs, consumables, etc. divided by the bed-days occupied by patients. The costs tend to be composite intermediate products, which are often large relative to the total value of resources. Since resources are often shared across multiple patients, interventions or services, the approach commonly requires assumptions about how to apportion expenditure to activity.

Bottom-up or micro costing is a process of aggregation. It involves collecting or estimating resource use data on all the component parts that contribute to the activity and obtaining a unit cost for each. These component costs are then added together to derive the cost for the activity. Using this approach, the cost per bed-day would be the total sum of each of the ward's cost elements for staff, equipment, drugs, consumables etc. Whilst the unit costs may be derived from actual expenditure, often standard costs are used such as the mid-point salary of national pay scales for staff.

Drummond et al (2005) suggested that top-down costing offers broad averages, and is cheap and quick, whereas bottom-up costing offers increased precision, richness of data, and allows greater insight into activities and costs. Furthermore, the top-down approach ensures that costs cover the total expenditure, whereas this may not be so for the bottom-up approach. The thesis examines costing advice and empirical evidence to assess the effect of choosing one approach over the other.

1.3 The importance of staff costs

Staff are a major health care expenditure as salaries constitute approximately 60% of NHS spending (2003, http://www.performance.doh.gov.uk/HPSSS/TBL_E3.HTM). The dominant staff group is nurses whose salaries account for about 70% of staff expenditure (Department of Health 2000b). Therefore, staff inputs are likely to be important considerations in many evaluations, including where:

- staff are a major part of the intervention e.g. psychological interventions such as counselling patients,
- the interventions compared require or result in different staff time inputs,
- specifically, there is substitution of roles, services or settings, since these may result in differences in use of staff time or skills.

A previous research project, the Renal Satellite Evaluation study (RSU study, Roderick et al 2005), was the motivation for the thesis. This evaluation compared central with local service provision that was designed to improve geographical access. It is described in detail in section 4.5.2. In summary, the study aimed to estimate the costs and outcomes of haemodialysis in two settings - renal satellite units (RSUs) and their parent main renal units (MRUs). Whilst haemodialysis involved a range of staff (including doctors, technicians and clerical workers), nurses provided the major routine input. RSU patients were relatively homogeneous whilst

MRU patients were heterogeneous since some were sicker or more unstable than those at RSUs. At that time, it was not possible to resolve how to allocate the cost of nurses' time to different patients, and so the evaluation averaged costs across all patients. The effect of not costing on a like-for-like basis was unclear. Averaging should have inflated the costs of the MRU patient sub-group and introduced bias in favour of treating patients at RSUs. Yet it was unclear whether differences in resource use between the patient groups were sufficient to affect costs. The thesis examines this and the related questions, namely i) do researchers measure staff time and if so how, or do they simply average costs across patients, and ii) do researchers acknowledge the influence of patient heterogeneity?

Evaluations of role substitution are further examples where it is vital to measure staff inputs. In nursing there has been much interest in substitution or extension of roles, new roles, generic workers, and the provision of specialist vs. generalist care (Royal College of Nursing 2003b). Moreover, the Department of Health has been looking at workforce capacity (numbers and skills) to improve productivity and find new ways of working (http://www.dh.gov.uk/en/Policyandguidance/Humanresourcesandtraining/Modernisingworkforceplanninghome/index.htm). Local initiatives have included Workforce Development Confederations to plan and develop the health care workforce. There has been emphasis on moving work from doctors to other health care professionals and from health care professionals to the support workers, partly to decrease doctors' working hours in response to the European Union Working Time Directive (The Working Time Regulations 1998). In addition, changes in nursing training and the advent of supernumerary status for students have led to use of support workers. Whilst some changes have been borne out of necessity, other role substitutions have been proposed to improve skill mix efficiency or to find cheaper ways of working (Gibbs et al 1991). Other changes have sought to expand the quality of services by altering staff roles, settings, or both (e.g. specialist rather than generalist care).

Given such interest in staff roles and their budgetary impact, it is surprising that economic evidence is generally lacking. Richardson et al (1998) found few cost-effectiveness studies on doctor-nurse substitution. Evidence was mostly dated (1970s and 1980s), from the US, and often from poorly designed studies with small sample sizes at single sites, thereby limiting the generalisability of the results. Systematic reviews by Horrocks et al (2002) and Laurant et al (2004) found similar methodological limitations in comparisons of nurse practitioners and doctors in primary care. In general, there were no differences in health outcomes, though patient satisfaction was greater for nurse practitioner care. However, nurses' consultation times were longer, they recalled patients more frequently, and requested more investigations. The effect on costs was variable, although few studies provided cost data; often these were underpowered and used different approaches. Indeed, others have challenged the assumption that nurses are a cheap alternative to doctors (Spilsbury and Meyer 2001, Watts et al 2001).

Substitutions of services or settings may also result in different usage patterns for staff time or skills. The 'NHS Plan' outlined increased investment in intermediate care facilitates to promote early discharge or prevent admission (Department of Health 2000a). The NHS Confederation (2006) made the case for the continuing reduction in hospital beds, increased care in the community and in some cases, care from specialists rather than the local hospital. There is an extensive literature on various intermediate care schemes for early discharge or admission prevention. These embrace 'hospital at home', early supported discharge, rapid response teams, community rehabilitation teams, day hospital and day care centres, community hospitals, nurse-led units, and care homes (Roe 2005). A Cochrane review of hospital at home schemes by Shepperd and Iliffe (2005) found that patient outcomes were generally similar. They concluded that whilst patient satisfaction may be higher for hospital at home, the carers' burden was greater and there was little evidence of cost savings to the health service, although relatively few included the cost impact. Overall, Little (2005) concluded that there remained a shortage of evidence on the economic benefits of intermediate care and such schemes do not always save money. In summary, whilst these evaluations demonstrated the need to assess staff inputs, this was often lacking. The next section considers how staff inputs might be measured.

1.4 Measuring staff inputs

Staff inputs are complex aspects to measure and value. The inputs are multi-faceted and encompass not just time but the personnel involved (i.e. nurses, doctors, etc.), grade mix (within a professional group), and skill mix (i.e. variations in experience or proficiency). Furthermore, there may be interaction between staff inputs, the quality of care and patient outcomes. Even more complexity arises in actually valuing staff inputs. Graham and McGregor (1997) reviewed papers that estimated the costs of a 10-minute GP consultation. They found 14 different estimates across 11 studies, although many studies were unclear about how they derived the costs. One reason why the costs differed was that the studies varied in the cost elements they included (e.g. heating, lighting, etc.). Moreover, most GPs are not salaried; they receive income to run the practice, which makes it more difficult to cost their time compared with other (salaried) health professionals. In contrast, standard salary costs are available for many staff in the "Unit costs of health and social care" (Curtis 2007), produced annually on behalf of the Department of Health.

Measurement of staff inputs may focus on tasks, aggregated activity or patients. Staff-patient interactions have inherently flexible or ill-defined boundaries, which can make consistent coding and measurement of activity difficult. Care for individual patients is either directly face-to-face, or indirect (i.e. away from the patient). Indirect care includes writing reports, phone calls, and arranging transport or support for the patient. Staff can record their own activity to capture such patient-level care, although this interrupts their work. Alternatively, an observer can record staff activity; however, it is difficult for an observer to link indirect care activities to patients. Furthermore, many indirect care activities may not be readily distinguishable from other general administrative activities that are not for any specific patient.

Conversely, some staff-patient interactions are relatively well defined. For example, whilst consultations booked into fixed appointment times may over or under-run, the scope for variability is likely to be less than staff-patient interactions for example on a ward. Moreover, an appointment schedule also locates members of staff, making it easier to track staff and monitor consultation times. Therefore, depending on the purpose of costing, measurement of the duration of some staff-patient interactions may not be worth the research effort.

Techniques to measure staff time include observation methods of time and motion and work sampling, self-recording, retrospective self-reporting and expert opinion. Nevertheless, Adam et al (2003a) identified that a major gap in the health economics literature was advice on data collection methods for resource use and costs, and especially practical guidance on how to cost staff inputs that are shared. The thesis evaluates the pros and cons of methods to measure staff time, especially at the patient-level, and reviews literature to determine how staff inputs have been quantified for economic purposes.

1.5 Aims and objectives of the thesis and empirical work

The previous sections have outlined the background to the thesis. The aim of the thesis is to examine methods to quantify and cost staff inputs for specific patients for economic evaluations. Based on the literature review, two methods of time measurement were selected for testing in the empirical work: nurses self-recording their time using barcode scanners, and observer work sampling. Chronic haemodialysis was used as a case study, both to address questions that arose from the RSU study, and because haemodialysis is an expensive treatment with ongoing pressure on the NHS to increase provision.

The objectives of the thesis are:

1. To give an overview of costing and review costing approaches for economic purposes.
2. To evaluate methods to measure staff time per patient.
3. To assess the effect of patient heterogeneity on nursing costs for chronic haemodialysis.

1.6 Outline and scope of the thesis

Chapters 2 to 4 provide the background and literature review. The search for studies in the literature review faced a major challenge as the terms 'cost' and 'time' are widely used and nonspecific. Consequently, within the scope of the thesis, some search strategies were restricted by publication year to limit papers to a manageable number of recent relevant ones. Additional information was gathered from other sources including expert opinion (supervisors' input, presentation of a discussion paper at the Health Economists' Study Group, and discussion with other researchers).

Chapter 2 addresses the first objective of the thesis. The first part provides an overview of costing, introducing key concepts, terms, principles, and issues. The second part examines literature on costing guidance and research. After an overview of costing in economic evaluations, it reviews the top-down and bottom-up costing approaches and classifications of patient heterogeneity.

Chapter 3 addresses the second objective of the thesis, evaluation of methods to measure staff time per patient. It discusses the categorisation of staff activity and outlines methods to measure staff time, and compares their advantages and disadvantages. Then it reviews literature on time measurement to examine use of the techniques in economic evaluations and health care more generally.

Chapter 4 sets the context for the third objective of the thesis and explains the rationale for using haemodialysis as a case study. It gives an overview of clinical and epidemiological aspects of renal failure, available treatments, and policy initiatives to meet the demand for treatment. Finally, it reviews economic evaluations of haemodialysis in different settings since patient heterogeneity could affect costs.

Chapters 5 to 9 describe the empirical work. Chapter 5 introduces the empirical work by drawing together findings from the earlier chapters. It gives the rationale for the methods chosen in relation to three aspects: use of haemodialysis as a case study, measurement of nursing time per patient, and classification of patient heterogeneity. Lastly, it gives the rationale for data collection sites and piloting.

Chapters 6 and 7 describe the pilot phase, which had three purposes: to establish data collection using barcode scanners, to assess work sampling, and to test the methods to classify patient heterogeneity. Chapter 6 describes the methods both for time measurement and three tools to classify patient heterogeneity in those undergoing chronic haemodialysis. Chapter 7 presents the findings.

Chapters 8 and 9 address the main data collection. Chapter 8 describes additional information about the methods - where they differed from those of the pilot phase - whilst Chapter 9 presents the findings.

Finally, Chapter 10 begins with an in-depth discussion of the findings from the empirical work and literature reviews. It then discusses the strengths and limitations of the empirical research, makes recommendations for future research, and presents overall conclusions and outputs from the PhD.

Chapter 2 Costing health care

This chapter addresses the first objective of the thesis. The first part provides an overview of costing. This includes key concepts - the production and cost functions - and costing terms. It then outlines principles and issues in costing: the scope of exercise, feasibility, sources of resource use and cost data, and routine costs available in the NHS. The second part examines literature on costing guidance and research. After an overview of literature on costing in economic evaluations, it reviews the top-down and bottom-up approaches and classifications of patient heterogeneity.

2.1 Overview of costing

This section gives an overview of key costing concepts, terms, and considerations as outlined in Table 2.1. A later section (2.2.2) discusses the top-down or bottom-up approaches to measure and value resources or unit costs.

Table 2.1 Outline of chapter sections addressing key aspects in costing

Aspect	Covers	Section
The production and cost functions	-	2.1.1
Key costing terms	Describes types of cost - fixed, variable, total, average, marginal and long-run marginal	2.1.2
Scope - purpose, perspective and timeframe	Guides the identification of resource use and relevant unit costs	2.1.3
Feasibility	Practicalities of data collection and targeting research effort	2.1.4
Resource use (i.e. inputs or factors)	Sources of data	2.1.5
Unit costs (i.e. input or factor prices)	Sources of variation in unit costs, sources of data (including routine costs in the NHS) and considerations about using or adapting existing unit costs	2.1.6

2.1.1 The production and cost functions

In economics, 'cost' incorporates two inter-related functions outlined in Table 2.2. The production function quantifies the overall output in terms of the component resources used (i.e. inputs, or so-called factors of production). The cost function values the output either directly or by adding together the costs of the component resources used. It is possible to achieve a given level of output using many combinations of the resources. For example, either a trained nurse or a health care assistant may deliver a patient treatment. Bottom-up costing involves establishing the full production function - identifying, measuring and then valuing the components (see equation 2). Top-down costing involves estimating the production function for items at a more general or aggregated level, such as outpatient visits or hospital days, before valuing them (see equation 1).

Table 2.2 Definition of the production and cost functions

	Production function	Cost function
Describes relationship between	inputs and outputs	output and total cost
Represented by	Q = f(L, K, M)	1) TC = f(Q, r) or 2) $TC = r_1L + r_2K + r_3M$
where	Q is the quantity of output L, K and M are the inputs (i.e. factors of production) of labour, capital and materials TC is total cost r is the price of the inputs (i.e. factor prices)	
Most efficient is the combination of ...	inputs that maximises output	inputs that minimises cost for the output with given factor prices

Source: Adapted from Jan et al (2005)

The production and cost functions are the basis for estimating costs, and for economic analyses of efficiencies and costs. Examples include an early study by Gunn and Douglas (1942) who assessed the relationship between labour, capital and output in 1914 across manufacturing industries in the United States. In health care, Reinhardt (1972) and Thurston and Libby (2002) examined the production function for medical services to assess how other workers might be used in addition to, or as substitutes for doctors to increase efficiency. Statistical or econometric analyses of the cost function have been used in primary care to investigate economies of scale and scope[2] (Giuffrida et al 2000). Other cost function analyses include investigations into why costs vary due to age, proximity to death or both (Seshamani and Gray 2004). Given the many inputs and sources of variation, it is unsurprising that such analyses have been found challenging (Lave and Lave 1970, Lave et al 1972, Smet 2002, Smet 2004).

[2] Scale efficiency is associated with the level of production of a single output and scope efficiency is associated with the range of outputs produced.

2.1.2 Key costing terms

Table 2.3 defines a number of costing terms salient both to sourcing and to using cost data. The definition of fixed or variable depends on the timeframe since in the long-term all costs are variable. Typically, the graph of average (total) cost and output is 'U' shaped. Average costs initially fall as output increases. Then, as output continues to increase, average costs start to rise because proportionally more resources are required. Similarly, the marginal cost initially falls, but then rises as output increases, and the curve intersects the average cost curve at the minimum average cost.

Table 2.3 Definitions of key costing terms

Term	Definition	
Fixed cost	Costs for inputs that do not vary by outputs (e.g. ground rent, capital).	FC
Variable cost	Costs for inputs that vary by quantity of output (e.g. drugs and consumables only used if treating a patient).	VC
Total cost	Market value of all inputs used in production.	$TC = FC + VC$
Average (total) cost	Total cost divided by quantity of output.	$ATC = TC / Q$
Marginal cost	The change in total cost for an extra unit of output (i.e. the gradient at each point of the graph of total costs and quantity)	$MC = \Delta TC / \Delta Q$ or
	Or the change in variable costs for an extra unit of output (because fixed costs are unchanged)*.	$MC = \Delta VC / \Delta Q$
Long-run marginal cost	The change in total cost for an extra unit of output when all costs are variable.	

Key: * Pindyck and Rubinfeld (2005)
Source: Adapted from Mankiw (2004)

Economists emphasise the difference between marginal and average costs. Marginal costs capture the effect of small increases or decreases in output and typically differ from average costs. Drummond et al (2005) noted that the cost of an extra day in hospital might be less than the overall average cost per day because treatment costs often decrease after the first few days of admission[3]. Jan et al (2005) argued that it is difficult to measure marginal costs without a large number of observations and so changes in costs may be measured over several units, i.e.

[3] In contrast, 'hotel' costs would be broadly similar throughout the hospital stay.

'incremental' cost[4]. Luce et al (1996) proposed that economic evaluations should use long-run marginal costs. Nonetheless, Drummond et al (2005) argued that in practice, marginal costs are context specific. They concluded that since economic evaluations assume resources freed will be redeployed efficiently, analysts should at least point out where this may not be the case, even it is not explored in detail.

2.1.3 Scope of costing exercise

In setting out to estimate costs, a number of aspects influence the scope, conduct and validity of the costing exercise:

1. The purpose or question addressed, which depends on both who commissions the evaluation and who will use the results.
2. Perspective, since it determines the resource components that should be included. Economists favour a societal perspective (Gold et al 1996, Drummond and Jefferson 1996), which by incorporating all stakeholders including patients and carers, maximises benefits to society and avoids inadvertent cost shifting. In contrast, for decision-making in the NHS, the National Institute for Clinical Excellence (2004a) recommends a narrower health and social care perspective. Alternatively, the perspective may be that of a specific budget holder.
3. Timeframe, as this affects whether costs are treated as variable or fixed (see previous section) and how far into the future to cost events.

Mogyorosy and Smith (2005) stated that selection of a costing methodology must find an optimal balance between the following competing objectives:

- Costing based on detailed, comprehensive and representative resource use and unit cost data.
- Accurate, precise cost measurement[5].
- Reliable and valid cost measurement that minimises the following biases:
    - Methodological (e.g. through incorrect treatment of fixed or variable costs, not costing options under the same rules, and inappropriate inclusion or exclusion of 'costs').

[4] Incremental cost also refers to the difference in costs between options in an economic evaluation.
[5] In the thesis, accurate is taken to mean correct and valid (i.e. measures what it purports to measure) and precise means exact (e.g. having narrow 95% confidence intervals).

  - Case mix and service mix (i.e. the over- or under-estimation of costs due to averaging across patients or services with different resource intensities).
  - Site selection (where costs at a single site are not representative of the 'standard' average or most efficient site).
- Value for money of the costing exercise.

Whilst these criteria chiefly relate to data quality, as shown in the next section, feasibility issues are also important.

2.1.4 Feasibility issues and targeting research effort

Feasibility is an overarching concern in costing. Data issues are fundamental. Data may be unavailable, or in an unsuitable format if collected for a different purpose. Alternatively, resource use data may not match available unit costs. Access to data may be limited by location, or because help is required (e.g. to decipher codes), or inaccessible (e.g. prices paid for drugs that are 'commercial in confidence'). Furthermore, primary data collection may not be possible because methods are unsuitable or difficult to implement in practice. Moreover, the need for the data must balance the burden and acceptability to staff or patients and so minimise the risk of missing or poor quality data.

Like all resources, those for research are limited and effort should be targeted to maximum impact. As a first step, the researcher needs to understand the relevant production process; how does the epidemiology of the condition before, during, and after the intervention affect resource use. Answers may come from existing data or it may be necessary to conduct a pilot study using one of the methods outlined later (see Table 2.4).

A second step is to work out the relative importance of the resources since some will be major contributors to the overall cost - so called cost drivers. This is challenging because, from the cost function, the impact depends on the interaction between resource use and unit costs. Consequently, an item may be influential because although of small monetary value it is used in large quantities, or *vice versa*. The effect may be at the level of either individual items or aggregate cost categories. Hence, cost drivers provide an overview of resource consumption that may be helpful in assessing overall expenditure and cost of illness studies, etc. Furthermore, the drivers of differences in costs are central to assessing technical efficiency and marginal analyses. Johnston et al (1999) called these drivers of cost differences key cost-generating events, defined by:

- actual or hypothesised variation in the frequency of patient-events across treatment groups, or within a treatment group,
- the impact on cost (either the relative magnitude of cost item or due to the effect on incremental costs),
- the consequences for study validity if data are not collected, both internally in reflecting actual costs, and externally for generalisability.

Identification of cost drivers is complicated in health care. Whilst clinical staff understand patient activities, these staff may lack the knowledge about the cost implications. Chilcott et al (2003) advocated 'pre-study' modelling; however, the degree to which this will help depends partly on the validity of estimates. Moreover, the judgment about what constitutes a large cost driver or meaningful cost difference is chiefly the domain of decision-makers who use the outputs, but who may not be involved in the commissioning or design stage.

Information on cost drivers can be used to target research effort, although with implications for the validity and use of results. If the purpose is to detect differences between interventions, items that are the same for each comparator may be excluded. These results are likely to be context-specific and therefore not comparable with other studies, and they cannot inform the overall budget impact. It seems sensible to focus effort on relevant major cost drivers, since these items will be crucial to the validity of the estimates. Yet decisions remain about small or unimportant cost components - should these simply receive less effort or it is justified to exclude them altogether?

Empirical support for restricted costing is limited. Knapp and Beecham (1993) examined the potential to predict total costs from resource use items. In five studies of mental health care services, use of each study's top five most expensive components accounted for 90-97% of the full cost. Using the top five most expensive global components (i.e. averaged across all the studies) accounted for 61-97% of the full cost. The authors suggested that such short cuts to costing might be useful, except when an evaluation requires comprehensive resource use and costs, or when there are inter-individual differences in resource use (i.e. where the group average is inappropriate).

A similar study by Whynes and Walker (1995) was less successful in predicting the costs of surgery for colorectal cancer. They agreed that reduced list costing was inappropriate when there was significant inter-patient variation. They concluded that such costing compromised accuracy and was less appropriate in acute care. Whilst the methods were relatively successful after a detailed costing exercise, they felt that identification of the high cost aspects *a priori* would have been difficult. These studies highlight the difficulties in identifying cost drivers and suggest that comprehensive costing is necessary where inter-patient differences are expected.

Having outlined important feasibility issues, the next two sections consider sources of resource use and unit cost data.

2.1.5 Resource use (inputs)

Production of an output typically combines multiple resource components (labour, capital, and materials). Dependent on the perspective, various resource categories may be relevant, such as intervention-related (staff, consumables, overheads[6]), patient out-of-pocket expenses, informal carers' time, productivity time for the patient's lost time.

Resource use data may come from staff, patients, or proxies (e.g. relatives). Sources include primary data collection using prospective or retrospective methods, or secondary sources - existing data collected for other purposes - but none is perfect. For example, a review by Evans and Crawford (1999) demonstrated that patients' recall of resource use declined over time. Recall was also affected by the salience of the event or resource item (e.g. hospitalisation, drug use, etc.), perceived social acceptability of the condition, and other factors such as educational status and mental condition. Moreover, the associated under- or over-reporting potentially affected the results of cost-effectiveness analyses.

The source of resource use data needs to be appropriate for the study's purpose and the choice depends on a trade-off between possible biases, validity[7], reliability[8], and feasibility issues. Table 2.4 gives an overview of data collection methods, their

[6] Shared resource use such as human resources and estates, which do not directly link to the output of interest.
[7] I.e. measures what it purports to measure.
[8] I.e. dependable and repeatable (e.g. if measured on more than one occasion, would give the same results providing all influences remained unchanged).

main sources of bias, and an approximate order from the least to most expensive cost to collect the data.

Table 2.4 Data collection methods for resource use and main sources of bias

Data collection method	Main sources of bias
Extraction from routine records (e.g. on to a case report form) • Routine medical or other health records (paper or electronic) • Administrative databases (e.g. number of patients, bed-days, etc.)	Information available inconsistent or missing (or not in format required).
Extraction from published literature	Publication bias (i.e. reporting of results depending on their result). Information may not be applicable to setting of interest.
Questionnaires (face-to-face, postal, email)	Recall and response bias (e.g. non-response and answering in ways thought to be socially desirable).
Interviews (face-to-face, telephone)	Recall and response bias as above.
Expert opinion (including Delphi methods)	Recall and response bias as above, choice / mix of experts.
Self-recording using diaries or logs (paper or electronic)	Altering subject behaviour (Hawthorne effect), compliance (accidental or conscious omissions).
Observation (e.g. time and motion study).	Altering subject behaviour, observer omission (accidental or conscious)

2.1.6 Unit costs

The unit cost is the cost to produce one unit of a product. This section outlines factors that influence unit costs, sources of data and considerations when estimating or using unit costs.

Table 2.5 shows the many aspects that influence an item's unit cost either directly through the monetary value or indirectly through the impact on resource use. As Adam et al (2003b) found, these influences lead to wide variation in unit costs between centres. Yet Sculpher et al (2005) noted a reliance on sensitivity analysis and little actual research into quantifying variability or uncertainty in unit costs.

Table 2.5 Direct or indirect influences on unit costs

Aspect	Description
Input prices	Amounts paid for resources - influenced by contract negotiations and life cycle or learning curve effects.
Outcomes (products)	Outcomes (an individual's health or welfare, or intermediate outcomes) affected by variations in service delivery or quality of care.
Capacity issues	Occupancy or throughput rates and economies of scale.
Patient characteristics / 'case mix'	E.g. background characteristics, need, dependency, and epidemiological factors (see section 2.2.3).
Clinical practice / service delivery variations	Skills or experience, payment systems / incentives and preferences, alternative treatment strategies and substitution effects affecting resources available.
Location	Wages and other local effects, e.g. London / non-London wages and property prices etc.
Sector of ownership	Public, voluntary or private.
Efficiency	Maximising outputs or minimising costs.

Source: Adapted from Beecham et al (1993)

Apart from primary estimation of unit costs for a specific study, Table 2.6 shows a variety of secondary sources. Again, the source needs to be appropriate to the study's purpose and the choice depends on trade-offs between various biases, validity, reliability, and feasibility issues.

Table 2.6 Examples of sources of unit costs and methods of derivation

Source	Main method of derivation
National reference costs or the national tariff (based on Healthcare Resource Groups). Typically aggregated across multiple conditions or procedures (see section 2.2.3.3)	Top-down
The annual "Unit costs of health and social care" (Curtis 2007) (see section 3.4.1.1)	Bottom-up
National pay scales, manufacturers or NHS list prices e.g. British National Formulary (BMA 2006)	Typically price
Local NHS sources, e.g. hospital finance departments (see section 2.1.6.1)	Top-down
Previous cost studies	Top-down or bottom-up

Several considerations are relevant when estimating or using unit costs. In economics, there is a fundamental distinction between costs and prices. Cost relates to production - the value of resources used. Prices reflect what consumers are willing to pay (i.e. their value of the product), which incorporates both production costs and profit. In health care, 'market' prices exist for goods such as consumables and equipment, whilst for other aspects 'market' pricing is limited or absent. Providers may have limited control over the payment levels set for tariffs, which in some cases may not even cover production costs. Other pricing mechanisms do not involve market forces and include cost-based strategies to add a profit margin to relevant production costs (Netten 1993). Hospitals in the more market-orientated United States have so called cost-to-charge ratios, which allow for the 'mark-up' charged on the actual resource expenditure. These can be used to adjust data to reflect the production costs, but there are no equivalents in the UK.

From an economic viewpoint, a cost should reflect the opportunity cost, but this may not be straightforward. Mogyorosy and Smith (2005) considered circumstances where market prices might vary from opportunity costs or where market prices do not exist (e.g. informal care). They summarised a number of arguments. One view was that, in practice opportunity costs may be context specific, calculation can be time consuming and expensive, and that in some circumstances routinely available accountancy costs may be reasonably good proxies. Alternatively, opportunity cost was not a type of costing system but an approach to decision-making where resources are scarce, and therefore the options considered should include all feasible alternatives. Indeed, Drummond et al (2005) noted

> "most studies use market prices unadjusted and it has often been remarked that health economists recognize that market imperfections exist in health care, unless they are undertaking an economic evaluation!" (p58).

These authors advocated adjustment of prices that would otherwise introduce substantial bias, and if adjustment could be done in an objective way.

The implications of using opportunity costs depend on the purpose of costing. Netten et al (1998) argued that sometimes (e.g. in role substitution), due to the training costs, salary costs alone do not reflect the true costs of making a professional available. For a health service provider, such costs may be irrelevant, but they have consequences for the NHS and society. A Department of Health study quantified the equivalent annual costs of educating health professionals based on their estimated working life. For doctors, the impact of education costs was especially dramatic, as it increased the unit cost by up to 66% (Netten et al 1998).

In summary, estimation or use of 'cost' information requires consideration of whether for a given purpose, the data are suitable as they are, or whether and how to adjust the data to reflect better the data required. Furthermore, to ensure comparability across data, other adjustments may be necessary. Cost data may come from different years and need to be inflated or deflated to bring them into a common year (the base year) using inflation indices[9]. Likewise, to compare interventions that incur costs across a number of years, the costs need to be adjusted to their present day value by discounting costs beyond the first year.

2.1.6.1 Routine costs available in the NHS

This section outlines the *routine accountancy-based (top-down) costing* in the NHS. Despite their shortcomings from an economic viewpoint, the availability of these costs makes them a principle source of data.

In the NHS, all care providers must comply with the NHS Costing Manual (Department of Health 2008b). This follows three broad principles that costs should:

1. be calculated on a full absorption basis (i.e. to recover all costs),
2. maximise direct charging or otherwise use standard apportionment methods,
3. avoid cross subsidisation by matching costs to services that generate them.

Ultimately, expenditure is broken down into costs for inpatients and day cases by assignment directly to patients or progressively through four levels of aggregation:

1. cost centres (i.e. treatment services and support services),
2. patients or costing pools (i.e. specialities, services, programmes),
3. at the point of delivery (e.g. day cases, outpatients etc.),
4. inpatient and day case activity. This level comprises costs for Healthcare Resource Groups (HRGs, described in detail in section 2.2.3.3) or other patient-related activity outside the HRG classification.

The manual uses three cost classifications outlined in Table 2.7.

[9] E.g. health-specific inflation indices available from the Department of Health (2007).

Table 2.7 Cost classifications in the NHS costing manual

Direct	Attributed directly to a service via cost centre or patient, e.g. drugs.
Indirect	Shared by a number of cost centres or patients with allocation based on activity data, e.g. linen allocated by patient days and pharmacists salaries allocated in proportion to the number of items dispensed.
Overhead	Support services typically not involved in face-to-face patient contact with apportionment based on a 'fair share' (not activity data). E.g. building maintenance apportioned by building volumes. As a last resort, other overheads (e.g. the Chief Executive), may be apportioned by the gross cost of patient treatment services.

Source: Department of Health (2008b)

The costing manual allows local flexibility to define costing centres or pools. Other aspects are imposed, such as which staff costs are direct, indirect or overheads. Within costing pools, indirect or time-based[10] costs are averaged across activity to derive a unit cost for a bed-day, theatre hour or session, or attendance, etc. The manual advocates extensive involvement of clinicians, nurses and other professionals to understand the patient activity and where necessary to estimate the level of staff input.

2.2 Literature review of costing guidance and research

Having looked at background information on costing, this section examines the costing literature. After an overview, it reviews guidance and research into the top-down and bottom-up costing approaches. Then, it examines the classification of patient heterogeneity through various 'case mix' measures, both to differentiate between patients and to define outputs of health care.

2.2.1 Overview of literature on costing in economic evaluations

Guidance on costing for economic purposes is available both within guidelines on economic evaluation and independently. An important example is the consensus statement from the US Public Health Services Panel on Cost-effectiveness in Health and Medicine (Gold et al 1996). Specific costing guidance has been issued in various manuals, reports and journals, for example in Canada (CCOHTA 1996) and the Netherlands (Oostenbrink et al 2002).

[10] I.e. related to contact time or duration of stay

In the UK, various sources of guidance are available. These include the National Institute for Health and Clinical Excellence in England and Wales (National Institute for Clinical Excellence 2004) and two key texts by Drummond (Drummond et al 2005, Drummond and McGuire 2001). In addition, a major source of information is the Personal & Social Services Research Unit (PSSRU) at the University of Kent. Since the early 1990s, the Department of Health has funded a PSSRU research programme to develop costing methods and to define good practice. Following the output of an initial workshop (Netten and Beecham 1993), the PSSRU has produced an annual schedule of costs called the "Unit costs of health and social care" (Curtis 2007). These include guidance on the derivation and limitations of the costs and general articles on costing issues (Netten and Beecham 1999, Netten 2001, Hutton 2001, Netten 2002, Netten 2003, Curtis and Netten 2004, Netten and Curtis 2005). An advisory group, which includes members of the Department of Health, PSSRU, Centre for Health Economics, and Centre for the Economics of Mental Health, meets once a year. Other costing guidance, for accounting and reimbursement purposes includes the NHS costing manual (Department of Health 2008b), national reference costs (Department of Health 2007f) and national tariff (Department of Health 2006f).

Despite available guidance, the choice of theoretically correct methods is not simple. Recommendations differ and they lack detail about how to apply principles. Adam et al (2003a) reviewed costing guidelines. They found agreement on some aspects such as choice of comparator (best current alternative practice) and adoption of the societal perspective whilst allowing other viewpoints. For other aspects, the recommendations were not uniform, a finding also reported by Jacobs et al (2005) and Mogyorosy and Smith (2005). Adam et al (2003a) categorised the variation between guidelines according to whether there was disagreement, agreement in principle but no practical guidance, and agreement but studies did not follow the recommendations. Table 2.8 summarises their findings. Disagreements included whether and how to include future non-health care costs for unrelated illnesses, handling of productivity changes (the effect on work time due to morbidity and mortality), and costing informal caregiver or volunteer time.

Table 2.8 Review of costing guidance for economic evaluations

Aspect	Comment
Overheads	Costing methods and validation not fully discussed.
Shared costs including labour	Costing methods briefly discussed, validation not discussed, no recommendations.
Productivity (time for work, leisure etc.) - patients, informal carers, etc.	No agreement on inclusion (although agreement that volunteer time should be included unless minimal). Costing methods discussed but no recommendations made. No discussion of methods applicable in developing countries.
Capital costs	Alternative costing methods discussed (e.g. rental, or historical or replacement annualised costs), but no recommendations. Variation between guidelines in recommendations about discount rates and lifespan of items.
Health care costs for unrelated illness in additional years of life	No agreement on inclusion.
Data sources	Some sources described, validity and reliability not discussed.
Approach	Methods and issues for top-down or bottom-up approaches, and in particular the importance of measuring and reporting capacity utilisation, were not discussed.
Price adjustments for market distortions	General agreement on theoretical need to adjust prices. Some guidelines considered this not worth the effort in practice, and others gave little detail on methods.
Prices and charges	Agreement that resource valuation should be based on costs, with adjustments necessary when charges used. Adjustment methods discussed for the US.
Exchange rates	No discussion about how and when to report in foreign currencies[11].

Source: Adam et al (2003a)

The authors noted significant omissions including a lack of detail about how to allocate shared costs such as staff time or overheads to specific interventions. Guidance was inadequate on how to address capacity utilisation and on use of shadow prices (i.e. for goods or services where no market exists). The authors specifically recommended future research to validate different techniques to collect

[11] Official exchange rates tend to fluctuate. One alternative is to use purchasing power parities that incorporate both the currency conversion and equalise differences in price levels between countries. These are available from the OECD (Organisation for Economic Co-operation and Development), but may be more accurate if developed for specific health areas as shown for renal services by Wordsworth and Ludbrook (2005).

data and to compare relatively rapid, low-cost methods with more expensive 'gold standard' methods. They also advocated examination of the trade-offs between the level of accuracy and the costs of undertaking each method, an issue discussed further in the next section.

The review showed that even when guidelines agreed, applied studies did not always follow the recommendations. Most studies adopted a health provider perspective rather than the recommended societal one. Studies omitted some costs completely (e.g. costs of recruitment for screening studies), or partially (e.g. donated items such as drugs and equipment). Studies often used charges without adjustment or comment about actual costs. Overall many studies failed to adopt the recommended format of reporting resource use separately from unit costs. Together, these findings limit the comparability and transferability of the results, which in turn restricts the value of such studies to decision-makers.

Discrepancies between applied studies were inevitable given the variation in guidance. Hence, it was unsurprising that Adam et al (2003a) found differences in methods used to allocate staff time. They also noted poor reporting of costing methods and analyses. This has been a consistent finding in other reviews (Halliday and Darba 2003, Stone et al 2000, Graves et al 2002, Doshi et al 2006). A weakness to these reviews, given the word limits imposed by most journals, was that the reviewers did not appear to contact the papers' authors to distinguish between inadequate reporting and lack of methodological rigour. Nevertheless, as shown later (section 3.4.1.2), methods of resource use data collection are often poorly described even when there are 'no' reporting restrictions.

Implications of poor methods and reporting are serious, as shown in Table 2.9. Graves et al (2002) assessed the quality of costing in economic evaluations conducted alongside clinical trials (patient-level cost data). The authors concluded that reporting of costing methods was poor and the majority of papers only satisfied two of 12 criteria. They judged more than half the papers to have deficiencies indicating potential 'gross errors' on two criteria - description of methods for allocating i) the time of human resources and ii) other resources between patients. The authors concluded that

> "no amount of statistical analysis can compensate for poor quality cost data" (page 739).

Table 2.9 Implications of poor methods and reporting

Aspect	Potential consequences
Exclusion of appropriate costs	Bias - wrong decision
Inaccurate or invalid estimates	Bias - wrong decision
Imprecise estimates	Inability to make decision
Lack of information on costing methods and estimates (e.g. lack of disaggregation of data)	Generalisability issues - inappropriate to use results to inform practice.

Overall, there is a perception amongst health economists that costing is easy and there has been less research into costing in economic evaluations compared with outcomes or benefits (Johannesson et al 1996, Sculpher et al 2005). In contrast, Brazier et al (1999) demonstrated there is an extensive literature on methods to measure health status and utilities used to weight length of life in the quality-adjusted-life-years (QALYs). Moreover, there are numerous direct comparisons of such outcome measures. Even one of the newest QALY measures, SF-6D, has been compared many times with the UK mainstay EQ-5D (Conner-Spady and Suarez-Almazor 2003, Longworth and Bryan 2003, and Gerard et al 2004, Brazier et al 2004, Tsuchiya et al 2006), or the Canadian equivalent HUI3 (O'Brien et al 2003). Conversely, comparisons of methods to measure resource use, and subsequently costs, are less frequent. Examples include comparisons of

- patient self-report questionnaires and medical records (Roberts et al 1996, Kenney et al 2002, Mistry et al 2005),
- patient self-report interviews and medical records (Petrou et al 2002),
- medical resource use data collected from case report forms and patient medical files (Standaert et al 2002),
- and one comparison of top-down and bottom-up costing (Wordsworth et al 2005, see section 2.2.2.3).

Some aspects of costing have received more attention than others have. Due to the challenges posed by cost data, there is a growing body of research into statistical and analytical issues. This includes research into how to deal with skewed cost data (Manning and Mullahy 2001, Manning et al 2005). Cost data are zero or positive and typically, because a small number of people consume a large amount of resources, the data have a positively skewed distribution (i.e. with a long tail). Usually, statisticians would transform such data to produce a more normal distribution that can be analysed using parametric statistics. Yet in economic evaluations, the interest is in the overall population effect - the arithmetic mean - so transformation of data causes problems with the interpretation of results. Alternatively, analyses can use less powerful non-parametric statistics.

There has been considerable work on handling missing data and censoring due to dropout (Briggs et al 2003, Raikou and McGuire 2004, Oostenbrink and Al 2005). These aspects are complicated because, in contrast to outcomes data (typically a single outcome), the cost of an option is usually an aggregation of multiple resource use items of which one or more may be missing.

The handling of uncertainty requires special techniques. This is because costs often incorporate both stochastic data from samples and deterministic data (point estimates). Whereas uncertainty in stochastic data is examined using statistical techniques, investigation of uncertainty in deterministic data is through sensitivity analyses. These can re-analyse data using alternative scenarios or estimates, use non-parametric bootstrapping, or use probabilistic sensitivity analyses with statistical sampling after assuming a distribution for parameters (Briggs and Gray 1999).

Another area of interest is costing methods for multi-centre studies (Schulman et al 1998, Raikou et al 2000, Glick et al 2001, Thompson et al 2006). Issues include whether to combine centre-specific resource use with either centre-specific or average unit cost data. According to Glick et al (2001), there remain good theoretical reasons to use centre-specific costs and resource use, as application of average (pooled) cost estimates may both reduce or increase the variance. Concern about the centre vs. average issue arises because patterns of resource use and prices may vary between countries or centres, partly due to possible interaction between resource use and unit costs. For example, Wordsworth et al (2005) suggested likely substitution between peritoneal dialysis and haemodialysis, with the former used more in countries where dialysis consumables were costly relative to staffing costs, and *vice versa*. Other issues concern the transferability of results between countries and debate over methods to convert currencies (Manca et al 2005, Wordsworth and Ludbrook 2005, Grieve et al 2005).

A systematic review by Johnston et al (1999), funded by the Health Technology Assessment Programme, examined the literature on the collection and analysis of cost data alongside clinical trials. A major finding was that data collection issues were under researched and that, in general, researchers did not report or test the validity and reliability of their methods. The authors recommended further research into various aspects - study design, data collection, data analysis and presentation of results. Design issues included the effect of alternative criteria for attribution of costs. Topics on data collection were the validity and reliability of data collection tools for resource use, appropriate recall periods for data collection, and development of standard patient-cost questionnaires. A further recommendation was for research into alternative methods to estimate unit costs.

Mogyorosy and Smith (2005) conducted an important systematic review as part of the HealthBASKET project funded by the European Commission. This examined literature on theoretical and practical approaches to estimating service costs and aimed to identify best practice for international and regional cost comparisons. The main recommendation was research to enhance standardisation of costing methods and patient classifications (subgroup definitions). Other research recommendations echoed those of Adam et al (2003a), namely the validation of widely used methods by comparison with the 'gold standard' to develop valid, reliable and inexpensive methods to collect and value resource use data.

2.2.2 Top-down and bottom-up costing - guidance and research

This section compares the top-down and bottom-up costing approaches and synthesises advice on the choice between them. Then, it presents the findings of a systematic literature search for empirical comparisons of the two approaches.

2.2.2.1 Comparison of top-down and bottom-up costing approaches

The top-down and bottom-up costing approaches follow different principles and each has advantages and disadvantages, as shown in Table 2.10. The approaches are seen as opposite ends of a spectrum of precision or specificity (Drummond et al 2005, Brouwer et al 2001, Luce et al 1996). Drummond et al (2005) presented a continuum in order of decreasing precision from micro-costing of each care component, cost per case mix group (i.e. by patient type), average daily cost for disease-specific groups, to the least precise top-down costs for daily costs averaged across all patients. Nonetheless, it is acknowledged that a study may combine top-down methods for some cost components and bottom-up methods for others (Luce et al 1996, Brouwer et al 2001, and Mogyorosy and Smith 2005). The approaches may be used for resource measurement and estimation of unit costs or both, with further complexity if bottom-up resource measurements are combined with top-down unit costs. Therefore, in practice the distinction between the approaches is blurred. In published papers, it is often difficult to work which approach has been adopted overall or for specific cost items; if reported, the approach typically only refers to resource measurement.

Table 2.10 Comparison of top-down and bottom-up costing approaches

	Top-down (gross) costing	**Bottom-up (micro) costing**
Principle	Process of disaggregation i.e. relevant expenditure divided by units of activity.	Process of aggregation, i.e. summing: Component 1 x Unit cost1 Component 2 x Unit cost2 Component 3 x Unit cost3 ....
Breakdown of intervention into	Composite intermediate products, often large relative to total value of resources.	'All' components.
Example - Cost of an inpatient day on a hospital ward	Total expenditure on the ward (including overheads) divided by the number of inpatient days used.	Sum of values for each component of ward care, including, e.g. • Staff time (nursing, medical, support, admin etc.) • Drugs • Consumables • Capital (land, buildings and equipment costs - including maintenance costs) • Overheads.
Level of detail / precision	Broad estimates, less precision	Fine detail, more precision
Research effort	Quick and cheap	Complicated, much effort
Generalisability issues	Typically more generalisable	Easier to check comparability between centres. Information may be useful locally to understand activity-cost relationships.
Challenges in relation to the other method	Heavily influenced by data quality that may be poor when not collected for this purpose. Risk of omission if items paid by different budgets e.g. staff ± agency nurses, drugs prescribed by hospital or GP. Aggregated data hampers between-centre comparisons. Averaging across patients may conceal differences (patients, interventions, etc.).	Requires much data. Easy to forget 'overheads' e.g. slack time. Tendency to apply approach to resource use with little attention to variation in unit costs. Health Economists believe this method is the 'gold standard', but in reality, it requires many assumptions (discussed further in section 2.2.2).

Source: Includes information synthesised from literature cited in section 2.2.2.

In reality, top-down and bottom-up approaches both rely on averaging costs across activity, though at different levels of aggregation. Therefore, an allied issue is how to handle shared costs or overheads. Such costs are usually automatically included in top-down estimates, but easily overlooked in bottom-up estimates. Furthermore, Drummond et al (2005) argued that although economists favour marginal analysis that might suggest exclusion of fixed overheads, this was unjustified unless the costs were common to all options[12]. However, there is no single correct way to attribute 'overheads' to patient care, and traditional approaches typically allocate overheads in proportion to direct costs, floor area, head count, etc.

There has been growing interest is activity based costing (ABC) to enhance methods to allocate costs (Drummond et al 2005). This process involves collection of data on direct costs (labour and materials) and examination of 'products' to trace their demands on other costs inputs. Three principles guide which resources to focus on - expensive resources, those where consumption varies significantly by product, or those where demand is uncorrelated with traditional allocation methods such as direct costs (Cooper and Kaplan 1988). Some see ABC as an improvement on conventional methods by tracing overheads to the activities that caused the costs to be incurred (Kaplan 1990, Aird 1996, Laurila et al 2000). Conversely, Itami and Kaplan (1980) argued that despite its apparent logic, implementation of ABC is not straightforward in multi-product or multi-service industries. Furthermore, Goddard and Ooi (1998) noted that the data requirements for such systems make them costly to administer.

ABC illustrates two aspects of confusion in the top-down / bottom-up debate. First, Ellwood (1996, as cited by Mogyorosy and Smith 2005) argued that sophisticated, complex costing methods such as ABC might be no more accurate than simple systems, not least because some overheads may not behave linearly in relation to a cost driver. Second, there is debate about what defines the top-down and bottom-up approaches. Mogyorosy and Smith (2005) define ABC as bottom-up, involving direct measurement of patient-specific resource use. Yet Cooper and Kaplan (1988), early proponents of ABC, did not demand actual measurement, but stated that tracing costs to activities and hence products could "not be done with surgical precision" (p98) and was better "basically correct" than "precisely wrong". They illustrated

[12] This argument relates to whether variable costs that exclude fixed overheads reflect marginal costs, i.e. (TC + x) - TC, where TC is the total costs and x is one unit of activity.

implementation of ABC by interviews with relevant staff to gather activity estimates to attribute costs.

Having compared the basic principles, the next section reviews advice on the choice between the two approaches and empirical comparisons.

2.2.2.2 Recommendations on which costing approach to adopt

Important health economics texts include recommendations about the selection of costing methods (Drummond et al 2005, Brouwer et al 2001, Luce et al 1996). They acknowledge three influences in the choice of costing approach, namely the importance of precise estimates, feasibility, and research costs (issues discussed in section 2.1). Table 2.11 gives an overview of the recommendations. In principle, the bottom-up method is the preferred approach by Luce et al (1996) in the consensus statement from the Panel on Cost-effectiveness in Health and Medicine (Gold et al 1996). They argued that the bottom-up approach allows analysts to check the applicability of inputs in other situations, and is especially suitable where the cost input is integral to the analysis. Similarly, Mogyorosy and Smith (2005) advocated the bottom-up method because

> "it is more reliable, accurate and flexible than more macro approaches" (p2).

Conversely, Mellett et al (1993), argued for top-down costing as, from an accounting viewpoint, figures could be proved by reconciliation with planned and actual costs for audit purposes. Nevertheless, they acknowledged that top-down costs might be less accurate and lead to credibility problems with clinical staff. A consensus statement from the US Department of Veterans Affairs advocated a hybrid of top-down and bottom-up costing (Swindle et al 1999). This proposed bottom-up costing when the routine computer databases did not track resource use or when the data were insufficiently disaggregated. It also called for data validation to be built into studies when cost is a significant decision aspect.

Table 2.11 Recommendations on choice of top-down or bottom-up costing

Top-down	Bottom-up	Source
When more exact bottom-up costing has minimal impact on results	Results sensitive to cost component	Luce et al (1996), Drummond et al (2005)
Future resource use as discounting diminishes impact	Current interventions and events	Brouwer et al (2001), Luce et al (1996)
Homogeneous production	Heterogeneous production	Brouwer et al (2001), Swindle et al (1999)
Small component of overall cost (low cost item)	Major cost drivers (large component of overall cost, or small cost but large volume differences)	Drummond et al (2005), Gruen and Howarth (2005), Luce et al (1996)
Infrequent resource use (e.g. side effects or less predictable conditions)		Drummond et al (2005), Gruen and Howarth (2005)
	When top-down costs not valid (i.e. do not reflect actual costs [13])	Luce et al (1996)
'High' level of aggregation needed (e.g. clinic, facility)		Swindle et al (1999)
Primary data collection too expensive		Swindle et al (1999)
Where major resource use is for marketed (traded) technologies (e.g. drugs, medical devices, consumables)	Significant staff input or overheads, considerable sharing (staff or facilities) between interventions or patient groups, and where systems do not routinely cost to the specific intervention (or patient group).	Mogyorosy and Smith (2005), Wordsworth et al (2005)

Three assumptions are implicit in these recommendations. First, bottom-up costing can uncover 'better' estimates (i.e. more valid, accurate, precise, or reliable). Second, bottom-up costing is less reliant on assumptions, although Luce et al (1996) acknowledged that handling of capacity and occupancy issues may be important. Third, bottom-up costing is worth the extra research effort.

In theory, estimates from top-down and bottom-up approaches might reconcile. Mogyorosy and Smith (2005) noted the approaches might yield similar results for

[13] I.e. when top-down cost corresponds poorly with resource cost (e.g. top-down costs for reimbursement); when different protocols are in use since previously derived costs would obscure differences; and when costs of options are likely to diverge.

marketed items (e.g. drugs, medical devices and consumables). In reality, reconciliation is unlikely due to numerous variations in resource use components, activity levels, capacity assumptions, and unit costs, etc. The published recommendations advocate choosing an approach that is unbiased and sensitive enough to detect true differences between options. Hence, the decisive test is whether the costing approaches lead to different results or decisions.

2.2.2.3 Review of empirical comparisons of costing approaches

Whilst the health economics texts made recommendations about the choice of costing approach, they did not provide evidence to support them. Therefore, a systematic literature search was conducted to find empirical comparisons of the two approaches[14]. This found only one relevant paper, which compared the top-down and bottom-up costing approaches in haemodialysis and peritoneal dialysis in ten centres across eight countries (Wordsworth et al 2005). Interestingly, in three centres[15] information was inadequate or too complex to undertake both costing approaches.

For haemodialysis, the overall top-down cost was greater than the bottom-up cost (by €330 to €11,800) at five of the seven centres, and less (by €770 to €1120) at two centres. For peritoneal dialysis, the top-down cost was greater than bottom-up cost (by €90 to €1,700) at all five centres where provided. In each case, exclusion of overheads tended to decrease the size of the difference. Across treatments, haemodialysis was more costly than peritoneal dialysis at two centres and the differences were greater for top-down costs. In contrast, peritoneal dialysis was more costly at three centres; at one centre the difference was greater for top-down costs, whereas at the other two centres the difference was greater for the bottom-up

[14] Inclusion criteria - English language, costing comparisons of approaches within centre(s). Search terms in title, abstract or keyword (where applicable): (bottom-up or bottom up or micro) and (top-down or top down or gross) and (cost$ or resource$).
Databases: Allied and Complementary Medicine 1985-July 2007; British Nursing Index & Archive 1985-July 2007; Cumulative Index to Nursing & Allied Health Literature 1982-July 2007 Wk 3; EMBASE 1980-2007 Wk 30; Health Management Information Consortium July 2007; MEDLINE 1950-July 2007 Wk 3; MEDLINE Daily Update and In-Process 26/7/07; EconLit 1969-06/2007. Topic = (top-down or top down or gross) and (bottom-up or bottom up or micro) and costing in: Science Citation Index & SCI Expanded 1970-2/8/07.
Search yielded 292 references, only one was relevant.

[15] Tirana (Albania) and St Petersburg (Russia) - limited information, and Nijmegen (Netherlands) - "finance data were too complex to collect information independently for the approaches". The other 7 centres were: Dundee and Aberdeen, Nantes (France), Thessaloniki and Veria (Greece), Tallinn (Estonia) and Debrecen (Hungary).

costs. Overall, the absolute difference between haemodialysis and peritoneal dialysis varied from €3,000 to €16,000. The authors noted that estimates were closer between the two costing approaches for peritoneal dialysis than haemodialysis. The greatest differences occurred in centres integrated within a hospital where staff did not work exclusively with dialysis patients.

The study also examined the differences between the approaches for the four main cost categories (staff, consumables, capital and overheads) at one Scottish centre and the Estonian centre. Table 2.12 shows the findings. Whilst overheads were consistently less using the bottom-up approach, capital was more across both centres and treatments. For staff and consumables, the pattern varied across centres or treatments.

Table 2.12 Comparison of top-down and bottom-up estimates by cost category at two centres (Scotland and Estonia)

	Bottom-up estimate relative to top-down estimate	
	Less by*	**More by***
Staff	21% Scotland HD 59% Scotland & 78% Estonia PD	12% Estonia HD
Consumables	13% Scotland HD 1% Estonian PD	5% Estonian HD 24% Scotland PD
Capital		6% Scotland & 28% Estonia HD n/a PD
Overheads	76% Scotland & 16% Estonia HD 70% Scotland & 35% Estonia PD	
Total	25% Scotland HD 4% Scotland & 4% Estonia PD	8% Estonia HD

Key / notes: * as % of top-down estimate HD = haemodialysis PD = peritoneal dialysis
n/a = not applicable as top-down information on capital not available for PD
Centres: Scotland (Aberdeen), Estonia (Tallinn).
Source: Data from Wordsworth et al (2005)

The authors' comments principally related to which cost categories contributed most to the overall differences, rather than why costs varied. They did explain that for peritoneal dialysis at the Scottish centre, the bottom-up estimate for consumables included the cost of EPO, an expensive drug, whilst the top-down estimate did not because the drug was prescribed by GPs, not the hospital. The bottom-up estimate of staff costs used data collected from nursing and medical staff's estimates of their time use by a time allocation questionnaire for the week ahead. The cost estimate excluded time spent with patients not receiving dialysis care, but also excluded time unrelated to patient care, which suggests the costs excluded 'overhead' staff time. Conversely, it is unclear whether the top-down costs attempted to exclude care for patients not receiving dialysis. In summary, the paper showed the differences between costing approaches could be substantial, but it also left many unanswered questions.

Whilst the literature search revealed a paucity of studies comparing top-down and bottom-up costing approaches, a few studies were found that compared bottom-up costs and tariffs (i.e. charges), as shown in Table 2.13. Some adjusted the tariff to reflect the 'cost' using cost-to-charge ratios to allow for the 'mark-up' charged on resource expenditures. The studies found differences between the approaches, but not consistently one way or the other, and interpretation is not simple. Glick et al (2007) noted that since costs differ by centre, comparison of a local cost and national tariff might not be appropriate. Furthermore, the comparisons operated at different levels of aggregation. It was for similar reasons that Little (2005) cautioned against using national average costs alongside costs derived locally within an evaluation. Overall, Glick et al (2007) argued that there is only limited evidence about whether the accuracy of tariffs has affected the conclusions of economic evaluations.

Table 2.13 Examples of studies comparing bottom-up costs and tariffs

Author	Location	Comparison	Comment
Using tariffs adjusted to reflect cost			
Riewpaiboon et al (2007)	Thailand	Micro-costing compared with i) charges multiplied by cost-to-charge ratios ii) relative value units. Six surgical procedures.	Compared to micro-costing, unit costs based on relative value units ranged from 25% less to 15% more, and unit costs based on ratio of cost-to-charge ranged from 85% less to 32% more. In five of the six procedures the costs based on both relative value units and ratio of cost-to-charge were less than micro-costing.
Taira et al (2003)	US	Comparison of 4 methods: i) hospital charges, ii) hospital charges converted to costs by hospital-level cost-to-charge ratios, iii) hospital charges converted to costs by department-level cost-to-charge ratios, iv) bottom-up costing for cath lab costs plus nonprocedural costs by method iii. Percutaneous coronary revascularisation.	Cost estimates and cost differences between treatment groups varied considerably by method (e.g. charges were about twice the hospital costs). At patient-level, only 5% of costs from method 1 were within 10% of those by method 4 (compared with 34% and 22% of patient costs with methods 2 and 3 respectively). Between-group cost differences were only consistently within $500 of the reference standard for method 3. However, overall the costing method did not affect the main results of the analyses for any of the three clinical trials.
Schwartz et al (1995)	US	US charges multiplied by Federal cost-to-charge ratios and relative value units.	Concluded patient-level 'costs' inaccurate, but average 'costs' per DRG group usually within 10% of relative value estimates.
Using unadjusted tariffs			
Heerey et al (2002)	Ireland	Irish DRGs and micro-costing. Acute myocardial infarction, cardiac failure and HIV.	Differences ranged from -9 to 66%.
Nisenbaum et al (2000)	US	Medicare payments (professional and technical) and costs based on activity-based costing. Computed tomography procedures.	All 16 Medicare professional payments were less than the professional costs, whereas technical payments exceeded costs in 14 of the 16 codes (Physicians' Current Procedural Terminology). The Medicare global payment (professional and technical) exceeded the costs for 10 of the 16 codes (mean $33, range $3-75), and for six codes Medicare under-reimbursed (mean $57, range $3-160).
Beck et al (1999)	England	Trust prices and service-specific (bottom-up) costing. Treatment of HIV infected children.	Trust prices consistently lower than unit cost estimates (69-88%) and the disparity was worse as service intensity increased.

2.2.3 Classification of patient heterogeneity

Patient heterogeneity is relevant both in defining health care outputs and as a source of variation in resource use and hence costs. This section gives an overview of the classification of patient heterogeneity drawing on literature about 'case mix', Diagnosis Related Groups, Healthcare Resource Groups, and workload measurement. Later (section 4.5.3), the thesis examines the handling of patient heterogeneity in economic evaluations of haemodialysis in different settings.

2.2.3.1 'Case mix'

The concept of 'case mix' is complex because it encompasses inter-related but distinct patient attributes and different perspectives, as shown in Table 2.14. Unsurprisingly, given the number of different aspects to 'case mix', there are numerous classifications. For example, severity rating using the Acute Physiological and Chronic Health Evaluation (APACHE 1, 2, and 3) uses physiological variables to predict both resource use and prognosis of death (Bardsley et al 1989b). Other systems use treatment scoring according to the number and type of interventions needed (e.g. TISS - Therapeutic Intervention Scoring System - de Keizer et al 1998).

Table 2.14 'Case mix' from different perspectives

Clinicians	Administrators / Regulators
Severity of illness Risk of mortality or relative loss of function for patients with a particular disease (may including degree of co-morbidity).	**Resource intensity** Relative volume and type of services (diagnostic, treatment and beds) to manage particular illness.
Prognosis Probable outcome, likelihood of improvement or deterioration by severity of illness, likelihood of recurrence and probable life span.	
Treatment difficulty Need for sophisticated or technically difficult procedures, close monitoring or supervision.	
Need for intervention Relates to severity of illness that would occur without immediate or continuing care.	

Source: Adapted from Averill (1991)

There are several salient points about classifications. They have different purposes and some classifications (especially clinical ones) may not be good predictors of resource use from an economic viewpoint (Bardsley et al 1989b, Averill 1991). For example, a terminally ill patient is severely ill and has a poor prognosis, yet may only need basic nursing care. The distinction between classifications is often blurred as terms have overlapping meanings. In the nursing literature, Harrison (2004) defines acuity as the patient's level of illness or likelihood to deteriorate, but this encompasses severity of illness, prognosis, and need for intervention. Furthermore, classifications can focus on patients, staff or a combination of both. Hence, measures of resource intensity may combine patient dependency (the amount of care required from staff), use of treatments, and workload.

The NHS Information Authority 2003 defines 'casemix' as

> "classification of people or treatment episodes into groups, using characteristics associated with the condition, treatment or outcome that can be used to predict need, resource use or outcome" (p25).

Nevertheless, 'case mix' or 'casemix' have become synonymous with the Diagnosis Related Groups (DRGs) system used in the United States, Australia, and some European countries, and the equivalent Healthcare Resource Groups (HRGs) in England and Wales. These are examined in the next two sections. To avoid confusion with these systems the thesis tends to use the term 'patient heterogeneity'.

2.2.3.2 Diagnosis Related Groups (DRGs)

Research into developing a variety of casemix classifications based on so called 'iso-resource' groups began in the late 1960s in the US. From this, the Diagnosis Related Groups (DRGs) evolved to become widely used, as from 1983 Medicare[16] adopted them for prospective payment of patient care at fixed price per DRG (Fetter 1989).

DRGs set out to define hospitals' products as collections of patients who received similar outputs or services (e.g. pathology tests, and care). The aim was to use the classification to evaluate hospital performance and apply industrial control methods similar to those in manufacturing (Fetter 1989). Initial versions of DRGs only

[16] The US government administered health insurance for people who are elderly, disabled or with end stage renal disease.

covered inpatient care. They were based on a hierarchy of diagnosis codes subdivided by either surgical procures or medical diagnoses, and further subdivided by age, co-morbidities or complications and discharge status. Four principles guided development of DRGs:

1. based on information routinely collected,
2. a manageable number of classes (i.e. hundreds rather than thousands),
3. similarity of resource intensity within a class (i.e. resource use averages and variation known and predictable), and
4. clinical similarity of patients in a class, (Fetter 1991).

DRGs have undergone numerous refinements over the years and those used by Medicare were version 25 in October 2007, but other variants exist (e.g. All-Patient DRGs used by Medicaid[17] and other government and commercial payers). Revisions have incorporated changes both in diagnostic and procedure codes, and in medical practice based on studies of predictive validity (Freeman 1991).

Bardsley et al (1989a, 1989b) highlighted five statistical issues concerning DRGs. First, resource use is expected to vary between patients within a DRG. Second, the distribution of resource use within a DRG is usually positively skewed (i.e. mean > median, see A in Figure 2.1). Third, a DRG may be less appropriate if it has a wide resource use distribution, especially if this results from the combined distributions of two or more patient types (see B in Figure 2.1). This would indicate heterogeneity and possible candidates for separate DRGs. In contrast, they defined homogeneity within DRGs as

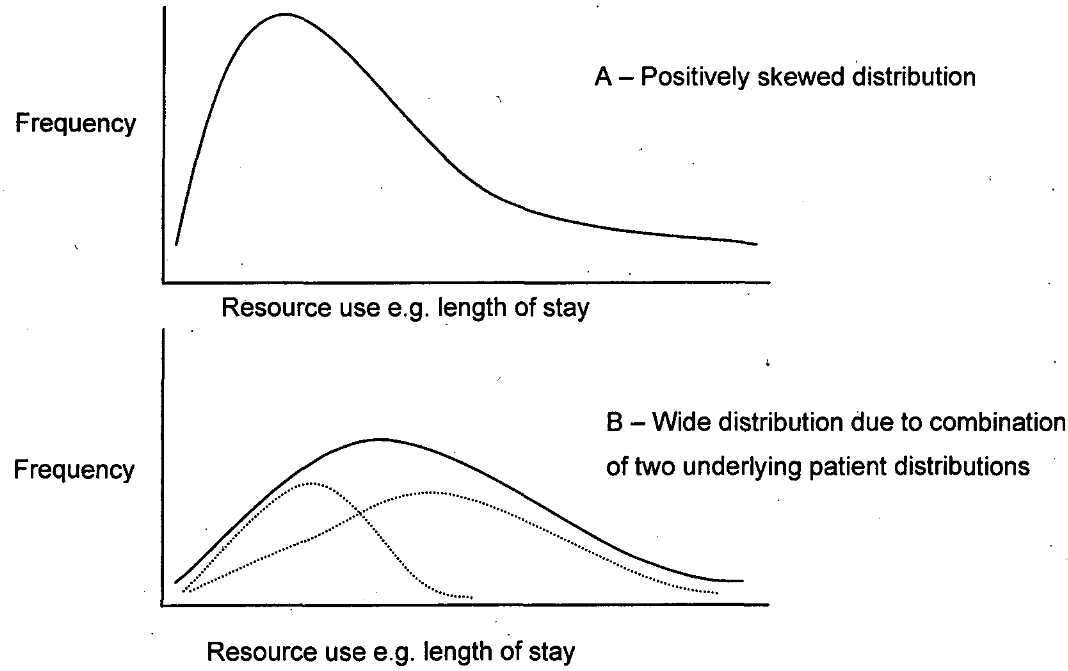
> "cases clustered around the mean and with few extreme cases or outliers. An acceptable level of homogeneity may be defined by a ratio of standard deviation to mean less than one." (p215)

Fourth, they advocated identification of patients with atypical resource use through trimming - the application of arbitrary statistical trim points to the tails of the distribution (e.g. mean ± a number of standard deviations). This allows review or removal of extreme cases and hence better description of the group's underlying characteristics. They suggested that removal of a few cases (e.g. 2-4%) might be acceptable if it greatly reduces the mean and standard deviation. Fifth, DRGs aim to be iso-resource, for example by length of stay, not iso-cost. This means that whilst DRGs aim to have statistically stable distributions of resource use, such stability is not expected when monetary values are attached (Coles et al 1989). Consequently,

[17] The US joint federal-state health insurance for people on low-incomes.

it is helpful to be able to assess the efficiency of resource use separately, without the extra variation introduced by valuing the resource use.

Figure 2.1 Distributions of resource use



Source: Bardsley et al (1989b)

In addition, Bardsley et al (1989b) argued that the cost aspect complicates diagnosis-based and iso-resource classifications. The classifications implicitly assume that patient outcomes and quality of care are constant between patients and hospitals. The authors noted there may be cost-quality trade-offs, additional to various factors that affect the relative costs such as wages and spare capacity for emergencies. Whilst this is a valid criticism, equally it could apply to most other classification systems.

2.2.3.3 Healthcare Resource Groups (HRGs) and reference costs

Healthcare Resource Groups (HRGs), the UK equivalent to DRGs, have become central to NHS costing, and more recently the basis for reimbursement of NHS health care providers. Sanderson (1989) reported that although UK research in the 1980s supported the overall validity of the DRGs, they did not perform as well statistically as in the US. Moreover, it was difficult to map between the diagnostic and procedure coding systems used in the two counties due to differences in medical practice and the organisation of care. For these reasons, the HRG system was developed using similar principles (i.e. standard groupings of clinically similar treatments that use comparable levels of health care resources).

Since 1998, the HRGs have been the basis for a national schedule of reference costs developed to assess NHS performance and tackle inefficiency. Dawson and Street (1998) noted that initially HRGs had multiple purposes: benchmarking costs, measuring relative efficiency, identifying best practice, remuneration, purchasing, internal management and central monitoring. They criticised the system for its lack of patient-level cost information and the discretion allowed in applying costing guidance. They also criticised the use of length of stay as a proxy for cost because it concealed differences in resource use by different patients, such as when delays in discharge arrangements rather than care needs led to longer stays.

HRGs and reference costs have evolved since their inception. There have been ongoing improvements to the HRG system. Version 4 is being phased in through use in reference costs from 2006/07 onwards. HRG4 involved major revision to increase coverage (from 650 to more than 1,400 groupings), to improve the use of complications and co-morbidities, and 'unbundle' high cost elements (e.g. drugs) to improve HRG performance and to make HRGs independent of the care setting. Furthermore, the previous version (HRG3.5) grouped patients by finished consultant episodes (FCEs)[18] and selected a dominant FCE if a patient received more than one intervention. In contrast, HRG4 uses spells[19]. The grouper software takes account of the FCEs, so patients with the same main procedure receive different HRGs based on their diagnoses.

Costing guidance has progressed with updates to the NHS costing manual (2003, 2005, 2007 and 2008) and specific guidance on reference costs collection each year. Whilst Mogyorosy and Smith (2005) refer to HRG based costs and Diagnostic Related Groups as top-down, this is changing. The current NHS Costing Manual (Department of Health 2008b), outlined in section 2.1.6.1, demands that health care providers use a top-down approach that bears many features of activity based costing. However, the Department of Health (2007f) already allows providers to use patient-level costing systems for HRG costing if the system has proven experience and can provide data at various levels of aggregation. Indeed, the NHS is being

[18] An episode of care under one consultant in one NHS Trust.
[19] Whole stays in hospital that may involve a number of FCEs.

encouraged to implement patient-level information and costing systems (Department of Health 2007b).

Since 2004, reference costs have been the basis for the national tariff and central to 'Payment by Results' (PbR), a reimbursement mechanism for NHS health care providers (Department of Health 2002b). PbR involves payment at fixed prices for activity undertaken, although there is some adjustment through a market forces factor for variations in 'case mix', regional wages and other costs (e.g. for land and buildings). PbR aims to avoid funding based on historic budgets or the negotiating skills of managers, and instead reward efficiency and support patient choice. The tariff's coverage has increased to include direct care services, clinical support services such as laboratory tests, follow-up and other care for inpatients, clinic visits, and accident and emergency services. Phased implementation of PbR started in 2004. Renal dialysis, along with some other services (e.g. community and mental health services, chemotherapy and radiotherapy) remains excluded from the mandatory tariff in 2008-2009 (Department of Health 2007e). The impact of migration from HRG3.5 to HRG4 will not be fully realised until the 2009 tariff. Furthermore, the Audit Commission (2008) found that the move to PbR has raised the profile of data quality for both NHS activity and costs, although major improvements are still required.

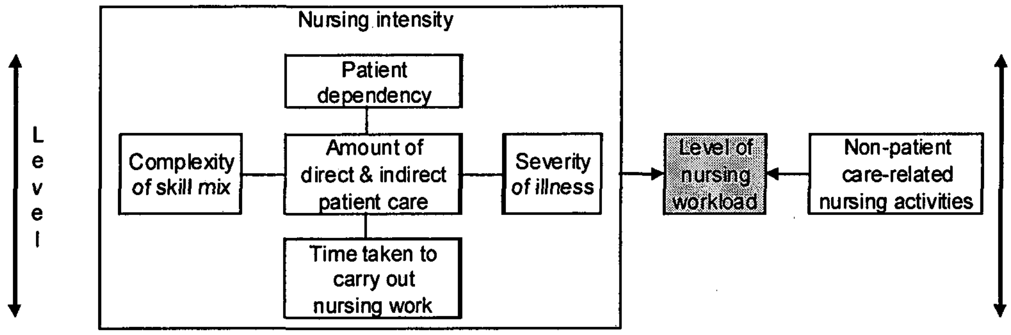
Despite being derived using accountancy principles, the accessibility of HRG-based reference costs means they are a useful source of unit costs for economic purposes. Conversely, given that HRGs typically cover multiple conditions and procedures, the reference costs are unsuitable for resource use at a disaggregate level within an HRG.

2.2.3.4 Patient heterogeneity in nursing inputs - nursing 'workload'

Patient heterogeneity can also be described by staff 'workload', although the concept is multifaceted and, given the focus of the thesis, this section focuses on nursing workload only.

Morris et al (2007) reasoned that although widely researched, nursing workload involves many different concepts that overlap but also differ subtly in their emphasis. For example, patient dependency includes the patient's care needs rather than the complexity of care required and skill mix etc. They noted the interchangeable use of terms, and variation in focus from purely patient-related to broad nursing activities. They proposed a model of nursing workload (shown in Figure 2.2) that acknowledged its multifaceted nature. Here, nursing intensity encompasses the amount of direct and indirect patient care activity and factors that affect the level of work required, namely time to deliver care, complexity of tasks or skill level (including salary), patient dependency and severity of illness. The overall impact on nursing workload is then a combination of nursing intensity and the non-patient care-related activities required during the nursing shift.

Figure 2.2 Model of nursing workload



Source: Morris et al (2007)

Workload measurement has many purposes. Examples include examination of ways of working, prediction of staff required in the long and short-term, grade mix (within a staff group e.g. nurses) or skill mix (across staff groups e.g. nurses and health care assistants), assessment of job satisfaction, and assessment of patient outcomes or cost implications of staffing establishments. For example, Rafferty et al (2007) found that higher nursing staffing levels were associated with better patient outcomes.

Hurst (2002) undertook a systematic review of literature on workforce planning methods on behalf of the Department of Health. The review found a vast literature on workload and staffing, and examined the five commonly used methods to estimate the size and mix of nursing teams. Each workforce planning method had advantages and disadvantages (summarised in Table 2.15). Hurst anticipated that users of the review would want to use triangulation (i.e. corroborate their results by using more than one method). He advocated that base data should be sourced from quality-assessed units to avoid possible confounding of workload measurement. However, an earlier paper by Balogh (1992) demonstrated that validation of tools to measure quality of care was notoriously difficult.

Table 2.15 Summary of commonly used workforce planning methods (adapted from Hurst 2002)

Approach	Description	Strengths	Weaknesses
Professional judgement	Expert judgement on appropriate size and mix of nursing team. Adjustment required for 'time-out' (paid and unpaid leave).	• Quick, simple, inexpensive. • Applicable to any speciality. • Easy to update. • Can deal with new or unmeasurable variables easily by simply agreeing change in nursing team.	• Hard to explain relationship between staffing and nursing quality. • Insensitive to changes in patient numbers or dependency. • Subjective.
Nurses per occupied bed	Uses average nurses per occupied bed for similar ward based on other study data. Adjustment required for 'time-out' (paid and unpaid leave).	• Simple. • Staffing and grade mix formulae derived empirically (from data collected routinely). • Data easily computerised for scenario analysis.	• Assumes staffing levels determined rationally. • Data quality important i.e. averages should be from wards that have met a quality standard. • Formulae insensitive to dependency changes. • Costly to update formulae. • Routine data collection (rather than specific) is more error prone. • Ward structures and processes masked.
Acuity-quality (dependency-activity-quality)	Requires measurement of i) average number of patients at each dependency and ii) amount of direct care time per day at each dependency (on quality assured wards). Time converted to ratios. The workload index (WLI) is the product of the ratios and average patient numbers at each dependency. Bed acuity is the WLI divided by bed occupancy. Adjustment required for 'ward overhead' time (indirect care and non-patient related activities), unpaid meal breaks and 'time-out'.	• Can use other study data or local values. • Can use bed occupancy based on patient whole-time equivalents rather than single time point estimates. • Easy to change ward variables (i.e. patient numbers and dependencies). • Can be used for scenario analysis e.g. on a daily basis. • Performance indicators easily derived.	• Complex. • Use of non-local data may be unpalatable to nurses. • Requires use of computer spreadsheets. • Problematic for staffing levels on small and/or low dependency units (i.e. suggests less than one nurse on duty). • Collection of necessary data may be expensive and time consuming. • Grade-mix proportions derived may be unsuitable locally. • Relationship between nursing activity and quality is complex and may be confounded.

Approach	Description	Strengths	Weaknesses
Timed task / activity	Initially requires recording of average time for activities. Subsequently, nursing hours derived from aggregation of times associated with activities on patients' care plans. Adjustment required for 'ward overhead' time (indirect care and non-patient related activities), unpaid meal breaks and 'time-out'.	• Easily computerised. • Commercial systems available (e.g. GRASP). • Base information easily updated. • Transferable to other care settings. • Useful where patients needs predictable (e.g. waiting list admissions)	• Increases work of care planning (each shift). • Commercial systems are expensive (especially initial implementation time and costs). • Work study approach may be unpalatable to nurses.
Regression-based	Initial statistical analysis of base data used to generate formulae to predict nursing numbers for activity level (from a number of independent variables).	• Useful where predictions possible. • 'Cheap' because data collection easy or aggregated from other sources. • Results corroborated by other methods and formulae judged valid and reliable. • Applicable across specialities.	• Initial number of variables for setting usually large and therefore needs input from statistician. • Interpretation and transferability of formulae (regression coefficients) problematic for qualitative / subjective independent variables (e.g. ward layout). • Assumes original data come from wards that operate efficiently. • Lack of understanding by nurses. • Unsafe to extrapolate beyond model's observed range (since linear relationship may no longer hold).

Within the workload literature, and of particular relevance to the thesis, there has been a sustained interest in how nurses spend their time. Hurst (2002) noted a decrease in direct care (face-to-face) over the previous 20 years and an associated questioning about the appropriateness of work undertaken by nurses. This was exemplified by Klein (2007) who called for exploration of nurses' deployment and how they apportion their time, in response to the findings by Rafferty et al (2007) about staffing levels and patient outcomes.

In terms of actual costing, a review by Sovie (1988) showed heterogeneity in nursing care and costs within DRGs in US hospitals. The commonest methods to allocate variable nursing costs to patients were nursing intensity or acuity classifications. She found inadequate reporting and many methodological inconsistencies between studies. Variations occurred in definition of similar terms, categorisation of nursing intensity, components and methods to assign costs (e.g. in some studies the nursing costs included other elements such as equipment and management costs).

2.3 Conclusions

This chapter addressed the first objective of the thesis through an overview of costing and review of literature. Health care has multiple purposes and outputs, and final patient-outputs are difficult to define. Furthermore, numerous combinations of resources can be used to produce multiple patient-outputs, posing a challenge when trying to establish the production function (resource use) and the cost function (value of resources).

The scope of the costing exercise (purpose, perspective and timeframe) is central to the choice of appropriate methods, which must balance data quality and feasibility issues. A difficult task is the identification of cost drivers so that research effort can be targeted efficiently. Whether from primary data collection or existing data (with or without adjustment to better reflect 'economic' costs), there are no perfect sources of resource use or unit cost data.

The overview of literature showed that there has been healthy scepticism of the benefit side of economic evaluations. Conversely, on the costing side, whilst there is a growing body of research into statistical aspects, there remain significant gaps in guidance and research on methods to collect resource use data and estimate unit costs. Research is needed into methods to attribute staff time that is shared across patients.

The review of guidance on costing for economic evaluations found the choice between the top-down and bottom-up approaches is not clear-cut in practice. In general, guidance seems to recommend the bottom-up approach, and implicitly assumes that it provides better quality or precise estimates, is less reliant on assumptions, and is worth the extra research effort. Nevertheless, there was minimal empirical evidence to support the recommendations or to assess the likely impact (i.e. whether the approach leads to different conclusions). Indeed, the literature search only found one empirical comparison, but the results were not consistent between centres or treatments.

Finally, the chapter gave an overview of methods to classify patient heterogeneity by various 'case mix' factors, DRG and HRG classifications, and workload tools. This showed that defining patient heterogeneity is complex, with many inter-related patient attributes and different perspectives. Nevertheless, these classifications have roles both in differentiating between patients and in defining the outputs of health care.

The next chapter addresses the second objective of the thesis - methods to measure staff time, especially at the patient-level.

Chapter 3 Measuring staff time

Staff time use has an opportunity cost. For example, a nurse undertaking care for one patient forgoes the opportunity to care for an alternative patient. Whilst the importance of time use applies to all other health care staff, the focus of the thesis is nurses' time.

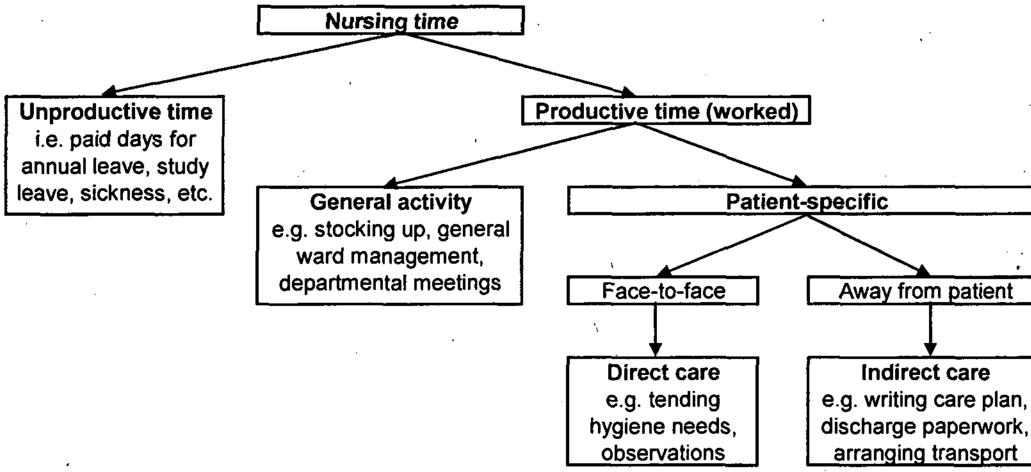
This chapter addresses the second objective of the thesis. It discusses the categorisation of staff activity, outlines methods to measure staff time, and compares their advantages and disadvantages. Lastly, it reviews literature on time measurement to examine use of the techniques in economic evaluations and health care more generally.

3.1 Types of staff time - categorisation of activity

It is necessary to categorise staff activity to measure staff time. Categories may be numerous and very specific (e.g. answering the phone), or more general (e.g. communication). Coding will become more onerous as the number of categories increases; it will also be more difficult to ensure consistency between coders. The focus may be on how staff use their time overall, or for specific tasks such as taking observations, patient hygiene, doing paperwork, attending meetings, etc. Alternatively, the focus may be at the patient-level to quantify the care received by individual patients.

Figure 3.1 shows a method to categorise nursing activity synthesised from general comments in the literature. Unproductive time is for annual leave, study leave, sickness, etc., that is covered by nurses' annual salaries, but not worked on the ward or unit. In the "Unit costs of health and social care" (Curtis 2007), this time equates to 52 days per year per nurse which is 20% of paid time. The remaining productive time is for shifts worked and comprises general activity and patient-specific time. General activity does not relate to any particular patient and includes clinically related time for staff handovers, stocking up, etc. and non-clinical time for staff meetings, paid meal breaks, etc. Patient-specific time comprises direct care for face-to-face contacts and indirect care for a specific patient but not face-to-face.

Figure 3.1 Approach to categorising nursing activity used for empirical work



Most nursing activities fall easily into the three categories of productive time. Other activities are less clear-cut and require decision rules, such as when staff multi-task. Some activities are difficult to define. Chatting to a patient may be 'therapeutic' direct care, or simply passing time and more appropriately coded as general activity. For other activities, the categories overlap. For example, a wound dressing comprises both direct care and indirect care when the nurse leaves the patient to dispose of soiled materials. Moreover, at the patient-level, it may be difficult for observers to link indirect care activities to the correct patient.

Categorisation of activity has direct implications for costing, which are discussed in section 3.4.1.1. Furthermore, to derive the cost per patient, staff time (and hence costs) must be allocated to categories of patient. To some degree, patients can be classified according to the hospital, ward or unit where they are treated. At this level, an average cost per patient may be available from routine NHS costs or easily calculated from the overall staff and grade mix. Within or across these locations, patients may be divided into smaller groups, for example using a classification of patient heterogeneity like those discussed in section 2.2.3. Within each patient group, the nursing times per patient will vary. Whilst some elements of nursing time may be relatively fixed or similar for each patient, other more variable elements depend on patients' or other characteristics. Having briefly examined the categorisation of staff activity and its implications, the next section examines techniques to measure staff time.

3.2 Methods to measure staff time

Methods to measure staff time are adaptations of the data collection methods outlined earlier (Table 2.4). They include traditional observer methods, 'motion and time studies', developed in the 1880s to measure productivity in manufacturing (Meyers and Stewart 2002), and work sampling that developed later (Hansen 1960). Other methods are self-recording, self-reporting and expert opinion. In addition, technology such as barcode scanning may be used to enhance data collection either for self-reporting or observer methods.

Each method has advantages and disadvantages. Alteration of staff behaviour, the Hawthorne effect, is an inevitable risk for observer or self-recording methods. Likewise, recall bias is a risk for self-reporting and expert opinion. Moreover, depending on the technique, data may be collected at the individual level (nurse or patient) or at an aggregate level (already averaged across a patient group).

3.2.1 Time and motion study (stopwatch study)

Time and motion or stopwatch time study was the first technique to measure staff time. There are two principal approaches. Continuous time measurement records from the start of the first activity until the end of the study and 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 zeros the watch before starting to record the next activity. The continuous technique, which accounts for the whole period, cannot conceal any elements and according to Meyers and Stewart (2002) is the preferred technique of trade unions. Time and motion studies use a variety of equipment from special single or multiple watches (mechanical or digital), computers, videos and tachometers. Traditionally, time measurement uses decimal minutes or hours. An advantage of time and motion study is that it allows very detailed recording of activities, though typically on a small sample of staff (Finkler et al 1993).

3.2.2 Work sampling

Work sampling also originated in production engineering (Hansen 1960), but has been used in health care settings (Sittig 1993, Oddone et al 1995, Pelletier and Duffield 2003). Work sampling applies the principle used in Gallup polls and other surveys based on random sampling and probability theory. Observations are made to determine how staff are spending their time at sampled times. Table 3.1 illustrates the calculation of times using data collected in the empirical study (see chapter 6). Observations comprised direct care (face-to-face) for 20 individually identified patients and 'other activity' for all remaining activity. The ratio is the percentages of total observations in each category. Ratios, when multiplied by the total hours worked by staff over the period, give the time for each category. The table shows each patient's direct care time and the remaining time for all other activity.

Table 3.1 Example of work sampling to estimate patient times

Activity	Patient number	Observations		Time (mins)
		Number	Ratio (%)*	
Direct care	1	10	1%	30
	2	17	2%	51
	3	22	3%	66
	4	14	2%	42
	... to			
	20	12	2%	36
Total direct care		255	35%	765 (13 hours)
Other activity		525	65%	1435 (24 hours)
Overall total		**733**		**2200 (37 hours)****

Key / Notes:
* Ratio is each patient's percentage of total observations.
** Hours worked were 37 hours for the six nurses observed (for period 7:00-15:00 at 20 observations per hour).
Source: Data from empirical study (see Chapter 6).

Work sampling starts by estimating the sample size of observations for a required level of confidence and precision. The category forming the smallest proportion of overall activity drives the sample size and affects the feasibility of the technique. This is illustrated in the Table 3.2. Here, direct care for an individual patient is assumed to take 2% of staff time. With greater accuracy (and a narrower band of detection around 2%), the observations required escalate rapidly. The observer time required shows the time necessary for recording, which depends on the frequency of sampling. As an activity becomes a greater proportion of overall activity, so the observations and time required reduce. The sample size is important, especially if comparing techniques. Finkler et al (1993) criticised work sampling when they found large disparities compared with estimates from time and motion study. They generated a work sampling dataset from their time and motion data. However, since they only sampled at four observations per hour, the dataset was vastly underpowered.

Table 3.2 Observations needed for 2% expected time with 95% confidence

Accuracy	Detection (%)	Observations required	Observer time required*			
			15 observations/hour		20 observations/hour	
			Hours	Days**	Hours	Days**
1%	1.98 - 2.01	1,882,384	20,915	2,789	15,687	2,092
5%	1.9 - 2.1	75,295	837	112	627	84
10%	1.8 - 2.2	18,824	209	28	157	21

Key and notes: * *Observation of 6 nurses at each time point* ** *Assuming 7.5 hour-day*
Calculated using the formula from Meyers and Stewart (2002)
$N = Z^2(1-P) / A^2P$ where N = total number of observations needed
Z = number of standard deviations required (Z = 1.96 corresponds to 95% CI)
P = expected % of time for smallest element A = accuracy required (SD of %), usually 5%.

Having estimated the number of observations needed, the next step is to produce a schedule for trips through the unit. Typically, start times for each trip are selected at random. Alternatively, de Keizer et al (1998) used observations at 10-minute intervals. Sittig (1993) proposed that fixed sampling was acceptable if work activities were random and suggested this was the case for most health care activities. Conversely, fixed sampling introduces systematic bias by under or over sampling activities occurring at regular or specific times. Traditionally an observer logs activity, although Finkler et al (1993) noted that workers themselves might log activity, but considered this less reliable.

Work sampling has a number of limitations. Usually the method is considered unsuitable to study single members of staff or staff located over wide areas (Barnes 1958). In addition, Lee et al (2003) argued that work sampling is not valid or reliable in health care settings because many tasks are performed infrequently. Finkler et al (1993) concluded that whilst time and motion study offers depth, work sampling offers breadth through limited recording of activities over a large sample of staff.

3.2.3 Self-recording

Self-recording (or self-registration) simply requires each member of staff to record their activities. Traditional time and motion study texts do not mention self-recording (Meyers and Stewart 2002, Barnes 1958). This is presumably because such logging interrupts work and is unacceptable during production processes. However, the health care setting is very different from manufacturing and, as shown later (section 3.4), self-recording is relatively widely used. Self-recording allows simultaneous recording by multiple staff, and for individual patient-level activities that would be difficult for an observer to code.

3.2.4 Opinion-based methods - self-reporting and expert opinion

Opinion-based methods offer an alternative to direct time measurement. They include retrospective self-reporting[20], or expert opinion. The latter draws on either personal experience or application of historical data. So-called Predetermined Time Standards comprise the aggregated times assigned to tasks from tables of previously measured reference values (Meyers and Stewart 2002). These may be used directly, adapted from similar tasks, or aggregated from components. In health care, an example is the GRASP system that originated in the US to assist in workload planning (Meyer 1984).

Texts on motion and time study do not appear to cover techniques for opinion-based methods. Data collection may be informal or through methods such as diaries, questionnaires and interviews. Indeed, beyond health care, there is a growing body of survey research into how people use their time using these methods singly or in combination. Examples include the "American Time Use Survey" (ATUS) sponsored by the Bureau of Labor Statistics (BLS) and conducted by the US Census Bureau (http://www.bls.gov/tus), the "Multinational Time Use Study" (http://www.timeuse.org/mtus), and the Office of National Statistics' "National Survey of Time Use and Omnibus Survey" (Lader et al 2006). In some cases, these data have been used to verify data from other sources. For example, Frazis and Stewart (2004) found that, on average, time estimates for actual hours worked from the ATUS and from the BLS "Current Population Survey" (CPS) in the US were similar during the CPS reference week but less accurate for the rest of the month. Whilst such comparisons are useful to help validate data, one would not expect exactly the same results as questions and reference periods vary slightly. These comparisons highlight the problem of there being no 'gold standard' method to measure time.

The survey data above involve large population samples. In contrast, estimates of staff time in health care are more likely to come from relatively small samples and

[20] The thesis considers self-reporting (retrospective) separately from prospective self-recording.

handling divergent estimates poses a potential challenge. As an alternative, opinions may be sought using consensus methods such as the Delphi approach (Mullen 2003).

The Delphi approach is a family of techniques, rather than a single procedure. It typically involves development of an initial questionnaire distributed to an expert panel. The panellists answer the questions, generate ideas, and return the questionnaire to an independent co-ordinator. The co-ordinator summarises the responses and produces feedback along with a second questionnaire. The panellists evaluate their earlier responses and vote on the second questionnaire. The co-ordinator develops the final summary and feedback for the group. Two aspects are crucial. First, the panellists have an opportunity to revise their judgements based on feedback. Second, panellists receive some degree of anonymity. There are many variations to the basic technique. These include the number of rounds, method of selection and size of the panel, scoring system and the rules used to aggregate the judgements, extent of anonymity, and definition of consensus when there is disagreement (Mullen 2000, Mullen 2003).

Compared with direct measurement techniques, there are a number of disadvantages of opinion-based methods. Data collection by questionnaire or interview is prone to recall and response biases (described in Table 2.4). They assume the overall time is simply a function of the frequency of activities and a standard time, but may overestimate time if tasks overlap, and 'overhead' time is easily overlooked. Previous time standards may conceal past inefficiencies, or alternatively, staff may report ideal rather than actual times. Lastly, there are transferability issues because available data may not be relevant to the patients or setting of interest.

3.2.5 Technology to help data collection - barcode scanning

Barcode scanning technology offers a way to enhance data collection, as a scanner captures coded information along with the date and time of each scan. Numerous types of barcodes and scanners exist (Appendix 1 describes the technical details). Barcodes have a wide range of uses - for identification, inventory, tracking, and point of sales - in the food industry, retail, manufacturing, warehousing, distribution and shipping. Increasingly barcodes are being used in health care. Indeed, the Department of Health (2007a) is promoting the use barcoding and similar technologies to increase patient safety and improve efficiency. Patient safety examples include using coding systems to match patients to their care to reduce medication errors and risk of wrong site surgery. Efficiency examples include the tracking of equipment, better record keeping, and electronic management of supplies and purchasing to cut costs. Other health care applications include laboratory management of specimens, medical records management, data collection for management purposes and, in a few hospitals, real-time tracking of patients (e.g. in theatres).

Compared with hand-written logs for self-reporting or observer time measurement, barcode scanning appears to offer many advantages. Data do not require hand-entry and automatic data processing is possible. In addition, it is difficult to falsify scanned data as the recorder can delete scans but cannot backdate them. Studies, discussed in section 3.4.2, have used barcode scanners for staff to self-rerecord their time. In addition, observers used barcode scanning to collect data on staff time for patients grouped by a case mix classification in a study by Eastham (2006). Having introduced each technique, the next section compares the trade-offs between them.

3.3 Considerations about choice of time measurement technique

The health care setting is very different from a production line and there is no 'gold standard' technique to measure staff time. Table 3.3 shows how the techniques vary on key aspects. Here, self-recording is presented using barcode scanning rather than paper logging. The table also includes the 'usual practice' of averaging resource use and costs across patients that is common in economic evaluations. Apart from the advantages of prospective over retrospective measurement, other relative advantages depend largely on whether the study focuses on staff activity, specific tasks, or patients. Theoretically, it should be possible to collect data on direct care (face-to-face contacts) for individual patients using any of the methods. However, it is likely to be challenging for an observer, especially at a distance, and opinion-based methods can only guesstimate direct care. Indirect care for individual patients is difficult to capture using observer techniques because it will not be obvious for which patient an activity applies.

Table 3.3 Comparison of methods to measure staff time

Characteristic	Time and motion (TAM)	Work sampling	Self-recording (barcode scanning)	Self-report or expert opinion	Usual practice (assumptions)
Method	Continual one-to-one observation ('close' proximity)	Random observation one observer, multiple subjects (at a distance)	Self-logging (multiple subjects)	Self-report or expert opinion (apportionment)	Researcher (apportionment)
Data collection	Prospective	Prospective	Prospective / Retrospective	Retrospective / guess (about self or others)	Retrospective
Staff involvement[21]	-	-	+++	++	+/-
Data collector	Observer(s)	Observer(s)	All staff	Multiple 'experts'	Researcher
Training required	++	+	+	+	-
Subjects	One-to-one	Multiple (simultaneously)	Multiple	N/A	N/A
Equipment	Special stopwatch / computer / video	Random number tables, watch, pen and paper	Barcode scanners and activity lists, computer, specialist software	Questionnaire / interview	-
Main cause of bias	Observation	Observation	Omissions (accidental or conscious)	Questionnaire / interview	Assumption
Main biases	Altering subject behaviour (Hawthorne effect). Observer omission.	Altering subject behaviour (Hawthorne effect, less than TAM). Observer omission.	Altering subject behaviour (Hawthorne effect, less than TAM)	E.g. recall, response and non-response[22], sample composition, question framing and ordering effects, interviewer effects etc.	Researcher bias

[21] Beyond agreeing activities to code.

[22] Response bias e.g. social acquiescence where respondents give the answer they think will be socially desirable. Non-response bias occurs when respondents and non-respondents differ systematically.

Characteristic	Time and motion (TAM)	Work sampling	Self-recording (barcode scanning)	Self-report or expert opinion	Usual practice (assumptions)
Measurement level					
Specific tasks	Yes	Yes	Yes	Guesstimate	Guesstimate
Direct care	Yes and pt-level	Yes and pt-level	Yes and pt-level	Guesstimate	Guesstimate
Indirect care	Difficult at pt-level	Difficult at pt-level	Yes	Guesstimate	Guesstimate
General activity	Yes	Yes	Yes	Guesstimate	Guesstimate
Time to be apportioned to patients	IC and GA	IC and GA	GA	Total shift hours (all DC, IC and GA)	Total shift hours (all DC, IC and GA)
Availability of data by staff grade	Yes	By group (e.g. RGN / HCA)	Yes	Possibly (more likely by group e.g. RGN / HCA)	No, unless assumption
Research costs					
Research staff	++++	++	++	+	-
Equipment	+	+	(+++)	+	+
Signed consent likely to be required	Staff (individual)	Staff (at unit level)	Staff (individual)	Staff (individual) or implied consent if return questionnaire	None
Other comments	Very time consuming and costly research	Potentially more efficient than TAM especially for long duration of observation and minimises effect of workload fluctuations. May be impractical if staff geographically dispersed.	Interferes to some degree with staff's daily work		

Key / Notes: DC = Direct care i.e. face-to-face IC = Indirect care i.e. for specific patient but not face-to-face GA = General activity i.e. not for specific patient
pt-level means possible to collect time data for individual patients
Source: Includes information synthesised from sources cited in section 3.2, and consideration of implications for data collection at the individual patient level.

Table 3.3 also outlines feasibility aspects in relation to staff involvement and research costs. In a given research setting, the following inter-related issues are also important considerations:

1. Nature of the staff-patient contact - whether face-to-face, by telephone or other (e.g. video conferencing, postal).
2. Degree of geographical dispersal of the relevant staff, patients, or both, which affects the ease of tracking them.
3. Targeting staff or patients for data collection. Targeting patients may be more efficient if their care involves multiple staff. This would capture all the patients' direct care (i.e. staff-patient contacts). It would miss their indirect care (that is not face-to-face) and would not provide information about overall staff time to use in calculating the unit cost of an hour of patient-related time.
4. Practicalities of data collection due to variability of activities of interest. Data collection will be more efficient where the activities are a major part of the workload since little data will be redundant. Likewise, activities may occur constantly or sporadically and capture of frequent activities of short duration will be more challenging and onerous.
5. Acceptability to staff and patients. These will influence the degree of co-operation, success of data collection and credibility of the results. Any examination of working practices, especially using observation techniques, is likely to cause suspicion amongst staff. Use of techniques such as video recording may require permission from patients.

Validity is a further consideration in the choice of technique. A valid technique is one that measures what it purports to measure. There is no single test of validity; it can be considered in several ways, as shown in Table 3.4. This follows the approach by Brazier and Deverill (1999) who adapted criteria used in the psychometric literature to assess the performance of measurement instruments. Validity depends on the study context, as the methods differ in their coverage of different care aspects and richness of data. The technique must fit the study's purpose, which relates to content validity. For example, if patient-specific times are required, the researcher may decide to rule out observer methods because of their limited ability to collect indirect care time. Alternatively, the researcher may decide that direct care time is a good enough proxy for overall patient-specific time and opt for an observational method that does not disrupt the staff's work. Additional aspects that affect validity include the completeness of data collection and representativeness of data for the population of interest (i.e. external validity).

Table 3.4 Validity of time measurement techniques

Assessment	Comments about time measurement methods
Face validity	Appears sensible and appropriate to capture what it intends to measure - all the methods have face validity to some degree.
Content validity	Covers dimensions of interest, is relevant to the study population and potentially sensitive to important changes. Coverage of different care aspects: • Time and motion study and work sampling: At patient-level, principally only direct care. • Self-recording: All aspects. • Self-reporting or expert opinion and usual practice assumptions: Theoretically all aspects, but unmeasured.
Criterion validity	Correlates with existing 'gold standard' or accepted measure, but no 'gold standard' is available for time measurement.
Construct or empirical validity	Able to detect or correlates with known or expected differences (hypothesis testing). E.g. in this thesis, hypothesis that time will be positively correlated with patient dependency.

Source: Using descriptions adapted from Fitzpatrick et al (1998) and Brazier and Deverill (1999).

Reliability, the consistency of measurement over time and between raters, is also important but difficult to assess for time measurement techniques. It can be argued that test-retest reliability is rarely relevant because in health care, conditions and especially staff-patient interactions will not be identical on both occasions. Inter-rater reliability assesses the consistency between raters of the same activity. Direct inter-rater reliability checks (i.e. using the same method) are possible for time and motion study or work sampling by two observers working simultaneously. Such checks for self-recording and opinion-based methods are complicated because the main rater is the subject, and so they require an alternative method that may not cover the same aspects. For example, self-recording can only be assessed by opinion or observation, but an observer cannot differentiate all indirect care at the patient-level. Moreover, close shadowing by an independent rater may influence both the activities carried out and their recording by the main rater (i.e. the Hawthorne effect) and so lead to spuriously good (or bad) reliability results.

Finally, like the sources of variation in unit costs (Table 2.5), Table 3.5 illustrates influences on the value of staff time, although it is conceivable that some factors may act to limit variability in time. One would expect nurses to use their time according to patients' needs. When less nursing hours are available, one would still expect the neediest patients to receive relatively greater input but, in absolute terms, the time per patient and variability between patients would be less. Furthermore, the overall variability in time may be reduced if large elements of care are relatively uniform across patients. Indeed, in each study context, it is necessary to weigh up the value of the data gathered in relation to the likely research effort required.

Table 3.5 Potential factors influencing variation in staff time and costs

Aspect	Description
Patient factors	Patient characteristics and 'case mix' (e.g. need or dependency, illness severity, etc.).
Capacity issues	Staff time available, occupancy rates and throughput. Typically (unless very flexible) staffing is a stepped function increasing in increments when an additional member of staff is needed.
Clinical practice / service delivery variations	Skills or experience, alternative treatment strategies, substitution effects (e.g. for different types of nurse), service quality.
Efficiency	Maximising outputs or minimising resource use.
Input prices	Amounts paid for staff, e.g. grade and pay point.

Source: Adapted from Beecham et al (1993)

3.4 Time measurement literature

Having compared the time measurement techniques, this section examines literature to assess their use both in costing and economic evaluations, and for patient-level measurement in other health care studies.

3.4.1 Measurement of staff time in costing and economic evaluations

Hughes (1991) noted that salaries are used to represent the economic cost of staff time. One source of costs for numerous health care staff, which is widely used by health economists, is the "Unit costs of health and social care" (Curtis 2007). This is reviewed, in particular for data on staff time use. Then the section examines the use of time measurement to cost staff inputs in economic evaluations.

3.4.1.1 Staff costs in the "Unit costs of health and social care"

The Department of Health funds the annual production of the "Unit costs of health and social care" (Curtis 2007). Schedules are based on national salary scales. In an earlier version, Netten et al (1999) stated that for the most part the costs were bottom-up estimates.

In the 2007 version (Curtis 2007), schemas for various health and social care professionals included cost information on the following:

- salary (for the assumed grade and salary point)
- salary oncosts (contributions by the employer for National Insurance and superannuation)
- qualifications (an equivalent annual cost to cover pre- and post-registration education)
- overheads
- capital overheads (for new buildings and land of NHS facilities)
- travel (if applicable)
- working time per year
- ratio of direct (face-to-face contacts) to other "indirect" time for "other clinical" and "non-clinical time"
- duration of contacts (e.g. at clinic or home, if applicable)
- London and non-London multipliers (for working inside or outside London).

Costs were given at various levels of aggregation from a cost "per hour" to the cost per hour for "patient-related" time and, where relevant, for various types of contact (e.g. consultation, clinic, phone, home). Schemas showed two sets of figures; one set included salary oncosts, overheads and capital overheads, the other set also included the additional costs for qualifications. Data were disaggregated so that users could develop their own estimates for salary elements alone.

Schemas presented the ratio of direct time (face-to-face contacts) to indirect time (the combination of other clinical activity and non-clinical activity) and Table 3.6 illustrates how activity varied across nursing grades. The categorisation differs from the one in Figure 3.1, where indirect care referred to clinical time for a specific patient (but not face-to-face) and general activity encompassed clinical and non-clinical activity for no specific patient.

Table 3.6 Proportions of working time for hospital nurses by activity

Nurse	Direct care*	Clinical	Non-clinical
Team manager or team leader (i.e. sister or senior staff nurse)	45%	35%	20%
24-hour ward	50%	40%	10%
Day ward	55%	25%	20%
Clinical support worker (i.e. HCA)	60%	15%	25%

Key / notes: * face-to-face Estimates from consultation with NHS Trusts
Source: Curtis (2007)

Over the years, the reports have repeatedly advised users that staff time use can have a major impact on the unit costs of staff (Netten 1996, Netten 2002, Netten 2003, Netten and Curtis 2005). In the 2003 edition, Patel et al (2003) produced costs for three intermediate care schemes[23] using a staff-completed 7-day event record of all patient-related activity. The authors made the following assumptions that,

- unrecorded time was not patient-specific,
- staff completed the records accurately (no validity checks were undertaken), and
- the 7-day period reflected usual working practices.

The study found major differences both in the composition of the teams (type and numbers of staff) and in the pattern of staff activities recorded. These translated into differences in the unit costs of an hour of face-to-face contact, with costs ranging from half to four times for different staff. The authors concluded that many differences were due to schemes' historical evolution, but these affected unit costs and hence the ability to compare relative cost-effectiveness.

Patel's study illustrated how the variation in the proportion of time for activities affected the cost. Yet data on staff time use can be used in different ways to derive unit costs. A first step is to decide whether to exclude some time components. For example, Kernick and Netten (2002) excluded study leave and travel time, and then attributed indirect care and other time in proportion to direct care to cost a GP consultation. Indeed, Table 3.7 shows how options for calculating the unit cost of nursing time vary according to how they attribute salary costs to patient time. Whether top-down or bottom-up, the cost per hour worked attributes 'unproductive' time to worked hours (actual or expected). Bottom-up methods to derive hourly rates

[23] In this case, a combination of supported discharge from hospital (after acute illness or surgery) and rapid response to avoid acute hospital admission.

for patients attribute time in proportion to direct care ± indirect care. The costs per hour progressively increase in magnitude (i.e. patient-contact > paid).

Table 3.7 Options in calculating the unit cost of nursing time

Unit	Description	Salary costs allocated *in proportion to*
Top-down options		
Expenditure per hour worked or per outcome	Total expenditure on nursing salaries divided by total hours worked. Automatically includes salary oncosts, overtime, agency staff, and salary enhancements (e.g. unsocial hours, London weighting, etc.) averaged across all grades.	Actual hours worked or allocated to other outcome such as bed-day, session or visit.
Bottom-up options		
"Unit costs of health and social care" (Curtis 2007)		
"per hour"	Annual salary divided by expected working hours per year (1560 hours for RGNs, 1597 hours for HCAs)[a].	Expected hours worked
"per hour of patient contact"	Rate per hour worked (expected) divided by the proportion of direct care.	Direct care
Per outcome	Using activity data and the hourly rates above, e.g. to calculate cost per: hour in surgery, consultation, home visit, etc.[b]	Direct care
Alternative bottom-up options		
Per hour paid (i.e. basic pay)	Annual salary divided by the hours paid per year (i.e. 1955 hours = 365/7 x 37.5). This is the basic hourly rate used to pay staff for working unsocial hours and overtime.	n/a
Per hour worked	Annual salary divided by actual or expected working hours per year[c].	Actual or expected hours worked
Per hour of patient-specific time	Rate per hour worked divided by the proportion of patient-specific time (direct and indirect care[d]).	Patient-specific care (direct & indirect care)
Per hour of patient-contact	Rate per hour worked divided by the proportion of direct care.	Direct care

Key / notes: n/a not applicable
[a] Excludes 29 days annual leave, 8 days statutory leave (i.e. bank holidays), 10 days sickness, and 5 training days for RGNs, and assumes a 37.5 hour working week.
[b] All non-contact time allocated to contact time (travel only allocated to home visits).
[c] As for note a, but also includes other absences e.g. maternity leave, compassionate leave, and could include 'management' time that is unrelated to running the particular ward or unit.
[d] Indirect care is for specific patients, but not face-to-face.

Consequently, although the resource use is unchanged, the value of differences in nursing time between patients varies simply by applying one of the three hourly rates. From Table 3.6, about 50% of nurses' working hours are spent on direct care. Therefore, in attributing nurse time to the cost per hour of patient-contact, this cost will be about double the cost per hour worked. For some outputs, such as outpatient visits or treatment sessions, the top-down approach attributes all nursing costs equally per patient. For these kinds of output, it could be argued that the bottom-up cost should share the elements of 'overhead' time (i.e. unproductive time and general activity) equally across patients, rather than attribute them in relation to patients' care time.

There was no advice in the "Unit costs of health and social care" about which hourly rate to apply, and it is unclear from an economics viewpoint which is correct to value the opportunity cost of staff time. Given the lack of guidance in these key reports, it is unsurprising that researchers have adopted various approaches. Waller (1999) used the cost per hour worked to cost GP and practice nurse consultations. Shepherd et al (2007), used an unspecified hourly rate that appeared to be the cost per hour paid to cost the nursing time for an outpatient clinic in an economic evaluation for NICE.

Other differences arise between unit costs estimated using the two approaches. The top-down hourly rate or unit cost typically includes all salary related expenditure such as payment for overtime and unsocial hours (i.e. nights, weekends and bank holidays), etc. The bottom-up method has to make assumptions about which grade and salary point to use, and how to handle extra allowances for staff working unsocial hours etc., which are important costs components for services with extended working hours[24]. Yet the "Unit costs of health and social care" only included such allowances for unsocial hours in three schemas (rapid response service, home care workers and doctors). Otherwise, it gave no advice about how to include them or the likely amounts. In addition, the hourly rates include capital and other overheads and will over-inflate costs if simply scaled up.

Table 3.8 summarises the data sources on staff time use. Of the 42 staff cost schedules, 23 used consultations or assumption and only 19 used data from studies - predominantly surveys. Data for hospital-based staff apart from doctors came entirely from consultation with NHS Trusts. Such evidence appears relatively weak given the potential for recall and response bias in opinion-based methods, and it is unclear whether data were checked for accuracy. In the 2005 version, Curtis and Netten (2005) cited the difficulty of obtaining nationally representative and up-to-date data on staff time use. This is reason why the cost schedules continue to use

[24] Typical additional payments to nurses are 30% for working Saturdays and night duty and 60% extra for working Sundays and bank holidays (NHS Whitley Council 2004).

data from old studies (e.g. 1980-90's). They argued that, in the absence of newer studies, small-scale exercises suggested the overall broad categories were sufficiently similar, although specific activities differed.

The overall conclusion from this review was that despite limitations, researchers use these cost schedules as they are easily accessible and the best available. However, it would help researchers if there was more guidance on how to use or interpret information in the schemas.

Table 3.8 Sources of information on time (use and unit costs) for health/social care staff in "Unit costs of health and social care"

Worker	Source of information on direct care and/or contact times (publication dates)
Doctors - Foundation house officer 1, Foundation house officer 2, Specialty registrars	Costs given per hour for 56 or 72 hour week using terms and conditions of service.
Community chiropodist	No information available. Costs based on number of visits per week from NHS Trusts.
Community and hospital pharmacist	Unclear - appears to be assumption.
Family support worker	Unclear
Community occupational therapist (local authority)	Assumption - as no information available they used NHS information from Government statistics (1994).
Social work assistant (SWA)	Assumption based on study by the National Institute for Social Work of 52 SWA (1997).
• Community and hospital occupational therapists • Community and hospital physiotherapists • Community and hospital speech and language therapists • Dietitian • Clinical support worker (hospital), higher level (Band 3) • Clinical support worker (basic, Band 2) • Nurses: i) manager, ii) team leader, iii) 24-hour ward, and iv) day ward • Radiographer	Consultation with NHS Trusts.
• Community nurse (district nurse) • Health visitor • Clinical support worker (community, Band 2)	Study (1982). Contact duration from discussions with a group of NHS Trusts.
Nurse specialist (community)	Study (1995).
Nurse advanced (includes lead specialist, clinical nurse specialist, senior specialist)	% activity and consultation duration from study - 27 nurse practitioners (2000)
Nurse (mental health)	Based on the National Child and Adolescent Mental Health Service Mapping data and returns from over 500 G grade nurses.

Worker	Source of information on direct care and/or contact times (publication dates)
GP practice nurse	Discussions with health service professionals. Contact duration based on a one-week survey of 4 Sheffield practices (1999). Contact duration at surgery based on the 2006/07 UK General Practice Survey. Number of procedures per week from survey (1995).
Consultants - medical and surgical	Audit Commission report (1996).
Consultant - psychiatric	Study by Institute of Psychiatry (500 consultants with 41% response rate) (2003).
General practitioner	2006/07 UK General Practice Workload Survey.
Clinical psychologist	Study (1993) and National Child and Adolescent Mental Health Service Mapping data and returns (Department of Health, 2002).
Social work team leader	Study for Department of Health by the National Institute for Social Work (1997).
Social worker (adult)	Four studies: 1991 Scottish Office, 1995 & 1997 PSSRU, and 1997 Department of Health.
Social worker (children)	Two studies: 1999 and 2001 for the Department of Health.
Approved Social Worker (mental health)	Study of 237 mental health social workers semi-structured questionnaire and diary.
Alcohol health workers (mental health nurses) in A & E	Survey (2004, unpublished).
NHS community multidisciplinary mental health team key worker for elderly people with mental health problems	Study of two teams (1995).
Home care worker (Local Authority)	Local Authority Social Care Workforce Survey (2007) and % activity from benchmarking club of 14 local authorities in the Midlands (from 1998/1999).

Source: Information from the "Unit costs of health and social care" by (Curtis 2007)

3.4.1.2 Use of time measurement techniques in economic evaluations

From Chapter 2, two key systematic reviews of costing methods and guidelines also offered insight into the use of time measurement techniques in the health economics literature. Mogyorosy and Smith (2005) found that over the last two decades few published evaluations have used time and motion methods. Adam et al (2003a) found wide variation between published evaluations in the methods used to allocate staff time. Apart from time and motion studies, other examples included interviews, self-administered time logs, and structured questionnaires.

Whilst these findings were part of broad reviews of costing methods, time study is not the primary focus in economic evaluations. Papers seldom mention use of these methods in the abstract and bibliographic databases rarely index them as time study. Hence, to assess the use of time measurement in economic evaluations it was deemed more efficient to target studies where staff inputs were likely to be crucial to the evaluation.

Two targeted systematic reviews were conducted. One examined comparisons of haemodialysis across different settings, as staff costs were potentially important to the service change. The review, described in full in section 4.5.3.2, found that most (11 of 17) studies simply ignored potential variations in staff inputs. Furthermore, whilst only three studies either measured direct nursing care hours or attributed nursing time using workload measures, the methods were not reported.

The second systematic review examined studies on role substitution in the journal "Health Technology Assessment"[25]. This single journal was chosen because it allows unrestricted reporting. For other journals, authors must work within word limits whilst trying to report the numerous aspects recommended for example by Drummond and Jefferson (1996). Hence, when methods are missing, it is usually difficult to know whether they were inadequate or simply badly reported. The search found five relevant papers that are summarised in the Table 3.9.

Boland et al (2003) compared two methods for nurses to insert special intravenous lines into cancer patients. It was the only paper to explain the rationale for the time

[25] Inclusion criteria: Reports based on titles and where costing staff time would be a key issue. All reports searched up to July 2007.

measurement methods, but the authors' reasoning was questionable. They argued that observation methods were inappropriate

> "given the *confidential nature of the service* provided by nurses" (p22).

Whilst this may reflect a local issue, it is not true in general. They argued against nurses recording individual patient times on the basis that nurses often treated more than one patient simultaneously, yet they asked nurses to estimate patient times.

Overall, the five papers illustrated four points. First, despite the lack of reporting restrictions, there was a lack of transparency and readers had to accept the results at face value. Kinley et al (2001) did not even state what method of time measurement they used. Authors rarely explained their methods for costing staff time, which suggests lack of concern or insight into the data used. There were disturbing reporting inconsistencies. Questions purported to be in the data collection instruments were missing from the paperwork in the studies by Caine et al (2002) and Townsend et al (2004). Second, the remaining four studies stated they used self-recording, but gave no insight into whether this was simply deemed the cheapest option. Some collected data retrospectively (i.e. self-reporting) and in others this was unclear. Third, although often true of the general study data, none of the studies attempted to check the validity of the time data (e.g. by observation). Fourth, the presentation of time data was not always appropriate or adequate for economic evaluations, with use of median rather than mean times, and staff costs not disaggregated into resource use and cost data. Overall, the measurement of staff time received insufficient attention and it was difficult to assess the accuracy of the time estimates and effect this might have on decision-making.

Table 3.9 Economic evaluations of role substitution in Health Technology Assessment

Paper	Setting / Comparison	Method of time measurement	Time aspects measured	Time measurement validity checks	Time - statistics and sensitivity analysis
Morrell et al (2000)	Community: Community support workers (CSW) cf. non-intervention control group for post natal care (first 28 days)	CSW self-logging (retrospective). Community midwives (CM) hospital survey using self-logging (unknown if retrospective)	CSW visit duration only (i.e. not contacts of other professionals). CM ante- and post-natal visit durations and travel. CM cost data randomly assigned to mothers' midwifery contacts (i.e. not measured on specific mother). Other resource use frequencies from mothers' self-complete questionnaires.	Not mentioned	Mean and SD. Some sensitivity analyses
Kinley et al (2001)	Hospital: Appropriately trained nurses cf. pre-registration house officers pre-operative assessment before elective general surgery	"Collected prospectively from the trial" (method not stated, ? observation)	Assessments by the Pre-registration house officers, Appropriately trained nurses and anaesthetist.	Not mentioned	Mean and SD. Sensitivity analyses (incl. probabilistic sensitivity analysis (PSA) using trial data, but unclear if PSA ranges data / researcher driven)
Caine et al (2002)	Hospital: Nurse practitioners cf. doctors outpatient care in bronchiectasis clinic	Patient self-recording (methods not described, no evidence on paperwork)	Consultation duration	Not mentioned	Mean (no SD)
Boland et al (2003)	Hospital: Blind and image-guided insertion of Hickman lines by nurses for adult cancer patients	Self-recording (start / finish on last page of case report form) and interviews with staff. Times for activities in X-ray suite from log book (retrospective)	Procedure, time in interventional X-ray suite to reposition misplaced catheters (but data not used in economic evaluation), and waiting time between insertion and repositioning.	Not mentioned	Mean 95% CI
Townsend et al (2004)	Hospital: Midwives cf. paediatric senior house officers (SHOs) routine examination of newborn babies	Self-recording on examination sheet (but no evidence on paperwork) and interviews with staff re admin time	Examination duration	Researchers could have partially validated time data from 39-videotaped examinations recorded for quality checking the assessments	Medians not means

3.4.2 Time measurement studies in health care

Given the limited use of time measurement in economic evaluations, it was important to evaluate other types of health care study that measured staff time. A starting point was the local NHS Trust that had co-ordinated several studies using barcode scanners to self-record staff activity. In addition, a systematic review was conducted of studies that specifically measured time at the patient-level. Lastly, health care studies that compared methods provided further insight into validity and reliability issues.

3.4.2.1 Barcoding studies at Southampton

Section 3.3 noted that self-recording is a useful method for collecting patient-specific time. Over about four years in the 1990s, Southampton University Hospitals NHS Trust co-ordinated five projects where staff self-recorded their activity using barcode scanners. Three projects that covered the neurosurgical unit, medical records staff, and medical and surgical consultants' workload were briefly reported in a journal article (Macfarlane and Lees 1997). Another study compared nursing on a nurse-led unit and an acute ward (Walsh 2003). A further, unpublished study examined activity of primary care staff (Taylor et al 1998).

The project on the neurosurgical unit was particularly important. Its first objective was to develop a decision support system (spreadsheet model) that clinicians and accountants could use to investigate the costs of different scenarios and contracting possibilities. Its second objective was to analyse the data to validate whether the relevant HRGs were suitable to classify and cost patient care episodes. The project, funded jointly by the Chartered Institute of Management Accountants and NHS, was described in detail in a book chapter and book (Connell et al 1996, Connell et al 1997).

The project needed accurate data on staff time for nurses, physiotherapists and senior house officers, and hence costs. The authors ruled out traditional time and motion study as impractical and too costly. Staff undertook an initial trial of self-recording using paper time sheets, which found they did not accurately measure their time, sometimes guessed, and filled in records retrospectively. Instead, the staff used hand-held barcode scanner 'pens' to record their work to unique codes developed for each activity, staff member, patient, and other resource use items (drugs, consumables, tests, etc.). Connell noted that use of barcode scanners was novel in health care at that time.

A 4-month study covered 80% of admissions to the neurosurgical unit (2 wards and neuro-intensive care). Data were cleaned to remove outlier data for activities (implausible data outside minimum-maximum values based on staff expert opinion), or patient episodes with a mean that exceeded six standard deviations. Incorrect or missing data were imputed using the mean of the 'valid' values.

Total direct costs for 498 complete episodes were analysed by patient episode characteristics[26]. Through empirical clustering techniques, patient episodes were grouped into iso-cost groups (using total direct costs) and iso-resource groups using a set of eight resources. The latter were in fact disaggregated cost groupings (costs for nurses, physiotherapists, senior house officers, consumables, drugs, theatre staff, other theatre costs, tests). Analyses found substantial differences in direct costs between patient episodes, including when grouped by HRG version 1. Furthermore, patient episodes could have the same direct costs and yet comprise very different patterns of resource use. After further analyses, the authors concluded that patient characteristics were superior to the HRGs in classifying patient episodes into the clusters[27], but that the project's validation process had been superseded by the introduction of updated and improved HRGs (version 2).

Barcode scanning was extended a further four months to record nursing activity (using times collected previously) to investigate nursing skill mix. It was also used for three four-week periods on the neuromedical ward to investigate various time measurement, time use and skill mix issues. Overall, the authors noted that adaptation of the results into a wider management information system was very time

[26] Significant differences in total direct costs were found for increasing length of stay (but at a decreasing rate), having an operation or tests, emergency rather than planned admission, increasing severity score on admission. Those with no improvement in severity score cost the least, followed by those who improved and the highest costs were for those whose severity score worsened. Costs showed considerable variation within HRGs, although this classification of patients accounted for a statistically significant amount of variation in total costs (figure not given). No significant differences were found for age and gender.

[27] Statistical discriminant analysis found that using the patient characteristics (reduced to operations, length of stay, tests and two HRGs) to classify patient episodes to clusters was not uniform and only accurate about 50% of the time, whilst there was no direct link between nine HRGs and the clusters. They examined relative performance using lambda. This statistic ranges from zero when of no use to predict a cluster, to one when the predictor is completely accurate. Lambda values were 0.20057 for HRGs and 0.27507 for the patient characteristics.

consuming. They also cited a number of advantages and disadvantages of barcode scanning to capture data, as shown below:

Advantages	Disadvantages
• Scanning, unlike other self-recording, prevented over-recording of times and hence total time exceeding the shift length. • Scanning allowed identification of the staff member, the time and duration of activities, and so it was possible to identify when and how many staff were simultaneously involved in a patient's care (which is difficult with other methods). • Electronic data capture eliminated transcription errors and enabled rapid data processing after the shift. • No keyboard or writing skills required. • Scanning prevented retrospective recording of durations (but not frequency of activities). • Scanning was a straightforward means to follow patients through different wards and theatre (within the neurosurgery department).	• Scanners could not record activities of less than one minute. • Accuracy of individual activities was dependent on diligent scanning. • Scanners sometimes broke down. Static electricity caused the clock and date to change, but activity time remained recorded. The problem was minimised by frequent downloading and by good training. • Administrative burden of producing new barcodes for admissions. • Specialised equipment and software required to produce, read and process barcode information. • Staff training needed. • Sometimes nursing staff working extra shifts as agency 'signed on' under their staff code (although rectified by checking staff rotas). • Some types of staff, e.g. agency staff, may have been less reliable in their recording due to a lack of training or motivation.

The authors concluded that staff acceptability of barcode scanning was high, although enthusiastic clinicians pioneered the projects. The data were valuable for a variety of purposes beyond the original objectives (e.g. initiated improvements in working practices). The authors considered that barcode scanning offered a cost-effective method to capture data, although it required 45 scanners at a cost of £300-400 each (i.e. approximately £16,000), with additional costs for software. They noted that the nursing time captured increased in use, but accuracy declined as interest diminished towards the end of the project, but did not specify how they assessed accuracy.

Anecdotal comments were also available from the co-ordinators of three projects. Frequent visits during data collection by the co-ordinator to monitor progress and give feedback appeared to increase staff involvement. Walsh (2003) reported poor recording and data only sufficient for contextual information. It transpired this occurred largely due to various 'political' and pragmatic difficulties at one site (including a flood) - problems not experienced in the other projects.

The unpublished study in primary care found that development of codes for each type of practice staff took longer than expected (six months rather than two) (Taylor et al 1998). The observer recorded more activities than staff, although proportions were similar, but suggested that further training might improve consistency of coding. Overall, staff under-reported their time by 2 to 11% compared with the observer, who in turn under-reported available time by up to 12%. Staff acceptability of barcode scanning was generally high, although the views differed slightly between staff groups. Staff were generally more negative towards data collection by an observer or video. Interestingly, GPs would not record direct contacts during consultations, although this appeared to have little impact on the overall time recorded.

Overall, barcode scanning appeared to be a feasible method to collect patient-specific data, despite various challenges. Taylor et al (1998) found problems downloading barcode scanners off-site from SUHT and had to seek the help of the management consultancy that developed the software. Whilst barcode scanning appeared to offer a useful enhancement to data collection, the next section provides a broader examination of studies of patient-level time measurement.

3.4.2.2 Studies of patient-level time measurement in health care

This section reports the results of a systematic review of studies that measured staff time at the patient-level[28]. The aim was to gather information on validity, reliability and feasibility issues from representative examples of recent papers. Searching for such papers is challenging. Bibliographic databases index "time and motion" and "work sampling" as time study, but self-recording, self-reporting and expert opinion are not similarly indexed. Given the nature of the terms, text word searches tend to retrieve large numbers of irrelevant citations (i.e. the searches are too sensitive with low precision). Consequently, some searches restricted the publication years.

The results of the literature searches and data extraction table are shown in Appendix 2. The 10 papers comprised 12 studies: three time and motion[29], two self-recording[30], and five papers (seven studies) of self-recording using barcode scanners[31]. No relevant papers were found on work sampling which suggested this method was rarely used to measure patient-specific time.

---

[28] Inclusion criteria: Time measurement studies: i) English language with abstract, and ii) Primary studies of time and motion, work sampling or barcode time, and iii) involved clinical staff (doctors, nurses, physiotherapists, or therapists), and iv) included actual measure of time (i.e. days, hours, minutes, or seconds) in title or abstract and publication 2001-2006, or included validity or reliability in title or abstract (no year restriction).
Exclusion criteria: Papers on sport, time to event or waiting times, where 'time' only mentioned, time for tasks rather than patient focus.
Additional barcode scanning studies: i) barcode papers related to time / work measurement.
Search strategy: Time measurement studies Set1 = exp 'Time and Motion Studies'* or (title or abstract: (time adj (motion or study or studies)) or (time and motion or time-and-motion) or work sampling or bar cod$ or barcod$)** AND (title or abstract: staff$ or nurse? or doctor? or clinician? or physician? or physio$ or therapist?) NOT (publication type: editorial or letter or review) LIMITED TO (humans and English language and abstracts)
Set1 AND (mesh: valid$ or reliabil$) OR Set1 LIMITED TO year='1996 - 2006' AND (title or abstract: second? or minute? or hours? or days?)
Where mesh=title, abstract, subject headings,
* Applicable in Medline & CINAHL ** Downloaded for BNI
Additional barcode scanning studies: Mesh: bar?code AND (time or staff or skill?mix or case?mix or work measure$ or workload or resource management or manpower plan$ or grade mix or staff$ level$).
Databases: Allied and Complementary Medicine 1985 to 2006 Jul; British Nursing Index (BNI) 1985 to 2006 Jul (2004 Oct); CINAHL 1982 to 2006 Jul Wk 4 (2004 Nov Wk 1); Embase 1980 to 2006 Wk 30 (2004 Wk 45); HMIC Jul 2006 (2004 Sept on 12/11/04); Medline 1966 to 2006 Jul Wk 3 (2004 Nov Wk 1) Dates in brackets relate to the search for additional barcode scanning studies.
[29] Oliver et al 2001, Zupanic and Richardson 2002, Larson-Lohr 2003
[30] Carpenter et al 2001, Cromwell et al 2004
[31] Walsh et al 2003, Martin 1990, Macfarlane and Lees 1997, Holmes et al 1997a, Blount 1999

The papers covered a wide variety of settings and staff. In common with earlier findings, reporting was often inadequate. Authors did not report the number of hours recorded in relation to available hours, and so it is difficult to judge how representative the time recordings were. Only Cromwell et al (2004) gave an indication of the missing data (~6% of patient days). In the time and motion studies, it was unclear how many observers were used or whether there was any assessment of inter-rater reliability.

Overall, validity, reliability and practicality issues received scant attention. Only Cromwell et al (2004) described thorough validity checks for time recording. This involved the co-ordinators checking the self-record forms for completeness and accuracy, but did not state how they assessed this. There was minimal discussion of validity issues in three papers (Larson-Lohr 2003, Walsh et al 2003, Macfarlane and Lees 1997), and two papers cited previous use of tools as evidence of validity (Carpenter et al 2001, Holmes et al 1997a). Four studies did not mention validity issues (Oliver et al 2001, Zupanic and Richardson 2002, Martin 1990, Blount 1999, and the latter two were PhD theses). In contrast to the time data, authors tended to report assessments of validity, reliability or both for other outcomes used (Oliver 2001, Martin 1990, Walsh 2003). These findings were interesting as three studies were using the time data to validate outcome measures (Martin 1990, Blount 1999, Carpenter et al 2003).

In terms of statistical issues, authors rarely reported the extent of missing data, although three studies did impute missing values (Cromwell et al 2004, Holmes et al 1997a, Macfarlane and Lees 1997). In addition, apart from Holmes et al (1997a), it was unclear whether researchers took account of repeated measurements from the same staff or patients, which would make results appear more precise than they should.

Papers contained some pointers about the practicalities of barcode scanning data collection. Holmes et al (1997a) noted that trainers were available everyday during data collection to give rapid feedback to staff and check anomalies. Macfarlane and Lees (1997) reported that whilst most staff took part, the appointments staff at one site were very reluctant to use barcode scanning and only recorded a small amount of the time, although no figures were given. At another site, a medical consultant found difficulty using the barcode pen and was excluded. Similarly, Walsh et al (2003) noted under recording by the staff on one unit, but no proportion was given.

Although a small sample of papers, barcode scanning studies tended to record data over a longer time than the time and motion studies (weeks rather than days). Overall, these findings confirmed the conclusion of a review by Holmes et al (1997b) that assessment of validity, reliability and practicality of time measurement methods was rare.

3.4.2.3 Comparative time measurement studies

Burke et al (2000) noted that few studies directly compared time measurement techniques. In addition, given the difficulties in identifying time studies, a specific search was not conducted to identify such studies. Instead, papers found as part of the previous search or referenced in other papers were examined. Again, key issues for assessment were validity and reliability.

Before discussing the findings, it is important to note a number of considerations in reviewing such studies. Section 3.3 reported that assessment of validity and reliability are complicated. For a given context, one method may be more appropriate or feasible than another, especially for data collection at the patient-level. Methods vary in coverage and so, unless activity coding is like-for-like, one would not expect different methods to produce the same results. Furthermore, Lee et al (2003) noted the potential for discrepancies in interpreting of codes between observers and participants. It seems likely this will escalate as the number of individuals involved increases. Lee et al (2003) also reasoned that for practical and cost reasons, simultaneous data collection by another method might only be possible for a small proportion of the data.

In two of the comparative studies, data collection was not concurrent. Oddone et al (1993) used self-recording followed by work sampling. Burke et al (2000) used time and motion study followed by self-recording. In the latter, proportions of total time in the main activity categories were comparable, but self-recording under-reported the frequency of activities (by up to three times), which led to large differences in mean time per activity. The authors noted that nurses found self-recording a burden, which may have contributed to coding differences.

Results varied for other comparative studies. A study by Finkler et al (1993) compared time and motion and work sampling, but was methodologically poor because work sampling observations were vastly underpowered. Bratt et al (1999) measured clinician time at three reproductive health services clinics. They used "time-motion" as the benchmark compared with three other techniques: patient flow analysis[32], structured interviews at the end of each shift, and self-recording. There were statistically significant differences between time-motion and each of the other three techniques. The interview method was poor because contact time was vastly over-estimated, although overall time reported was similar. Overall time was under-estimated by over 30% using patient flow analysis and self-recording, particularly due to under-estimation of "non-contact" time and "non-contact non-productive" time. A criticism of the study is that it used the time-motion method as the 'gold standard', but it appeared to be a work sampling study (as observer-recorded activities at 3-minute intervals). It was unclear whether the researchers assumed that the activity observed had lasted the previous 3-minutes, and as a work sampling study it may have been underpowered[33].

A more robust study by Stewart and Short (1999) evaluated hand-written and barcode logged events observed on video tapes of simulated resuscitations. Compared with the videotaped time, the barcode method was more accurate than hand-written logs, as assessed by the mean absolute errors and their standard deviations ($p<0.01$). Omission of events was not significantly different. In conclusion, these studies highlighted the challenge faced in like-for-like comparisons to assess the validity of time measurement techniques.

3.5 Conclusions

Through review of literature and other sources, this chapter has considered the second objective of the thesis - evaluation of methods to measure staff time per patient. Measurement of staff time may focus on tasks, aggregated activity or patients. A first step is categorisation of staff activity, although it may be difficult to ensure consistent coding for staff-patient interactions. Furthermore, to derive the cost per patient, staff time must be allocated to types of patients, however defined. Within patient-groups, some elements of nursing time may be relatively fixed (i.e. similar for each patient), whilst others may depend on patients' or other characteristics.

---

[32] This involved staff recording start and end times of contacts with patients on a sheet that the patient kept for the duration of their visit. Non-contact time was ascertained from structured interviews with staff at the end of the shift.
[33] The researchers did 20 observations per hour for 10 four-hour shifts at each of the three clinics (i.e. approximately 2,400 observations in total).

Time measurement techniques originated in production engineering and their application in health care is challenging. Methods involve observation, self-recording, self-reporting or expert opinion, but none is a 'gold standard'. The choice of method for a specific context requires trade-offs, since they vary in the aspects covered, feasibility of implementation, and validity and reliability (or the ability to assess these). Moreover, depending on the purpose for costing, measurement of staff time may be not worthwhile. For example, sophisticated time measurement may not be a good use of research effort where variability of staff time is restricted through use of fixed appointments.

Staff time includes unproductive time and general activity. These are important in calculating the unit cost of staff time. In top-down costing, unless weighted, all staff costs are attributed equally across patients. In contrast, the bottom-up approach may use a variety of methods to attribute time to patients. The annual "Unit costs of health and social care" (Curtis 2007) is a key source of data on staff costs. It includes other cost aspects such as qualifications and overheads, but presents data so that users can select specific elements. Costs are presented at various levels of aggregation (depending on assumptions used). Limitations are that data on staff time use are predominantly from consultations, assumption or surveys, which appear relatively weak sources of evidence given the potential for recall and response bias. In addition, it does not include costs for staff working unsocial hours. Although the cost schedules are the best available nationally, it would help researchers if there was more guidance on how to use or interpret information in the schemas.

Appraisal of literature revealed that time measurement techniques have not been widely used for economic purposes and self-reporting or opinion-based methods dominate. In these and other time measurement studies in health care, reporting of methods was generally poor. Although difficult to assess, most studies paid little attention to validity and reliability.

In terms of measuring time at the patient-level, observation techniques have limited ability to capture indirect care. Conversely, self-recording can capture all aspects of care, although it interferes with staff work. The review suggested that barcode scanning might enhance self-recording. The next chapter gives an overview of haemodialysis and explains why it offered a useful case study to test methods to measure staff time.

Chapter 4 Renal failure and renal services

This chapter sets the context for the third objective of the thesis, to assess the impact of patient heterogeneity on nursing costs for chronic haemodialysis. A key purpose is to explain the rationale for using haemodialysis as a case study. It gives an overview of clinical and epidemiological aspects of renal failure, available treatments, service provision, and policy initiatives to meet the demand for treatment. Lastly, it reviews economic evaluations of haemodialysis in different settings, as examples of analyses where patient heterogeneity could affect the costs.

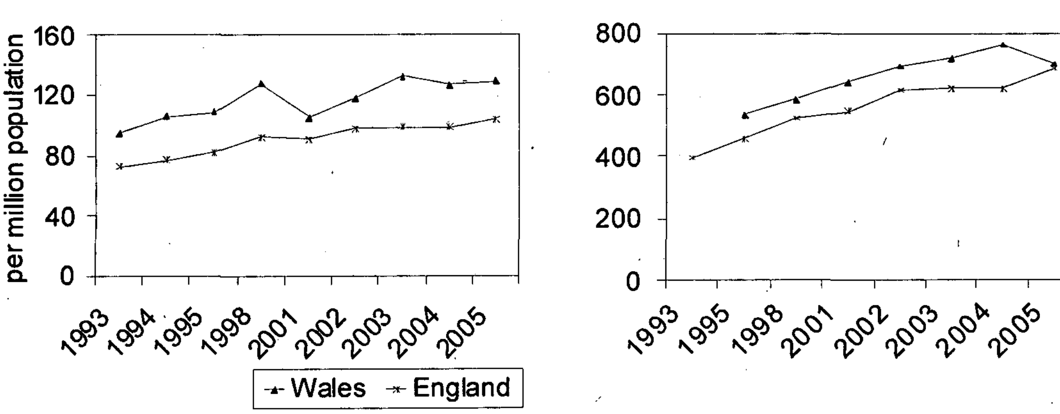
4.1 Epidemiology of established renal failure

Chronic kidney disease (CKD) occurs when there is progressive loss of kidney function due to irreversible damage, often over months or years. The kidneys have various functions in terms of removing excess fluid, minerals, and waste material from the blood. They also secrete hormones: erythropoietin (EPO) involved the production of red blood cells; calcitriol, a form of vitamin D involved in regulation of calcium and phosphorus, and hence bone metabolism; and renin involved in blood volume and blood pressure control (Ansell et al 2006). Although escalating loss of kidney function is detectable by a simple blood test, patients tend to present late because initially there are no symptoms or unspecific ones such as tiredness and anaemia. Otherwise, early detection is typically secondary, for example, to finding that a patient has hypertension (high blood pressure).

Eventually, CKD may progress to the terminal phase called established renal failure (ERF), end stage renal disease (ESRD) or end stage renal failure (ESRF). Established renal failure is fatal if not treated with a renal replacement therapy (RRT, described in the next section). Ansell et al (2006) noted that the presence of other diseases (co-morbidity) affects the choice between treatments and effectiveness. They outlined three causes of such co-morbidity. First, the primary disease; for example, diabetes damages nerves and blood vessels and so causes blindness and cardiovascular disease. Second, ERF leads to conditions such as anaemia, bone disease and heart failure. Third, other diseases, for example chronic bronchitis and arthritis are common in older people with ERF.

The true incidence and prevalence of ERF are unknown because not all patients with ERF are identified, referred and accepted onto treatment. Instead, the rate of acceptance of new patients onto RRT and maintenance of patients on RRT are proxies for the incidence and prevalence of ERF. Since 1998, the UK Renal Registry has produced annual reports. From these, Figure 4.1 shows that the incidence and prevalence of RRT in England and Wales has been increasing since 1993. In 2005, there were an estimated 36,660 patients in England and Wales on RRT; a prevalence rate of 0.07%, or 687 people per million population (pmp) (Ansell et al 2006).

Figure 4.1 Incidence and prevalence of adult patients on RRT 1993-2005 (England and Wales)



Notes: Data not available for all years.
Sources: Data from Ansell et al (2002), Ansell et al (2003), Ansell et al (2004), Ansell et al (2005), Ansell et al (2006).

Increasing demand for RRT is expected to continue. Simulation modelling by Roderick et al (2004) predicts that demand for RRT, and particularly haemodialysis for elderly patients, will grow for at least 25 years. They estimate that in 2010 the prevalence of RRT will be 42,000-51,000 or equivalent to 900-1000 per million population.

Several inter-related factors are driving the increasing demand for RRT. Demographic change, specifically the ageing population, is important because CKD and ERF are more common in the elderly. A related factor is the increasing prevalence of Type 2 diabetes, a major cause of ERF (Ansell et al 2006). Furthermore, diabetes contributes to the higher rates of CKD and ERF seen in some ethnic populations such as those of Indo-Asian and African/Caribbean origin (Ansell et al 2005). Other factors include the increasing survival time for patients on RRT and continuing liberalisation of acceptance on to RRT.

4.2 Treatment options for ERF

Treatment options for patients with ERF comprise renal replacement therapy (RRT) by two main modalities - kidney transplantation and dialysis - or alternatively specialist palliative care. The latter is supportive care to control symptoms and complications of ERF using drugs and dietary interventions. It is particularly relevant for the very elderly or those with extensive co-morbidities for whom dialysis may not improve quality or length of life.

Kidney transplantation is the treatment of choice because it completely replaces all kidney functions. Aside from the requirement for immunosuppressant drugs to prevent rejection of the transplant, the patient can lead a 'normal' life (Department of Health Renal Team 2004). Moreover, it is the most cost-effective treatment because after initial surgery and follow-up, the costs reduce considerably (Department of Health Renal Team 2004). However, the availability of kidney transplant is limited by a shortage of organs.

Dialysis, the alternative RRT, involves filtering the blood across a semi-permeable membrane to remove waste products into the sterile fluid (dialysate). It does not restore the loss of hormones secreted by the kidneys and so supplementation with drugs (erythropoietin, vitamin D and anti-hypertensives) may be necessary. There are two modalities of dialysis - peritoneal dialysis and haemodialysis.

Peritoneal dialysis uses the peritoneal membrane within the peritoneal (abdominal) cavity for manual or machine-driven fluid exchanges. Treatment is daily, typically at the patient's home. Since the patient or carer must take responsibility, peritoneal dialysis is not practical for some patients. In addition, a common complication is peritonitis (infection and inflammation of the peritoneum) that may require a temporary or permanent change in treatment modality.

Haemodialysis involves a dialysis machine that pumps the patient's blood through a dialyser (a chamber containing a membrane) so that waste products pass into circulating dialysate. It requires permanent vascular access via needles into a fistula or venous graft, or venous catheter. Treatment takes place predominantly within a renal unit and most patients dialyse for about four hours, three times a week.

4.3 RRT service provision

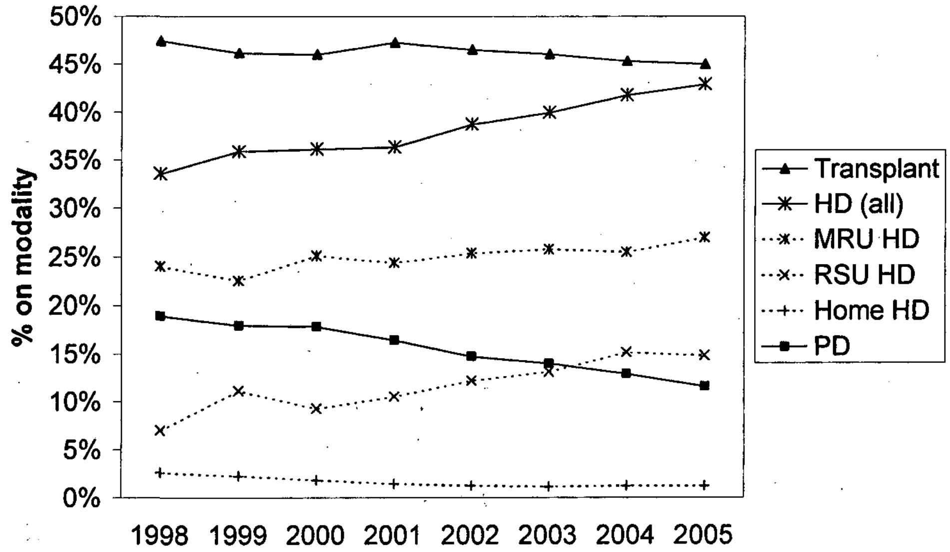
This section focuses on historic changes in RRT modalities in the UK and the trend to provide expansion of RRT through haemodialysis. At the outset, RRT provision in the UK was through a small number of renal units in teaching hospitals and so patients often had to travel long distances. RRT was mostly haemodialysis and restricted to younger, fitter patients (Stanton 2005). In the 1980s, renal services expanded due to a national target (Patten 1984), and the advent of peritoneal dialysis. Subsequently there was major congestion in hospital renal units. This was due to a decrease in the use of home haemodialysis, the limited life span of peritoneal dialysis as a treatment, and increasing acceptance of the elderly with co-morbidity who were unable to manage peritoneal dialysis (Roderick et al 2005).

In the 1990s, there was growing concern about the mismatch between the estimated need and provision of RRT (Mallick 1994). In addition, there was pressure to achieve RRT rates comparable with other countries (Stanton 2005). A study in Wales found that in patients aged over 60 years, there was a negative relationship between referral rates and patients' travel distances to the renal unit (Boyle et al 1996). Roderick et al (1999) found a similar negative relationship in England, irrespective of patient age. Moreover, geographic variations in services did not simply reflect differences in population need by age or ethnicity.

Main renal units (MRUs) offered full renal services including inpatient beds for CKD and ERF, treatment of acute renal failure, and some units provided transplantation. Expansion of existing MRUs was problematic due to space constraints within their hospitals. An alternative was renal satellite units (RSUs) where nurses provided long-term haemodialysis under the overall clinical management of the parent MRU, but with limited or no on-site specialist medical help. RSUs were located as freestanding units or within other hospitals, so reduced patient journey distances and improved access to RRT. In consequence, the National Renal Purchasing Guidelines suggested expansion of RRT provision through RSUs (Department of Health 1996).

Figure 4.2 shows there have been continuing shifts in the proportions of patients on the RRT modalities in England and Wales between 1998 and 2005. Transplantation and peritoneal dialysis decreased by 2% and 7% respectively, whilst haemodialysis increased 9% largely by the doubling of RSU patients (Ansell et al 2003). Home haemodialysis declined from 3% to 1% despite policy guidance encouraging its use by the National Institute for Health and Clinical Excellence (NICE 2002, Mowatt et al 2003) and the National Service Framework (NSF) for Renal Services (Department of Health Renal Team 2004).

Figure 4.2 Dialysis modalities of adult patients 1998-2005 (England and Wales)



Notes:
HD = haemodialysis, PD = peritoneal dialysis
Sources: Ansell et al (2005) and Ansell et al (2006)

Provision of specialist palliative care as an alternative to RRT is growing, although access is often restricted to patients with cancer not ERF. A survey found that only 39% of UK renal units had staff with a formal specialist palliative care role and the amount of time spent delivering such care was small - mostly less than 4 hours per week (Gunda et al 2005). However, most renal units follow up patients who choose not to have dialysis and the NSF for Renal Services specifically recommends an increase in palliative care provision.

Overall, in 2005, 45% of patients on RRT had a functioning kidney transplant, whilst the remaining 55% were treated by dialysis. Peritoneal dialysis accounted for 12% of RRT provision, or 21% of dialysis patients, a relatively high proportion compared with other developed countries (Ansell et al 2006). Furthermore, the growing proportion of patients on haemodialysis includes an increasing percentage of patients who are elderly and have co-morbidity (Feest et al 2005).

4.4 Challenges for haemodialysis provision - policy initiatives

Due to various factors, including demographic changes, there are continuing pressures to increase the provision of RRT and haemodialysis. This section discusses policies relevant to haemodialysis provision and the payment mechanism for service providers.

4.4.1 Policy initiatives

Numerous policy initiatives influence the provision of health services in England and Wales. Initiatives of relevance to improving the quantity and quality of renal services include: the NSF for renal services, the Renal Association and NICE guidelines, monitoring by the UK Renal Registry, and incentives for care of patients with CKD in primary care (the "Quality and Outcomes Framework"). These are discussed in Appendix 3, a published review of the organisation and financing of renal services in England and Wales - part of an international study conducted alongside the PhD.

Another policy objective, mentioned in the last section, has been to improve geographical access to treatment through satellite and home haemodialysis. Despite improved proximity to units, patients still face challenges in travelling to and from their dialysis sessions. Many patients use NHS transport including ambulances, hospital cars, and occasionally taxis to attend dialysis sessions, but cannot guarantee that they will arrive on time and many units experience difficulties in coordinating patients' transport home. This has necessitated initiatives to improve patient transport (Department of Health Renal Team 2004, Cheshire and Merseyside Renal Transport Action Learning Set 2006).

Further initiatives include the "NHS Plan". Through this, the Government proposed 450 new and replacement haemodialysis stations to treat another 1,850 patients and 1,200 existing ones (Department of Health 2000a). "Delivering the NHS Plan" describes the implementation, which included financial reforms such as Payment by Results (PbR), discussed in the next section (Department of Health 2002a). In addition, the "NHS Plan" sought to promote patient choice within the NHS (Department of Health 2000a). Furthermore, increased patient choice was advocated in renal services by increased provision of home haemodialysis and palliative care (Mowatt et al 2003, NICE 2002).

In practice, improvements in patient choice pose challenges. In most areas of England and Wales there is no real competition between providers and so patients have little choice over facilities. Moreover, aside from some medical and practical considerations (e.g. ability of the patient to self-care), the availability of services affects the choice of modality. Service commissioners try to meet the increasing demand for RRT, but treatment costs are high (discussed in the next section). The NHS does not dedicate funds for renal services in general or to implement the improvement policies, although overall national health care expenditures increased 44% in real terms between 1999 and 2004 (Yuen 2005). There is inevitable tension between providers trying to develop services to meet markers of good practice and commissioners trying to prioritise services for funding.

4.4.2 Payment for haemodialysis

Although the number of people needing RRT is relatively small, the treatment is costly. Each patient on haemodialysis costs £21,000-31,500 per year, based on the national reference costs and indicative tariff shown in Table 4.1. The national tariff is central to the Payment by Results (PbR) reimbursement mechanism described in section 2.2.3.3. Some commissioners plan to use the tariff to reimburse the MRUs, although it is not compulsory for renal services (Nicholson and Roderick 2007). Others continue to negotiate a local tariff based on an agreed number of haemodialysis sessions. Through service level agreements, a different payment rate may apply for any extra haemodialysis sessions delivered.

Table 4.1 Haemodialysis reference costs and indicative tariff (England, 2006)

Haemodialysis codes	Unit cost (£)		Annual cost [a] (£)		Indicative tariff [b]
	Mean	IQR (25-75%)	Mean	IQR (25-75%)	
HRG 3.5 [c]					
MRU [d]	182	154-203	28,383	23,979-31,646	159
RSU	158	136-174	24,632	21,206-27,078	132
Home	98	87-138	15,334	13,539-21,532	102
MRU [d] patients with infectious diseases	184	169-186	28,710	26,426-29,036	189
Holiday	193	154-186	n/a	n/a	155
HRG 4 [e]					
HD / Filtration	158	129-197	24,590	20,180-30,690	
HD / Filtration with Hep B	175	134-178	27,380	20,900-27,830	

Notes: n/a = not applicable

a. Calculated assuming haemodialysis three times per week, 52 weeks per year.

b. Department of Health (2006a)

c. 2005-06 reference costs (Department of Health 2006c) inflated to 2006-07 prices using pay and prices index from Curtis (2007). Mean costs are weighted national averages, whereas 25% and 75% use providers' submissions and are not weighted and not comparable with mean.

d. Including inpatients, outpatients, ward attenders etc.

e. 2006-07 reference costs (Department of Health 2008a)

Table 4.1 shows the two most recent sets of reference costs. The HRG 3.5 costs are 2005-06 reference costs inflated to 2006-07 and include separate costs for MRU and RSU haemodialysis. In contrast, the move to make HRGs independent of setting means that version 4 HRGs only split adult haemodialysis if the patient has Hepatitis B, as such patients require extra resources (e.g. sole use of a dialysis machine). The split by setting (MRU, RSU, home) for the indicative tariff will disappear in 2008-09, as the tariff is based on the reference costs two years earlier.

For the HRG 3.5 costs, simple comparison of the mean figures suggests that haemodialysis at a RSU was only 87% of the cost at a MRU, a difference of only £24 per session but £3,750 annually. Similarly, haemodialysis at home appeared to be 62% of the cost at a RSU, a more dramatic difference of £60 per session and £9,300 annually. As shown later (section 4.5.3), it is common to use costs averaged in this way in economic evaluations, but the comparisons are not like-for-like. The cost of haemodialysis at the MRU was the average across a mixed patient group (including those both suitable and unsuitable for RSU care). Likewise, not all patients at MRUs or RSUs were capable of self-care at home. Simple comparisons make haemodialysis at a RSU look more favourable than at MRU, and haemodialysis at home more attractive than at a RSU.

Cost escalation is a major concern. Mallick (1997) predicts that when the RRT programme approaches a steady state in the UK, RRT will consume 2% of the national health care budget. Indeed, Winkelmayer et al (2002) noted that the high cost of this life saving treatment is a reason why, in the United States, RRT is the only treatment for which Medicare provides universal coverage in a largely private health care system.

Whilst the UK uses flat-rate reimbursement for haemodialysis, in the United States, Medicare has changed from a flat rate because it produced disincentives to care for costly patients. Wheeler et al (2006) detailed the differences between the old basic "composite-rate" and new case mix adjusted payments (see Table 4.2). A report by the Kidney Epidemiology and Cost Center (2004) described the modelling to identify the case mix factors. The data were at the facility level for both costs and an average case mix measure that used patient-level co-morbidity data from claims or bills (from 2000-2002). For various reasons (due to poor data or atypical units), data from 14% of the 2978 RSUs and 45% of the 214 MRUs were excluded. The authors noted that case mix adjustment needed routine, objective data, which precluded many co-morbidity variables. However, there has been controversy over the rationale given for the specific case mix adjustments (Himmelfarb and Chertow 2005). In addition, various projects are investigating inclusion of other fee-for-service items (e.g. drugs) within an increased composite rate, to reduce cross-subsidisation and remove perverse incentives (Hirth 2007).

Table 4.2 Medicare reimbursement for haemodialysis

Basic composite-rate	**New case mix adjusted system (from 1/4/2005)**
Flat rate with limited adjustments: • Geographic, for wage rate differences. • Facility - slightly larger payments for MRUs compared with RSUs. • Additional payments when high proportion of paediatric patients or geographical isolation.	'Case mix' adjusted for: • Age, due to the U shaped relationship found between age and cost. Younger patients were more likely to skip treatment and had higher prevalence of AIDS/HIV. • Body surface area as larger patients took longer to dialyse, needed larger dialysers and limited the unit's capacity. • Low body mass index (more costly due to various factors including unmeasured co-morbidity, increased frequency of admissions and missed sessions).
Services outside bundle (fee-for-service) include injectable medications, non-routine laboratory tests, and vascular access procedures.	

Source: Adapted from Wheeler et al (2006)

4.5 Economic evaluations of RRT

Having looked at the background to renal failure and RRT, this section examines costing in economic evaluations of RRT. After an overview of the literature, it describes how the RSU study (Roderick et al 2005) was the stimulus for the thesis and empirical work. The final part reviews the handling of patient heterogeneity in evaluations of haemodialysis in different settings.

4.5.1 Overview of economic evaluations of RRT

There have been many evaluations of the modalities of RRT, including some of the earliest published economic evaluations (Klarman et al 1968, Buxton et al 1975, Churchill et al 1984). Winkelmayer et al (2002) suggested one reason for the prolific analyses was the US government's decision in 1972 to reimburse all patients with ERF via Medicare. As shown later (section 4.5.3), and concluded in a review by Peeters et al (2000), most evaluations present inadequate information on costs.

Winkelmayer et al (2002) found that the cost per life year gained for in-centre haemodialysis had been relatively stable over 30 years. They undertook a meta-analysis of economic evaluations of RRT (published 1968 to 1998). The 13 studies varied by methods and by costs covered, and many omitted aspects such as patient transport costs and the time for patients or informal caregivers. Despite this, the costs per life year gained (in 2000 US dollars) were in a narrow range: $55,000 to $80,000. The authors concluded the stability over time was due to 'case mix' changes, as sicker and older patients underwent treatment. The authors also noted that the frequently cited thresholds of $50,000 per life year gained or $61,000 per QALY[34] originated from RRT.

Winkelmayer et al (2002) also argued for the cessation of analyses of single modalities because they were not true substitutes (due to medical reasons, provider preference or availability of donor organs). Instead, the authors made the case for analysis of the whole ERF programme or modality sequences, though the results might be less transferable between settings. This echoed an earlier call for 'life cycle' costing by Mallick (1997). Life cycle costs were defined as the total costs that individual patients accrued during their RRT, including modality changes, inpatient stays and co-morbidity. Mallick advocated this approach to aid comparison of treatment by type of renal disease and co-morbidity, although he acknowledged it would be difficult and time consuming. From a societal perspective, such analyses have merit to help determine the best (efficient) treatment pathways. They still require good information about the costs of modalities.

In 1998, the Department of Health commissioned the RSU study to evaluate the provision of satellite compared with MRU haemodialysis (Roderick et al 2005). The study found no significant differences in care processes, most clinical outcomes were similar, and it concluded that patients could be safely dialysed without on-site medical input. For reasons explained below, due to difficulties in comparing resource use on a like-for-basis between the two settings, it was unclear whether there were cost differences. Researchers faced the same challenge when they compared haemodialysis at home or hospital for NICE (see section 4.5.3).

4.5.2 Renal satellite evaluation (RSU) study and background to thesis

Motivation for the thesis came from the RSU study (Roderick et al 2005). A key question was how to attribute staff time at each MRU between patients deemed eligible and ineligible for RSU care by the senior nurse. Whilst this problem applied to all staff time, it was important for nurses because they undertook the bulk of day-

[34] Equivalent to approximately UK £34,500 and £38,300 respectively using Purchasing Power Parties to convert from $US to £UK at 2000 prices from http://www.oecd.org/dataoecd/61/56/39653523.xls.

to-day care. Figure 4.3 illustrates the challenge faced. The left side of the figure shows a renal satellite unit (RSU), a unit geographically distant from its parent main renal unit (MRU, shown on the right). Typically RSUs had little or no medical help available on-site and consequently the patients were relatively similar to care for (homogeneous) and straightforward to cost. MRUs had medical help on-site and the patient population was mixed (heterogeneous). Some patients were eligible for placement at a RSU. Others were ineligible for RSU care, for instance because they were too sick or unstable, and might have needed extra input from nurses. A like-for-like comparison between MRUs and RSUs therefore required the cost per patient for the two shaded areas in the figure.

Figure 4.3 RSU and MRU patient populations

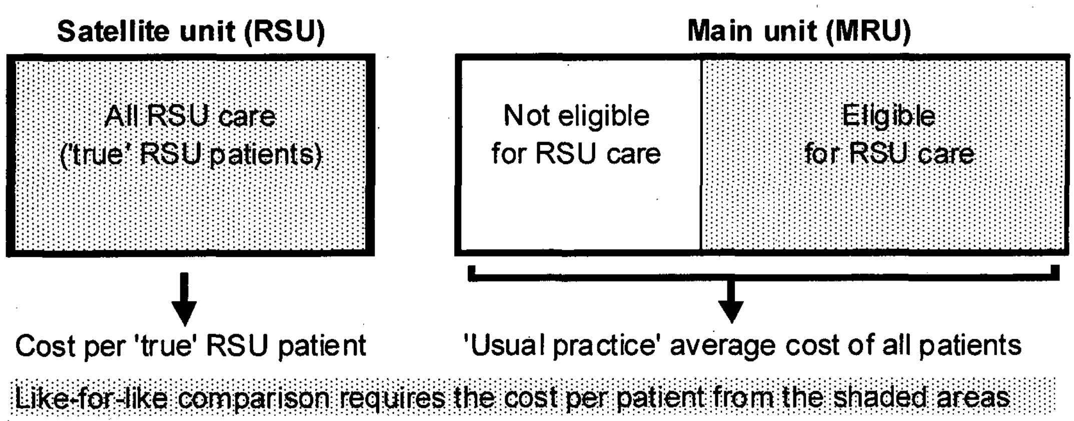


Table 4.3 shows the options considered about how to handle the potential patient heterogeneity. The study reverted to the 'usual practice' of averaging costs across all patients as the alternatives were unworkable. Since it was unlikely that patients needed the same nursing inputs, the average cost for both patient groups overestimated the cost for patients eligible for RSU care.

Table 4.3 Options for handling patient variation in nursing inputs in RSU study

Possible approach	Decision in RSU study
Directly measure nursing time.	Unfeasible. Data not routinely collected on nursing time by patient. Primary data collection would have incurred extra costs beyond those budgeted for the study (especially as there were 12 MRU-RSU pairs).
Attribute nursing time - weighting total nursing hours by the mix of patients using a classification tool.	Unworkable. Study units did not use a dependency or workload tool. Haemodialysis nurses were not aware of a suitable tool and considered scoring tools designed for inpatients inappropriate (personal communications in RSU study).
Attribute nursing time - weighting total nursing hours by workload factors from nurses in study.	Attempted. Opinions elicited from five key senior nurses at different units, but no consensus. Tried to identify and quantify time implications of potential factors demarcating patients who required more than average nursing time and link these to eligibility / ineligibility for RSU care. See Appendix 4.
'Usual practice' averaging costs across all patients (i.e. cost of eligible + ineligible instead of eligible alone).	Adopted as last resort because other options unfeasible in RSU study. Not ideal, but a typical straightforward procedure used in economic evaluations.

Straight comparisons of MRU and RSU patients supported the concerns about patient heterogeneity. An early study, albeit from the US, found MRUs had a higher percentage of patients in higher-severity groups than RSUs did (Plough et al 1984). The five severity groups were based on age, race, primary renal diagnosis, co-morbidity, and risk of mortality. Another study in the US compared nursing in MRUs and RSUs (Jones 1992). Data collection comprised patients' resource use and frequency of non-routine events over three months, and costs for events from four senior nurses' estimates of nursing time. Over the short study period, there were no differences in outcomes (deaths or hospitalisations). Non-routine events were more common in patients at MRUs than RSUs, but confounded by marked variations in clinical practice between MRUs and RSUs (e.g. type of vascular access, hours on haemodialysis, dialysers and fluids used, and medications). (In contrast, in the UK, a MRU manages one or more RSUs, typically using the same protocols and so nursing practice is more consistent.) The author concluded that patients at MRUs received more intensive and costly haemodialysis than patients at RSUs did. Conversely, she showed that (unidentified) patient factors interacted with clinical practice to affect outcomes, which in turn had an impact on nursing time.

Table 4.4 shows examples of associations between patient heterogeneity and outcomes or resource use. These links are expected but complex. Miskulin (2005) noted the lack of consensus on how to define a patient's risk or severity due to co-morbidity; tools vary by conditions covered, definitions, and weightings. She argued that studies should risk-adjust for co-morbidity (and other factors) since these could affect outcomes and quality of care. Although much evidence is from the US where patient and practice characteristics may vary from the UK, patient heterogeneity is clearly important. Apart from the study by Freund et al (1998) discussed below, none of the studies examined how patient factors affected staff time within haemodialysis sessions.

Table 4.4 Affect of patient heterogeneity in haemodialysis

Classification*	Outcomes for patients on haemodialysis
Functional activity: Karnofsky Performance Scale (KPS, Karnofsky and Burcherval 1949)	Predicted survival in patients beginning HD or PD (US: McClellan et al 1991 and Keane and Collins 1994; UK: Chandna et al 1999). Predicted survival in patients on maintenance HD (US: Ifudu et al 1998).
Co-morbidity: Wright / Khan index	Predicted survival in patients starting HD or PD (US: Wright 1991; UK: Khan et al 1993).
Co-morbidity: Charlson Co-morbidity Index (Charlson et al 1987)	Predicted admission rates, hospital days and costs in patients on HD or PD (US: Beddhu et al 2000). Predicted survival in patients starting HD or PD (Canada: Hemmelgarn et al 2003).
Co-morbidity: Lister score	Predicted survival, morbidity and higher hospitalisation rates in patients starting HD or PD (UK: Chandna et al 1999).
Multifaceted: Risk of Outcomes Adverse to Dialysis (ROAD)	ROAD included co-morbidity, functional status, social support, psychological status and health behaviour. High scores associated with hospitalisations and poorer quality of life (US: Lamb et al 2004).
Dependency / acuity: ANNA/MECON Patient Classification System	Predicted caregiver time during dialysis treatment sessions (US: Freund et al 1998, scale unobtainable).

Key: * Reference for tool cited if different from outcomes
HD = haemodialysis PD = peritoneal dialysis

Freund et al (1998) developed a patient classification system (PCS) based on time required from nurses or technicians for a chronic haemodialysis session. It is examined in detail here because it tried to involve all ESRD centres in the US in the first stage to develop the new PCS items. Nevertheless, reporting of the inter-rater reliability checks was poor. Twelve haemodialysis centres were chosen to cover a high and low cost per treatment unit from i) for-profit RSUs, ii) not-for-profit RSUs, and iii) MRUs. For the second stage, a co-ordinator at each centre estimated the time needed for each PCS item. These were averaged to develop preliminary acuity weights. Then, following a cross-sectional survey of patient's acuity scores, the researchers defined the five acuity levels and selected 10 patients at each acuity level. The third stage entailed nurses and technicians recording their time use using barcode scanners. Over 2-week blocks, the researchers aimed to collect patient acuity and total staff time for 50 designated patient-sessions per centre.

Data comprised 610 patient-sessions, as planned. Some patients' acuity changed between planning and data collection, and so data did not achieve the desired stratification by acuity (especially for levels IV and V). Staff time per patient increased with patient acuity at each centre, although not all the relationships appeared linear, possibly influenced by the low numbers at higher acuity. The researchers stated that, due to the unbalanced data, they used "log likelihood ratio analyses" based on percentages rather than minutes. Separate models found significant effects for acuity, centre and an interaction between type of centre and high-low cost unit; models of the other combinations were not significant. Whilst the authors acknowledged it was difficult to interpret these conflicting results, the explanation given was confusing. Overall, the average staff time per session increased sequentially from 61 to 97 minutes across the five acuity levels. One conclusion was that patients on haemodialysis were not a stable population, rather their needs for staff time varied considerably over short time intervals.

Limitations of the study were that it did not assess the validity of the time data, did not give standard deviations for the acuity-times, and ignored clustering within centres. It is unclear why multiple regression was not used to assess the effects of the variables simultaneously. The data collected did not include centre-level information that might have helped explain the variation in patient times within acuity levels across centres. For example, the mix of staff comprised registered and licensed nurses (equivalent to the old-fashioned state enrolled nurse in UK) and technicians, and was unlikely to be uniform. It was an ambitious study, but it is unclear if it was sufficiently powered for the analyses, given it only had two units for each centre type / high-low cost centre combination.

These examples have shown that patient characteristics are associated with variations in resource use. Since the RSU study was commissioned, the need for definitive cost-effectiveness information about MRU and RSU care has changed. RSUs offer improved geographical access and most existing MRUs do not have the space to expand. There is still a need to understand the resource use and cost implications of service and patient changes. For example, Ansell et al (2003) advocated development of some larger RSUs in England into MRUs. The predicted demographic changes are likely to increase the proportion of patients with co-morbidities. Moreover, Beech et al (2004) argued for attempts to reduce the cost per case (at the same quality), for example by skill mix changes. The RSU study faced the interlinked challenge of how to attribute staff inputs shared across patients and how to assess the impact of patient heterogeneity. Haemodialysis provides a useful example to test methods to cost services on a like-for-basis, but the challenges faced are relevant to many other evaluations introduced in Chapter 1. The next section reviews how other researchers had tried to handle these costing issues.

4.5.3 Handling of patient heterogeneity in renal economic evaluations

This section reviews how researchers have handled patient heterogeneity in economic evaluations of haemodialysis across different settings. First, it considers the appraisal of haemodialysis at home or hospital produced for NICE. Such evaluations are important because, since 2002, the NHS in England and Wales has been legally obliged to fund treatments recommended by NICE guidance (NICE 2006).

4.5.3.1 NICE appraisal of home and hospital haemodialysis

Mowatt et al (2003) undertook a systematic review and modelling for NICE to compare home and hospital haemodialysis. The evaluation faced similar problems to the RSU study, since not all patients at a MRU or RSU would be eligible for home haemodialysis. The reviewers gave strong warnings that the effect of setting was confounded due to patient and treatment factors. Home haemodialysis sub-groups tended to be younger and have less co-morbidity than MRU or RSU patients. Though home haemodialysis could be longer and more frequent than at units, many papers did not report the treatment regime. Despite noting major shortcomings in the evidence, the review concluded that home haemodialysis was more effective than MRU haemodialysis and "modestly" more effective than RSU haemodialysis was.

A general criticism of the evaluation is that it did not reflect current practice in England and Wales. Most studies were old (pre-1990), minimal evidence came from the UK, and self-care at hospital units was included as RSU haemodialysis even though it has no equivalent in the UK. Likewise, costs may have been unrepresentative. They came from a European study, but it is unclear whether from a single centre (Scotland) or the average of all centres including those in Eastern Europe[35]. In relation to patient heterogeneity, specific criticisms are that the cost per QALY model used average nursing costs at the MRU and RSU and ignored the acknowledged patient selection factors. Despite reporting increased carer burden for home haemodialysis, the model prevented patients from switching from home to MRU or RSU. It therefore omitted set-up costs to adapt the homes of patients who then changed setting.

Given the weak evidence, the NICE guidance was surprising. It ordered that all suitable patients (existing and new) be offered choice between haemodialysis at home or MRU/RSU (NICE 2002). The guidance declared that in the absence of robust evidence, haemodialysis at home was at least as effective as at hospital. It also stated that despite uncertainty in the RSU costs, overall cost differences should have been even more in favour of home haemodialysis.

4.5.3.2 Economic evaluations of haemodialysis in different settings

This section presents the systematic review of economic evaluations of haemodialysis in different settings (main unit, satellite, and home)[36] that was introduced in section 3.4.1.2. It examined whether researchers adjusted for potential

[35] The source cited, Valderrábano et al (1996), was incorrect and does not refer to the EURODICE study.

[36] Inclusion criteria: i) English language with abstract, ii) cost or economic evaluation studies that compared types of haemodialysis (e.g. hospital, satellite, home, self-care), iii) primary studies or modelling (i.e. excluded simple literature reviews), iv) published 1996 - 2007 (i.e. recent papers because of technology changes and reporting expected to be better).
Exclusion criteria: Papers that purely compared haemodialysis with peritoneal dialysis and/or transplant.
Search terms: (mesh: haemodialysis or hemodialysis) and (title or abstract: satellite or stand-alone or free-standing or hub and spoke or out-center or out-centre or home or self-care or self-assisted or (minimal adj care) or (limited adj care)) and (mesh: cost$ or economic$) where mesh=title, abstract, subject headings.
Databases: British Nursing Index & Archive 1985-Aug 2007; Cumulative Index to Nursing & Allied Health Literature 1982-Sept 2007 Wk 2; EMBASE 1980-2007 Wk 38; Health Management Information Consortium Sept 2007; MEDLINE 1950-Sept 2007 Wk 2; MEDLINE Daily Update and In-Process 21/9/07.
Results: 135 bibliographic details downloaded after exclusion of duplicates (all abstracts examined), 40 potential papers for data extraction examined, 20 papers for data extraction comprising 17 studies.

heterogeneity in patients' resource use or simply averaged costs across all patients. Appendix 5 shows the full data extraction table.

Table 4.5 summarises the key data extracted. The search found 17 studies (in 20 papers). Four were models based on secondary data sources (Gonzalez-Perez et al 2005, McFarlane et al 2006, Mohr 2001a and Mohr et al 2001b, Mowatt et al 2003). It was not clear whether Lim et al (1999) used primary or secondary data. The numbers of patients in analysis groups were often small and unbalanced. Overall, as found in the two previous chapters, standards of reporting varied. Information presented was rarely enough to gain a complete picture of the costing undertaken.

Whilst the majority of studies (14) included nursing costs, inclusion of medical and technician (maintenance) costs was inconsistent and less clear. Despite potential differences in staff travel between the settings, it was uncertain whether such costs were incurred or included in most studies (13). Most (13) studies included at least some costs of patient complications, although chiefly for hospitalisations.

Attention to possible patient heterogeneity varied. Five studies tried to ensure comparability of resource use between patient groups through choice of units or matching patients. Three studies adjusted for outcomes or co-morbidity, but not costs. Three studies mentioned patient heterogeneity issues, but made no adjustment. It was unclear whether patient heterogeneity was considered in the remaining six studies.

Most studies ignored the effect of patient heterogeneity on the costs of a haemodialysis session. Eleven studies simply used the unit's average costs per session. Only three studies mentioned apportionment of nursing costs across patients. Lee et al (2002) used unpublished information on nursing time from a patient management database, however, the source paper (Manns et al 2001) gave no details about the methods used to allocate nurses time. The study by Lindsay et al (2003) was reported in two other papers (Lindsay 2004 and Kroeker et al 2003), but none contained details of the workload tool used to assess nursing time. The third study used an unspecified method to measure direct care hours (Soroka et al 2005). In conclusion, it remained uncertain whether variation between patients affected the nursing costs for a dialysis session.

Table 4.5 Economic evaluations of haemodialysis across settings

Component	Costs included	Comment on methods to adjust for patient heterogeneity
Patients	Cost of complications at patient-level 13 Yes, at least some (predominantly hospitalisations) 3 No (Agar 2005, Bjorvatn 2005, Tediosi 2001) 1 Unknown (Piccoli 2004)	Actions to ensure comparability of resource use between patient groups 1 study ensured comparability between patients by the choice of satellite (Agar 2005) 4 studies used matched controls or cohorts (Lindsay 2003 and 2004 and Kroeker 2003, McFarlane 2002 and 2003, but only for drug and travel costs in Soroka 2005) 3 studies adjusted for outcomes or co-morbidity (De Wit 1998, Gonzalez-Perez et al 2005, Lee 2002) 3 studies mentioned patient heterogeneity issues, but made no adjustment (Roderick 2005, Tediosi 2001, Mowatt 2003) 6 studies unclear whether patient heterogeneity was considered (Bjorvatn 2005, Jassal 1998, Lim 1999, McFarlane 2006, Mohr 2001a and 2001b, Piccoli 2004).
Nurses	14 Yes 3 Unknown (Bjorvatn 2005, McFarlane 2006, Mohr 2001a and 2001b)	2 studies used workload measurement tool (Lee 2002, Lindsay 2003 and 2004 and Kroeker 2003) 1 study measured direct care hours, but method unspecified (Soroka 2005) No information in remaining 11 studies (assumed average costs used).
Doctors	10 Yes 7 Not included or unclear (Agar 2005, Bjorvatn 2005, Jassal 1998, Lim 1999, McFarlane 2006, Mohr 2001a, 2001b, Roderick 2005)	Costs mostly pay per contact or average treatment costs. Agar (2005) excluded costs on the basis that doctors' coverage of dialysis patients was part of their salary.
Technicians (maintenance)	6 Yes (Agar 2005, Lee 2002, Lim 1999, Lindsay 2003 and 2004 and Kroeker 2003, Soroka 2005, Tediosi 2001, Mowatt 2003) 2 No (Bjorvatn 2005, Roderick 2005) 8 Unknown	Average costs Lee (2002) Billed costs Soroka (2005). Remainder method unknown. No discussion about differences in maintenance costs for machines off-site, but these would incur technician travel time and costs.
Travel	Travel costs for staff included 1 Yes (Bjorvatn 2005) 1 Described (Roderick 2005) 2 Excluded (Gonzalez-Perez 2005, Tediosi 2001) 13 Unclear if incurred	Travel costs for patients included 6 Yes (Bjorvatn 2005, De Wit 1998, Jassal 1998, Mohr 2001a and 2001b, Mowatt 2003, Soroka 2005) 1 Described (Roderick 2005) 6 Excluded (Agar 2005, Gonzalez-Perez 2005, Lee 2002, Lim 1999, Lindsay 2003 and 2004 and Kroeker 2003, Tediosi 2001) 4 Unknown (McFarlane 2002 and 2003, McFarlane 2006, Piccoli 2004)

All papers: Agar et al 2005, Bjorvatn 2005, De Wit et al 1998, Gonzalez-Perez et al 2005, Jassal et al 1998, Lee et al 2002, Lim et al 1999, Lindsay et al 2003 and Lindsay 2004 and Kroeker et al 2003, McFarlane et al 2002 and McFarlane et al 2003, McFarlane et al 2006, Mohr 2001a and Mohr et al 2001b, Mowatt et al 2003, Piccoli et al 2004, Roderick et al 2005, Soroka et al 2005, Tediosi et al 2001.

Only first author shown in table

4.6 Conclusions

Haemodialysis provides a useful case study to assess the impact of patient heterogeneity on nursing costs. The treatment is life saving and costly. Both service commissioners and providers must work with limited resources to meet the expected increase in demand for haemodialysis. However, policy initiatives to improve the quantity and quality of haemodialysis provision conflict with other initiatives to improve patient choice. Moreover, renal services face further pressures given the increasing proportion of elderly patients with co-morbidities. Consequently, efficient use of staff, a major cost in haemodialysis, is important.

In general, costing staff in economic evaluations of haemodialysis has been poor. The effect of patient heterogeneity on staff time is unclear as most evaluations ignored the issue and simply used average costs. Given that patient factors such as co-morbidity and dependency affect hospitalisations, it is likely that patient factors may also be important in costing haemodialysis sessions. Having set the context, the next chapter outlines the empirical research.

Chapter 5 Introduction to empirical work

As an introduction to the empirical work, this chapter draws together findings from the earlier chapters. It gives the rationale for the methods chosen in relation to three aspects: use of haemodialysis as a case study, measurement of nursing time per patient, and classification of patient heterogeneity. After setting out the specific research questions addressed, it gives the rationale for data collection sites and piloting.

5.1 Haemodialysis as a case study

A key issue has been that resource use varies across patients with important consequences for costing comparisons. Chapter 4 outlined why chronic haemodialysis (HD) was chosen as a case study for the empirical work. HD is expensive, there is pressure to expand provision and nursing inputs are a major cost driver in HD. Yet most evaluations of HD in different settings (MRU, RSU, and home) have ignored patient heterogeneity and the affect it may have on nursing costs. The RSU study (Roderick et al 2005) suggested that at a MRU, compared with those eligible for RSU care, patients who were ineligible for RSU care would be sicker or more dependent and require more nursing input. Appropriate NHS costs were not available for the two patient groups and the study did not find a suitable method to weight the average (top-down) cost. The lack of suitable costs alone is a reason to adopt a bottom-up costing approach. HD also met criteria for which costing advice (section 2.2.2.2) proposed bottom-up costing, namely major staff inputs (cost driver) and staff shared between patient groups. Moreover, Wordsworth et al (2005) specifically recommended bottom-up costing for HD. For these reasons, the empirical work adopted a bottom-up approach to measure resource use.

5.2 Measurement of nursing time per patient

Routine chronic HD is characterised by intensive bursts of nursing activity at the start and end of each patient's session to gain vascular access, and connect or disconnect the patient from the dialysis machine. When on dialysis, nurse involvement usually drops, although patients require regular monitoring since HD alters the blood chemistry and can lead to sudden circulatory collapse or other adverse events. Poor vascular access and increased co-morbidity therefore pose additional challenges to getting the patient on to dialysis and keeping the patient stable. In addition, haemodialysis nurses require technical skills beyond those received in basic training.

The empirical study required an appropriate method to measure or attribute nursing inputs shared across patients. The primary focus was patients, rather than how nurses spend their time or the time taken for individual tasks. Chapter 3 showed there was no 'gold standard' technique to measure staff time; each method had advantages and disadvantages. With this in mind, self-recording by barcode scanning was selected for the empirical study. The method covers all aspects of nursing care: direct face-to-face care, indirect care (away from the specific patient), and general activity that does not relate to any particular patient. In addition, it allows collection of data by nursing grade, and so offers insight into the grade or skill mix of care delivered and hence costs.

From Chapter 3, barcode scanning appeared feasible, although it required time to perfect the downloading and data processing systems and studies often failed to undertake even basic validity checks. Therefore, the plan was to incorporate observer 'spot checks' to validate the data collected by the nurses. A question then arose - could the observer use work sampling both to validate the barcode data and as an alternative means to collect time data. Work sampling can cover all staff, rather than just individuals as in time and motion study and (unlike self-recording) does not disrupt nurses' work. The main disadvantages are that observers cannot easily capture indirect care for specific patients or activity by nursing grade. Since no examples of work sampling at the patient-level had been found, it was important to assess whether work sampling could link data to specific patients.

5.3 Classification of patient heterogeneity

The empirical work set out to investigate differences in resource use between patients deemed eligible or not for RSU care by the MRU senior nurse. The same method of designation had been used in the RSU study (Roderick et al 2005). The study had found no uniform eligibility criteria; RSUs varied by staffing, the presence of on-site medical cover, and geographical distance from the MRU. However, each MRU-RSU pair had acted as its own control. Whilst planning the empirical work, it became apparent that eligibility for RSU care even varied between RSUs managed by the same parent MRU. From anecdotal reports, eligibility was more restricted for some privately managed RSUs than their NHS counterparts. Moreover, eligibility for RSU care, although entered on computer, was not a label used in the nurses' daily practice. Hence, it appeared it would be useful to have another more objective method to categorise patients.

Methods to classify patient heterogeneity cover a wide variety of attributes, but need to be appropriate to the setting and acceptable to staff. The RSU study had not found a patient dependency system for outpatient HD in routine use, and this did not appear to have changed at the start of the PhD. Local nurses[37] judged there was still no accepted HD dependency tool and that other generic or specialist tools were unsuitable. Consequently, the initial plan was to investigate alternative tools used in the RSU study.

One tool was a functional assessment (the Karnofsky Performance Scale (KPS), Karnofsky and Burcherval 1949). The KPS was quick to complete and had face validity because it indicated patients who required increased levels of assistance. Other tools were the co-morbidity indices. These looked interesting since many version 3.5 HRGs and the reference costs used co-morbidity, although HD costs split predominantly by setting (MRU, RSU, home). The KPS and co-morbidity indices offered a practical means to assess the designation of eligibility for RSU care. Yet ratings were expected to be stable in the short-term and so their usefulness in relation to nursing inputs was less clear. It was also unclear whether local data (computer and nursing notes) were sufficient to generate the co-morbidity indices.

During 2005, plans changed because the local nurses developed a dependency-scoring tool and started to use it routinely (see next chapter). Although not validated, the tool seemed worth investigating.

5.4 Research questions addressed by empirical work

The previous sections have presented the rationale for the empirical work and data collection methods selected. Specifically, the empirical set out to answer the following research questions:

[37] At the Wessex Renal and Transplant Unit encompassing the main unit at Portsmouth (Queen Alexandra Hospital) and its satellite units.

1. What can be learned about the feasibility and data quality of barcode scanning to self-record nursing time per patient?
2. Is it feasible to use work sampling to measure nursing time per patient?
3. In chronic HD, is the nursing time per patient statistically or economically different between:
   i) patients who are eligible and ineligible for care at a renal satellite unit (RSU),
   ii) patients of different dependency?

5.5 Rationale for data collection sites and piloting

This chapter has raised a number of issues about the data collection methods and tools chosen for the empirical work. A pilot phase was therefore important to check the viability of collecting the relevant data. The purpose of piloting was:

1. To establish data collection using barcode scanners.
2. To assess whether work sampling could link data to specific patients and offer
   i) an alternative method to measure nursing time inputs, and
   ii) a means to validate the barcode data.
3. To test tools to distinguish between patients undergoing chronic HD (KPS, co-morbidity indices, dependency scoring-tool).

Based on a number of factors (see Table 5.1), two sites were chosen for piloting. The first site was local - the Southampton University Hospitals NHS Trust (SUHT) - on a 24-hour, 24-bedded surgical ward (F9) for patients undergoing upper gastrointestinal surgery (oesophagus, stomach, liver, etc). For technical reasons, it was preferable to test barcode scanning on-site with full access to research resources. In addition, the researcher needed both to gain experience of work sampling and to test its feasibility whilst supporting nurse barcode scanning. The ward was attractive because patient throughput was less and slower than in the HD units. It also offered insight into implementation of barcode scanning in another setting, however, the timing of piloting at SUHT was not ideal. The ward was in the midst of a redesign in preparation for a planned, but repeatedly postponed, ward move.

Table 5.1 Factors influencing choice of data collection sites

	Pilot (SUHT)	Pilot (Totton)	Main (Portsmouth)
Setting/patients	Ward (not HD)	HD (RSU only)	HD (MRU i.e. RSU eligible and ineligible)
Location	On-site	5 miles away	20 miles away
Delivery of care	24-hours 7 days a week	6:15-22:30 Mon-Sat	6:30-24:00 Sun-Fri 6:30-14:00 Sat
Size/layout	2 bays of 4 beds 2 bays of 6 beds 4 side rooms	Large room with 9 bays 1 side room	Large room with 22 bays 3 side rooms
Access to research facilities	Full	'None' (use of staff room)	'None' (occasional desk space in unit, room in residences)
Patient throughput	24 patients Some admissions / discharges daily e.g. for 2-3 patients at variable times	10 patients in 3 cohorts per day (i.e. 30 per day) with 'fixed' changeover times of 20 patients twice daily	25 patients in 3 cohorts per day (i.e. 75 per day) with 'fixed' changeover times of 50 patients twice daily
Duration of stay	'24-hours'	Typically three times per week for HD sessions lasting 4-5 hours	
Nurses and working patterns	~20 mostly 12-hour shifts	~20 mixture mostly 12 or 7.5-hour shifts	~40 all 7.5-hour shifts
Workload (patient-nurse interactions)	Fairly constant (less at night and staff handover times)	Sporadic when getting patients on/off dialysis, then low intensity	

The second pilot site was the Totton haemodialysis unit (Southampton). This 10-bay RSU provided outpatient chronic HD for 54 patients, without on-site medical support, but under the management of the MRU at Portsmouth. The aim was to assess the feasibility of data collection in a small local unit before trying to apply the techniques in the larger geographically distant parent unit. It was an opportunity to test various patient dependency and co-morbidity tools to differentiate between patients undergoing HD, although patients were expected to show less variability than those at the MRU.

It was not possible to proceed with the original plans. By the time piloting could start at SUHT, nearly a year had lapsed since the HD units had agreed to the study. Due to financial pressures on the renal service, the HD nurses felt the research might be of immediate mutual benefit. Therefore, since barcode scanning required active involvement of the nurses, set-up work for Totton started 'immediately' after piloting at SUHT. Consequently, only brief inspection of the SUHT data was possible.

The main data collection (Portsmouth study) took place in the MRU at Queen Alexandra Hospital, Portsmouth. This was a 25-bay unit providing outpatient chronic HD to approximately 145 patients, with on-site medical support.

The study was approved by the Southampton and South West Hampshire Research Ethics Committee (A) (05/Q1702/83) (see Appendix 6 for details). The researcher and observer referred to in the empirical work was Tricia Nicholson.

5.6 Summary

This chapter has provided the rationale for the empirical work and methods chosen based on the findings from earlier chapters. In summary, the purpose of the empirical work was both to evaluate data collection methods and to measure the nursing time and costs per patient for patients at a main renal unit who were either eligible or ineligible for renal satellite unit (RSU) care. The next chapter describes the pilot phase.

Chapter 6 Methods for piloting

It was important to test the systems to measure nursing time and to classify patient heterogeneity. This chapter describes the methods tried in the pilot phase: barcode scanning and work sampling to measure nursing time, and three tools to classify patient heterogeneity in those undergoing chronic HD.

6.1 Measurement of nursing time inputs for haemodialysis

This section details the methods for barcode scanning and work sampling. Nurses were the study subjects in order to capture all their working hours and hence all patient care. Here, the term nurse encompasses staff such as health care support workers. All patients were included.

6.1.1 Barcode scanning

Preliminary work involved sourcing and setting up both the barcode scanners and the necessary software. Appendix 7 reports the initial lessons learned. In each setting, a first step was to agree with nurses how to code activities to the three activity categories (direct care, indirect care, and general activity - see Appendix 8 for examples). The aim was to maximise consistency across nursing shifts and minimise disruption to the nurses' workload. Figure 6.1 shows an example of the barcodes[38] that were produced using Loftware Label Manager 7 (2003). Barcodes were available beside each patient, with complete patient-lists in the clinical areas and temporary unallocated barcodes for new patients. Nurses 'pre-piloted' the barcode scanning and, based on their feedback, changes were made to the paperwork.

Figure 6.1 Example of barcode label

FirstName Surname

Direct care 11001156 Indirect care 11002156

156

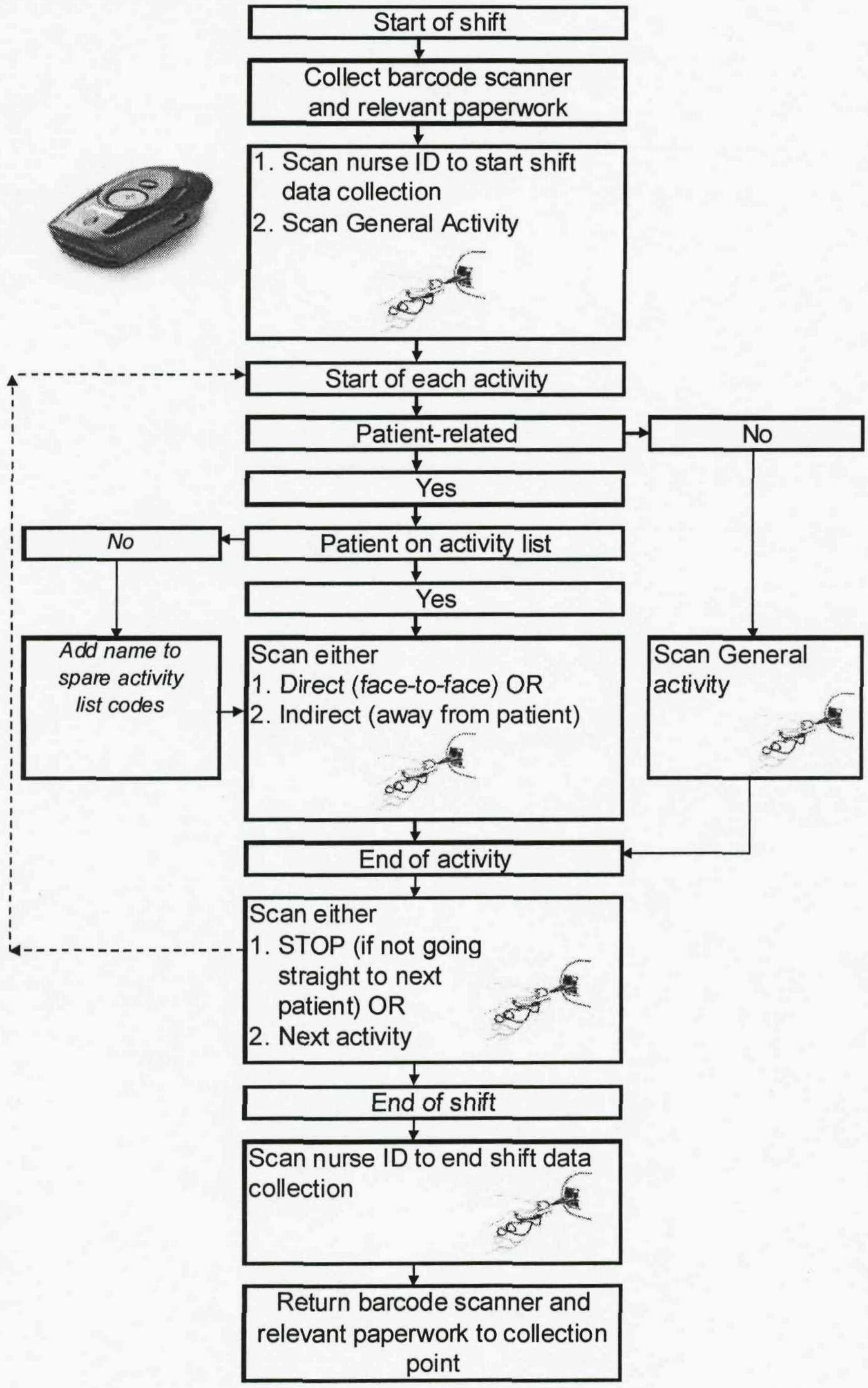
STOP or General activity 10003000

[38] Appendix 1 provides technical details about the type of barcode chosen (Code 128).

The barcode scanners were Symbol CS 1504 Consumer Memory Scanners. Before starting data collection, each barcode scanner's internal clock was synchronised to a radio controlled watch and time keeping was checked at each data download.

Figure 6.2 gives an overview of the barcode scanning process. At the start of the shift, nurses collected a scanner and relevant paperwork and scanned their identifier barcode. When nurses started an activity for a different patient, they scanned the relevant patient's barcode for direct or indirect care as appropriate (according to whether they were with the patient or not). For tasks that were not patient-specific, the nurses scanned the single 'Stop or General activity' barcode. Scanning the next barcode signalled the end of the previous task. At the end of the shift, the nurses returned the scanners and 'scanned off' using their identifier barcode. All nurses had access to individual or communal 'comments' sheets to record data corrections (e.g. for incorrect barcodes scanned, missed meal breaks or scanner problems). These comments sheets enabled each nurse's working hours to be calculated taking into account typical meal breaks.

Figure 6.2 Barcode scanning process



6.1.1.1 Recruitment and consent to barcode scanning

The researcher held meetings with the nurses to explain about the study and distribute an information sheet (see Appendix 9). At least 24-hours later, each nurse was asked to consent to enter the study (see Appendix 10). All nurses were eligible to enter the study unless there was insufficient time for the 24-hour 'cool-off' period. This excluded temporary nurses (agency or students), those who worked infrequently, and some nurses returning from long-term sick leave. The researcher held brief training sessions for the nurses, in addition to providing written instructions. Nurses' feedback was sought during visits to the ward or unit as this had proved important in previous studies.

6.1.1.2 Data management and statistical analyses

Barcodes were downloaded as text files using MiniPro 1.0 (2004) and then imported into Excel (2003) for processing and cleaning. Other quantitative data were entered into Excel spreadsheets. Subsequently, data were exported by Stat/Transfer 8 (2003) into SPSS 14 (2005) and Stata 9 (2005) for analysis.

Standard summary (e.g each patient's mean) or descriptive statistics and graphs were used to describe the baseline data and perform validity checks. In common with usual practice in health economics, data were presented as means rather than medians even though positively skewed. The mean, unlike the median, can be used to calculate the budgetary impact (i.e. the number of patients treated multiplied by the mean cost). Tests of associations between time or dependency and other variables used non-parametric correlation coefficients (Spearman's rho). These were interpreted using Altman's (1991) guidance on strength of agreement.

6.1.1.3 Data validity

The first step to check the validity of barcode scanning was to assess the completeness of data collected. In addition, the researcher planned to use the work sampling observations (see section 6.1.2) to check the accuracy of the nurses' recordings. Potential threats to data validity were if study nurses:

- Failed to collect scanner at the start of shift
- Failed to scan off at the end of shift
- Failed to scan when first starting an activity (as this 'stopped' the last activity), with the following possible consequences:
    - Previous activity too long or next activity too short
    - Possible missed activities
- Scanned the incorrect barcode:
    - Direct instead of indirect care or *vice versa* (for the correct patient)

- General activity instead of patient-specific time (direct or indirect care)
- Wrong patient
- Or multi-tasking and had to decide which patient to attribute the time to.

6.1.1.4 Feasibility of barcode scanning

Data on feasibility issues comprised descriptive information on the practicalities of implementation, research effort and equipment costs. In addition, a self-complete questionnaire and general feedback comments were used to assess the acceptability to nurses of barcode scanning (and work sampling or observation). The questionnaire included questions adapted from a questionnaire used in a previous unpublished study in primary care (Taylor et al 1998). Appendix 11 describes how this was developed further during the SUHT study. All study nurses at SUHT and Totton received the questionnaire at the end of the data collection period.

6.1.2 Work sampling

Typically, work sampling starts by estimating the number of observations needed based on relative workload from informed guesses or piloting, as here. Therefore, data collection was pragmatic and geared to what was feasible in the time available.

Due to the size and layout of the ward and unit, sampling frequencies of 15 and 20 times per hour were achievable. These varied as it took approximately two minutes to walk through the ward, but the HD unit was geographically smaller. Random observation times for one-hour blocks were generated in Microsoft Office Excel (2003). Table 6.1 outlines the components of the work sampling schedule.

Table 6.1 Work sampling schedule

	SUHT	Totton
Overall duration	1 week	2 weeks
Observation periods	8 hours out of 24 in blocks of 1-5 hours to capture part of each day & night shift	'Continuously' for 7 to 8½ hours out of 16 hours (1st week 'mornings', 2nd week 'evenings')

Usually work sampling involves recording observations on paper, but in this study barcode scanners were used to facilitate management of the large volume of data. To validate the barcode scanning time data, it was necessary to record when a non-study nurse was with a patient. Without this, it was impossible to work out whether disagreements were due to unrecorded activity, or care delivered by a non-study nurse. Appendix 12 shows the work sampling barcodes and paperwork used in the Totton study.

The researcher tested work sampling over several single hour blocks. At the designated start time, the observer scanned the most appropriate barcode for each nurse. At SUHT, observations were made whilst walking through the ward, whereas at Totton it was necessary to stay mainly in one position from where the clinical area was visible. Additional notes recorded which nurses and patients were present at the time of work sampling. Tracking patients was fundamental to ensure that the work sampling observations linked to the relevant patient IDs. For example, the work sampling data recorded:

- Observation time 1
    - 1 nurse with patients A, C and D
    - 2 nurses with patient F
    - 1 nurse not with a patient
    - 6 nurses expected (according to duty rota)
- Observation time 2
    - 1 nurse with patient B
    - 2 nurses with patient F
    - 3 nurses not with patients
    - 6 nurses expected

It was planned to estimate the nursing time per patient from the work sampling observations following the procedure outlined in Table 3.1.

Potential threats to data validity for work sampling were if the observer:

- Deviated from a random observation schedule e.g.
    - systematically started observations early or late (other than by chance)
    - systematically missed observations at certain times
    - failed to undertake sufficient observations
- Coded the activity wrongly, e.g.
    - missed seeing a nurse (especially if with a patient)
    - coded wrongly (patient or activity)
- Inaccurately coded or calculated nursing hours, patient hours of both if
    - observer could not find expected number of nurses, or patients, or both

- observer could not directly see either the patient or nurse (e.g. if behind curtains), leading to wrong coding if incorrect assumptions made rather than relying on activity or contacts actually observed
- patient away from expected location and therefore could not be identified.

Work sampling received unit-level consent from the senior ward sister and Modern Matron for HD. Although individual nurse consent was not required (see Appendix 6), posters informed staff and patients that a study was underway.

6.2 Classification of patient heterogeneity

The following sections describe the methods to classify patient heterogeneity - the patient dependency-scoring tool, Karnofsky Performance Scale and co-morbidity indices. These only applied to the patients on HD (i.e. at Totton).

6.2.1 Patient dependency-scoring tool for outpatient HD

The haemodialysis nurses[39] had developed the dependency-scoring tool for outpatient HD from first principles. Although termed patient dependency, based on the model described by Morris et al (2007, section 2.2.3.4), the tool also included nursing intensity. Nineteen items covered three aspects - "HD Assessment" (i.e. pre-dialysis), "Risk Assessment", and "During HD Treatment". Some items were about actual care delivered (e.g. ease of vascular access, level of assistance required to mobilise, wound dressings undertaken). Other items were about the potential need for nursing care by severity of illness, or the potential to deteriorate (e.g. grading of blood pressure control). Scoring levels on individual items varied, usually from zero (routine care) to three (greatest actual or potential input required).

There were no validity or reliability data for the dependency tool. It appeared to have face validity because it covered aspects deemed important in the RSU study (Roderick et al 2005). Validity and reliability checks carried out as part of the main data collection are described in section 8.3.

[39] At the Wessex Renal and Transplant Unit encompassing the main unit at Portsmouth and its satellite units.

In routine use, a registered nurse (RGN) rated the patient's dependency at each HD session and entered the total score on computer (Proton). During the study, the nurses also completed a paper record for each patient to provide the component information (see Appendix 13). Table 6.2 shows the nurses' interpretation of the overall dependency scores.

Table 6.2 Interpretation of dependency scores

Dependency scores	Level	Recommended nurse to patient ratio
0 to 5	Low	1 to 3
6 to 10	Mid	1 to 2.5
≥ 11	High	1 to 2

Source: Wessex Renal and Transplant Unit (unpublished)

6.2.2 Karnofsky Performance Scale (KPS)

The Karnofsky Performance Scale (KPS) measures functional activity and scores range from normal to dead in 11 categories (Karnofsky and Burcherval 1949). Although developed to assess patients with cancer, the KPS has been used for patients with established renal failure (described below). Previous research showed the KPS was a useful predictor of survival in patients beginning HD or peritoneal dialysis (McClellan et al 1991, Keane and Collins 1994, Chandna et al 1999); and predicted survival in patients on maintenance HD (Ifudu et al 1998).

Hutchinson et al (1979) assessed inter-observer variability between two pairs of doctors rating the KPS in 29 patients in Accident and Emergency (A&E) and 31 patients on chronic HD. They found moderate agreement for the three scoring groups[40] (kappa statistic 50% for A&E and 46% for HD patients). A further test found poor agreement between each doctor's ratings and those of patients on HD (kappa 17% and 11% for each doctor). The authors argued the scoring problems were due to lack of operational definitions and aggregation of multiple aspects within score levels. Whilst valid criticisms, they may be less serious for patients on HD as the doctors' ratings mostly varied by only one category. Conversely, for patients in A& E the variation between the doctors' ratings was more diverse. This was unsurprising since the doctors may have known the patients on chronic HD, whereas the doctors may have just met the patients in A&E. Furthermore, the doctors and patients did not use the same scoring system. Instead of direct rating,

[40] Scoring groups A= 0-40, B = 50-70, C= 80-100.

patients' responses to three questions about their ability to carry out normal activity, work and care for themselves were mapped to the scale. Differences were also predictable given that doctors had to judge patients' abilities that were unobserved in the consultation. Ifudu et al (1994) developed a modified version of the KPS with different intervals and 14 categories; however, the original score remains widely used (Roderick et al 2005).

The KPS had face validity, and it had been quick and simple to rate in the RSU study (Roderick et al 2005). The main purpose for the KPS was to assess both the designation of eligibility for RSU care and the validity of the dependency-scoring tool (see section 8.3). Compared with the dependency-scoring tool, the KPS contained fewer markers of nursing inputs. The ratings were likely to be stable in the short-term and therefore insensitive to small (daily) changes in the patient's condition. Therefore, unless validity problems precluded use of the dependency-scoring tool, there were no plans to examine nursing time by KPS.

At Totton, a senior nurse rated each patient's KPS once a week at the patient's first dialysis session after their weekend break (see Appendix 14).

6.2.3 Co-morbidity indices

Two co-morbidity scores had been used in the RSU study[41] (Roderick et al 2005). The Wright/Khan score, shown in Appendix 15, has three categories - low, medium and high risk. It was found to predict survival in patients starting HD or peritoneal dialysis (Wright 1991, Khan et al 1993).

The Charlson Co-morbidity Index, shown in Appendix 15, weights the number and the severity of co-morbid diseases and covers more conditions than the Wright/Khan index. Initial research in medical patients showed the Charlson index predicted the risk of mortality from co-morbid disease (Charlson et al 1987). In patients on HD or peritoneal dialysis, the index predicted admission rates, hospital days and costs (Beddhu et al 2000). Hemmelgarn et al (2003) found that modifications improved its ability to predict survival in Canadian patients starting HD or peritoneal dialysis.

[41] A third co-morbidity score used in the RSU study, the Lister score (Chandna et al 1999), was excluded here because it required detailed information on each patient and access to medical notes.

Inter-rater reliability between two nurses was very good (kappa 0.93) in a study by Bernardini et al (2004).

Similarly to the KPS, co-morbidity scores were likely to be stable in the short-term but they offered a practical means to assess the designation of eligibility for RSU care. Both indices included age, which from the RSU study (Roderick et al 2005), had not been considered a good marker for nursing inputs in itself. Since it was unclear whether the HD unit could provide accurate scores from summary data, a first step was an audit of data quality for 30 patients at Totton, which was undertaken by the Modern Matron for HD.

6.3 Summary

The key purpose of piloting was to establish the data collection methods. This chapter has detailed how it was planned to use barcode scanning and work sampling to measure the nursing time per patient. It has also described the three methods selected to classify patient heterogeneity in patients on HD. The next chapter presents findings about the feasibility of implementing the methods from piloting on the SUHT ward and Totton HD unit.

Chapter 7 Results of piloting (SUHT and Totton)

Piloting had three purposes: to establish data collection using barcode scanners, to assess work sampling, and to test the methods to classify patient heterogeneity. To avoid repetition, the findings from both SUHT and Totton are presented together rather than sequentially. Table 7.1 gives an overview of the patients at each setting.

Table 7.1 Patients' characteristics (SUHT and Toton)

	SUHT	Totton
Number	64	54
Type of patient	Surgical	Chronic HD
Male	48%	61%
Age	Data not collected	Mean age 59 years (SD 16) Range 20-82 years *39% aged 65 years or over*

7.1 Establishing barcode scanning data collection

The preliminary lessons learned about sourcing and setting up barcode scanners are reported in Appendix 7. In summary, the initial set-up was technically challenging and more time consuming than expected. Obtaining full ethical and research governance approval took over 7 months, in addition to the challenges highlighted in Chapter 5 about gaining access to the ward and the HD units.

Nurses were under considerable work pressure at both SUHT and Totton. Staff shortages on the SUHT ward necessitated extensive use of agency nurses and overtime. At Totton, there was unease amongst staff due to the 'imminent', but repeatedly postponed, opening of a new RSU to which some patients and nurses were due to move. Overtime was restricted due to financial pressures and available working hours had decreased due to nurses' increased annual leave entitlement through the national "Agenda for Change" re-grading.

This section presents an overview of the barcode scanning data collected, recruitment of nurses and completeness of data. It discusses the lessons learned from piloting about implementing barcode scanning, and presents the results of the nurses' acceptability questionnaire.

7.1.1 Overview of barcode scanning and data validity

Table 7.2 summarises the barcode scanning data collection. The planned 1-week 'pre-piloting' phase at Totton was extended by 2 weeks, though on an informal basis, as pressures on staffing levels delayed the start of the data collection.

Table 7.2 Barcode scanning data collection (SUHT and Toton)

	SUHT	Totton
Ward / unit operation	24-hours a day 7-days a week	16.25 hours per day 6:15-22:30 Mon-Sat
'Pre-piloting'	3 nurses, 2.5 hours each 2 Nov 2005	All nurses, 3-weeks 13 Jan - 4 Feb 2006
Data collection dates	5-12 Nov 2005 & 16-23 Nov 2005 i.e. 14 days with 3 day gap	6-18 Feb 2006 i.e. 12 consecutive days

Staff at both sites were a mixture of registered general nurses (RGNs) and health care assistants (HCAs). All 'available' nurses were recruited to barcode scanning as shown overall in Figure 7.1, and by nursing grade in Table 7.3.

Figure 7.1 Recruitment of nurses (SUHT and Toton)

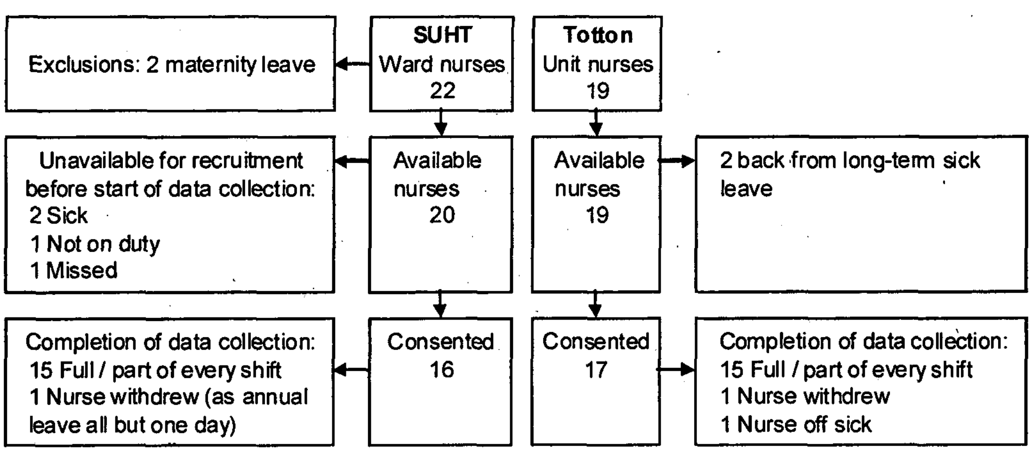
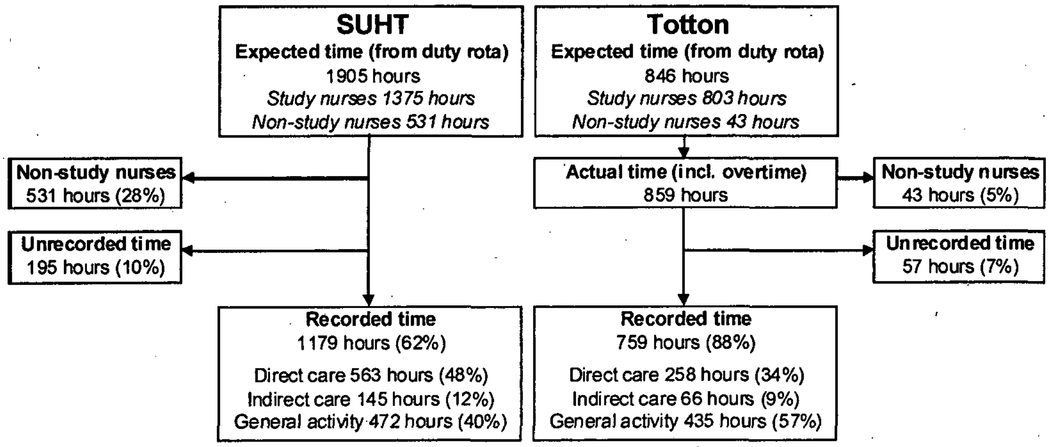


Table 7.3 Recruitment of nurses by grade (SUHT and Toton)

Grade	Available nurses (number, % in study)	
	SUHT	Totton
RGNs (D to G)	14 (13, 93%)	9 (9, 100%)
HCAs (A to C)	6 (3, 50%)	10 (8, 80%)
Total	20 (16, 80%)	19 (17, 89%)

Figure 7.2 shows the nursing hours expected from the duty rota and the hours recorded by nurses. At SUHT, completeness of data recording by study nurses was good; they recorded 1179 hours or 86% of their expected hours. Yet this only represented 62% of the overall hours, mainly due to reliance on agency staff who contributed to the 28% of hours for non-study nurses.

Figure 7.2 Barcode scanning time recorded (SUHT and Toton)

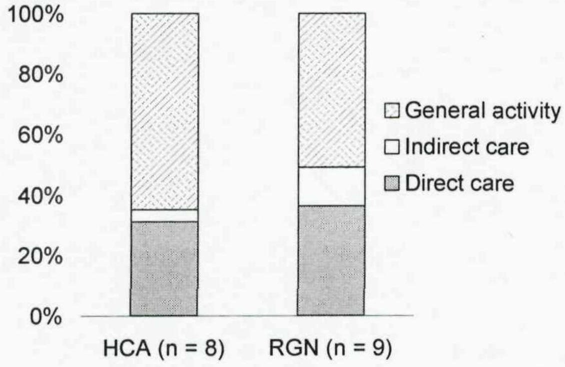


Completeness of data recording by study nurses was also good at Totton; they recorded 88% of actual hours. Actual time was derived by comparing each nurse's scanned data with the duty rota. It identified unrecorded time at the start of the shift, for example, when the nurse forgot to start scanning, or end of the shift when the nurse finished early. It also identified recorded, but unscheduled 'overtime' when the nurse finished late and so differed from expected time. Missing data, as a percentage of actual hours, comprised the following:

- 5% for non-study nurses
- a small amount of time (7%) unrecorded:
    - 1% when study nurses consciously stopped scanning,
    - 2% for the study nurse who withdrew from the study,
    - 4% that should have been recorded as general activity at the start or end of shifts and occurred when a study nurse forgot to start barcode scanning, stopped scanning early, or went home early (particularly on the smaller dialysis rotations on Tuesdays, Thursdays and Saturdays evenings).

At Totton, patient-specific time (i.e. direct and indirect care) accounted for 43% of overall time recorded, comprising 35% of HCAs' time and 49% of RGNs' time, as shown in Figure 7.3. A difference between nurses was predictable as the HCAs undertook much of the general activity tasks such as stock control and preparation of dialysis consumables and equipment.

Figure 7.3 HCAs and RGNs - proportions of time spent on tasks (Totton)



For reasons explained later (section 7.2), it was not possible to use the work sampling data to check the accuracy of nurses' recordings, although other validity checks were planned for the main data collection (see section 8.2).

7.1.2 Feasibility of barcode scanning

During piloting, a number of issues became apparent about the practicality of using barcoding technology to measure nursing time. Some missing data or scanning mistakes[42] had been expected, as outlined in section 6.1.1.3. Whilst it was possible to quantify unrecorded time at the start or end of shifts, other missing data or mistakes were less easy to detect because the nurses rarely used the comments sheets. Under reporting seemed likely as nurses tended to report mistakes in person and assumed the researcher was aware of 'emergencies' even when absent. At Totton, approximately 0.5% of barcode scans may have been mistakes as they were direct care for patients on non-dialysis days. Some may have been due to confusion over names, as four patients with the same first name attended at the same time. Attempts to improve reporting included increased training, regular visits by the researcher, and revision of the comments sheet with more tick boxes and prompts.

Although the activity codes (direct care, indirect care and general activity), appeared simple, they were not always clear-cut to apply. As expected from section 3.1, to

[42] The original plan was to correct the dataset for reported mistakes and hence avoid attributing time to the wrong patient, wrong activity, or both.

avoid contamination of scanners, some activities coded as direct care encompassed indirect care too. Despite the agreed decision rules, coding errors or inconsistencies were apparent though infrequent. Differences in working practices meant it was not possible to define activity codes uniformly across settings.

Other unforeseen issues arose. It was very challenging for one researcher to manage the barcode scanning alongside work sampling. At SUHT, printing of barcode lists (at least daily) required a ward visit to check patients and locations, and then typically 30-45 minutes for printing. At Totton, printing of barcodes was far less onerous because the patient population was stable. However, the researcher had to speak to every patient because, in the small unit, patients were highly aware of an 'outsider' and some feared being forced to go to another unit.

Scanner problems proved very time consuming. Although new, 18 (36%) scanners appeared to need new batteries before use at SUHT. The scanners did not give the low battery warning, a few batteries were found to be corroding, and some scanners' clocks became fixed at '63' seconds and then only recorded to the nearest minute. At Totton, 14% of barcode scans had fixed '63' seconds, which caused 2% of scans to have zero duration. All batteries were changed before data collection at Totton and then as required, but this incurred extra unexpected costs. Later it became apparent that fixation at '63' seconds, an ongoing problem in 11 scanners, was not simply a battery problem. Minor scanner problems continued and one failed completely. Static electricity, reported by Connell et al (1997), may have caused problems since occasionally scanners recorded 'random' times (and at Portsmouth, data for one nurse's shift was lost because the scanner would not download.)

Another problem was the scanners' poor time keeping; some ran slow (typically by up to one minute) and a few ran fast (by less than one minute). This was surprising, as one would not expect even a cheap wristwatch to lose or gain more than a few seconds in a week. The problem had no major effect on the nurses' time data per se. Conversely, it prevented use of the observer-recorded data (also collected by barcode scanner) to check the accuracy of the nurses' recordings (discussed further in section 7.2). To alleviate the problem, more frequent re-setting of the scanners' clocks was planned at Portsmouth.

Lastly, neither the SUHT ward nor Totton HD unit kept records of which nurses cared for which patients. This hindered attempts to resolve data discrepancies and made it impossible to attribute the non-study nurses' time to specific patients. It was expected to be less problematic at Portsmouth where nurse-patient allocation sheets were routine.

Barcode scanning required specific equipment and software that cost £6260 (see Appendix 7). Of this, £5260 was expected, but £640 (11%) was unexpected expenditure mainly for software support and the numerous battery changes. Due to the ward and unit hours, the researcher stayed on-site in the hospital residences at SUHT and Portsmouth to manage data collection, which cost a further £630. Additional costs incurred, but not quantified, were for consumables (paper or stationery, printer cartridges, CDs to backup data, etc.), computer equipment (laptop, USB hubs, laser printer, etc.), travel to meetings to set-up data collection, and researcher time.

7.1.2.1 Acceptability of data collection methods to nurses

The final feasibility issue assessed was acceptability of the data collection methods to nurses. A questionnaire was distributed to 30 nurses (15 at SUHT and 15 at Totton). All questionnaires were returned and the full responses are shown in Appendix 16. All nurses found the barcode scanners easy to use and generally reliable, and over 80% of nurses were confident using the barcode scanners after two shifts. Over 80% nurses stated they usually scanned at the start of the activity, although 73% stated they sometimes scanned during an activity. From anecdotal reports, nurses felt their scanning became less accurate as the shift progressed (especially as most SUHT nurses worked 12-hour shifts), as they forgot whether they had scanned activities.

Nurses did not feel that either barcode scanning or observation intruded unacceptably in their relationship with patients. Over 70% nurses felt they acted normally in the presence of an observer, although three nurses felt that observation reminded them to scan. Everyone thought the potential information was useful to the team, colleagues or hospital managers. One nurse was not happy to take part in another barcode scanning project, whilst three nurses (10%) were not happy to be observed again for work sampling. In conclusion, nurses' acceptance of both barcoding scanning and work sampling was high.

7.2 Assessment of work sampling

Piloting tested the feasibility of work sampling to measure nursing time per patient both as an alternative method to barcode scanning and as a means to validate the barcode data. At SUHT, due to the set-up time for barcode scanning, it was only possible to collect one-week's worth of work sampling data (covering 56 hours with 15 sets of observations per hour). At Totton, work sampling observations covered 94 hours with 1885 sets of observations (i.e. 20 per hour). Whilst Appendix 17 gives fuller details of the lessons learned from piloting, this section summarises the findings.

Work sampling proved unworkable for the main data collection. Four factors made it hard to link observations to patients. First, the researcher did not know the patients and so could not identify them if they were not at their bedside or allocated dialysis station. Second, a related problem was that in the HD unit there were large and rapid changeovers of patients (20 at a time). One solution, though beyond the scope of the study, was to use off-duty nurses from the ward or unit, although the lack of independence might make observations more prone to bias from mistakes or omissions. Third, the geographical layout and use of curtains for patient privacy made it difficult to observe and hence track both nurses and patients. Fourth, in order to minimise mistakes, the work sampling observation periods needed to be shorter. This was achievable if undertaken by more than one observer or by extending the data collection period, but neither option was possible in the current study.

Even if work sampling had successfully linked data to specific patients, the data were still unsuitable to validate the nurses' barcode scanning data. This arose due to the scanners' time keeping (discussed in section 7.1.2) and meant it would have been inaccurate to compare data recorded on different scanners. One option was to use a wide time window to assess the agreement between the observer-recorded and nurse-recorded barcode data, but this posed two problems. First, since the scanners' internal clocks did not remain synchronised, it was unclear which scanner should be the benchmark. Second, a wide time window risked finding multiple activities recorded, making it difficult to decide what constituted agreement. Furthermore, the researcher could not find a method to automate the data matching and so each scan needed checking manually, a very time consuming process.

Therefore, work sampling was dropped for data collection at Portsmouth and a different data validation procedure was used instead (see section 8.2).

Although unfeasible in this context, responses to the acceptability questionnaire (section 7.1.2.1) showed that work sampling was generally acceptable to nurses; only three of 30 nurses were unhappy to be observed in future research. On the other hand, nurses knew that work sampling was being used to check their data and not purely a different means to collect time data. In common with the unpublished study in primary care (Taylor et al 1998), nurses were slightly more negative towards data collection by an observer than self-recording using barcode scanners.

Overall, work sampling had been a secondary choice to measure nursing time per patient and was unfeasible for the current study. Compared with barcode scanning, key limitations were the inability to record most indirect patient care and to collect data by nursing grade. Yet key advantages were that it only required unit-level, not individual consent and took less time to set-up partly because nurses did not need training. Thus, in multi-centre studies, work sampling might be more practical than barcode scanning for measurements not at the patient-level. It was concluded that it might be possible to use work sampling to collect patient-level data under very restrictive conditions. Examples might be where patients are immobile (e.g. intensive care, special care baby units, etc.), where the observer can identify the patients (e.g. in stable patient populations (low throughput) such as long-term care or through use of photos). Nonetheless, it would be extremely challenging.

7.3 Testing methods to categorise patient heterogeneity (Totton)

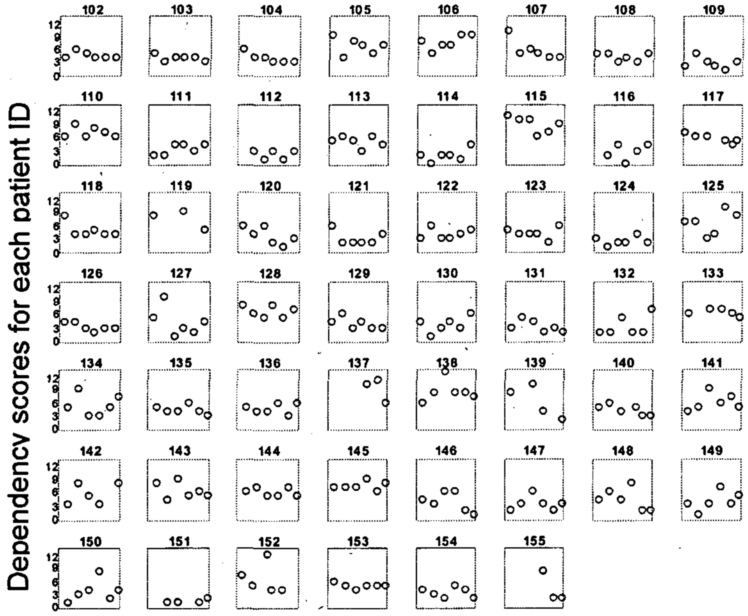
The final purpose of piloting was to test methods to classify patient heterogeneity in patients on chronic HD (at Totton) using three types of tool.

7.3.1 Patient dependency-scoring tool for outpatient HD

Patients were rated at each HD session using the patient dependency-scoring tool. In total, there were 305 dependency ratings (only 1 missing) by eight RGNs across the 54 patients, and 81% of patients had the maximum six possible ratings.

Figure 7.4 shows the profiles of each patient's dependency scores, which ranged from 0 to 13, where higher scores represent increased dependency. Whilst dependency was relatively stable for some patients, for others it varied widely across HD sessions. As predicted for a RSU, patients had relatively low scores and there were only four instances of scores above 10 (i.e. high dependency). The mean dependency score using summary statistics (i.e. from the mean for each patient) was 4.6 (SD 1.8). The next section shows the preliminary results for the tool's construct or empirical validity (i.e. ability to detect or correlate with known or expected differences or changes).

Figure 7.4 Scatter plots of dependency scores by visit (Totton)

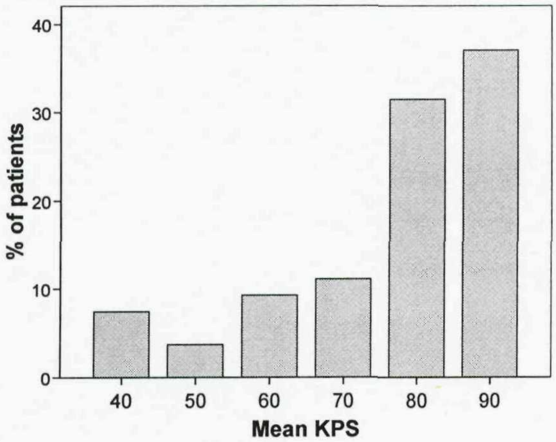


A number of concerns arose about the scoring as detailed in Appendix 18 and summarised here. Some theoretical issues remained unresolved despite discussion with senior nurses from the HD units. Prime concerns were the scaling levels assigned to dependency items, discriminatory power of items, and for some items the link with patients' needs for nursing input was unclear. Apart from using the total score, the tool presented problems for analyses due to the large number of items and multiple (unequal) levels on each item. Other concerns were the nurses' mistakes in totalling the scores and inconsistencies in ratings. Consequently, for use at Portsmouth, it was necessary to revise the tool and increase training (details given in section 8.3).

7.3.2 Karnofsky Performance Scale (KPS)

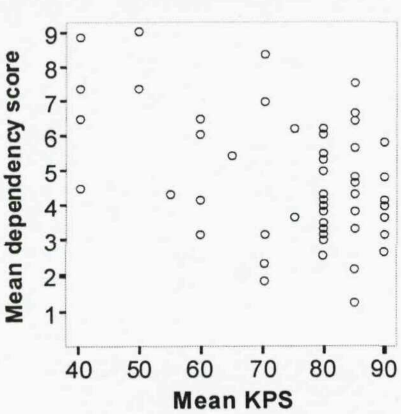
The Karnofsky Performance Scale (KPS) was rated each week and 52 patients (of the 54) had ratings on both occasions. KPS scores were consistent between weeks; for 88% of patients their scores were the same or only differed by one category despite little continuity in the raters. Figure 7.5 shows the distribution of patient summary (mean) KPS scores. A lower KPS indicates a worse score and since all patients on HD have evidence of disease they could not score 100. As predicted for a RSU, few patients had severe functional impairment (i.e. scored 40 or less). Overall, the mean KPS score was 75 (SD 14) using each patient's summary mean.

Figure 7.5 Distribution of patients' mean KPS (n=54, Totton)



The KPS was used as a preliminary test of the dependency-scoring tool's construct or empirical validity. As predicted, the graph of patients' mean dependency score and mean KPS showed a negative correlation, though weak (Spearman's rho -0.31, p<0.05, see Figure 7.6).

Figure 7.6 Graph of patients' mean dependency score and mean KPS (Totton)



7.3.3 Co-morbidity indices

To test the feasibility of co-morbidity scoring, the Modern Matron audited the data available on computer and in the HD notes for 29 patients at the Totton unit. She found scoring the Wright/Khan and Charlson indices unfeasible since many aspects were recorded inconsistently or missing. Accurate scoring needed data extraction from patients' full medical notes, ideally by a doctor. This required additional ethical and research governance approval, patient consent, and extra research resources, all of which were beyond the scope of the study.

The finding was disappointing, but unsurprising. Ansell et al (2007) noted that in 2006, Portsmouth only submitted co-morbidity data on 34% of new patients to the UK Renal Registry. The authors were optimistic that similar poor reporting by units would improve. From 2006, Registry reports have identified each centre's co-morbidity adjusted survival rates, and the Registry is exploring linking into Hospital Episode Statistics (HES data) to access co-morbidity data from inpatient care. Similar problems with data quality in the US meant that modellers excluded co-morbidity variables in producing the case mix adjustments for HD (the Kidney Epidemiology and Cost Center 2004).

Piloting showed that co-morbidity scoring to characterise the patients on HD was unfeasible for the main data collection. From a nursing viewpoint, the indices excluded many aspects (e.g. mobility) that were indicators of nursing care. Therefore, without further research, it is unclear whether co-morbidity scores are useful in assessing nursing time and hence costs for routine HD.

7.4 Conclusions from piloting

Barcode scanning was successful, especially given its reliance on nurses who faced many challenges due to service reorganisation and staffing constraints. The nurses collected a substantial amount of data and their acceptance was high. On the negative side, barcode scanning required considerable effort in data processing, confirming other SUHT researchers' experiences in previous studies (described in section 3.4.2.1). Indeed, despite theoretical improvements in the technology, barcode scanning took a lot more time to support than expected.

Piloting found patient-level work sampling unfeasible because data could not be linked accurately to patients. Work sampling was also difficult for a single researcher whilst supporting the nurses' barcode scanning. Although one solution was to recruit local nurses for work sampling, this was not possible in the current study. In addition, work sampling data from piloting was unsuitable to validate the barcode scanning data. Work sampling using barcode scanners appeared an efficient way to capture data and automate crosschecking compared with simple paper-based recording. Yet problems with the scanners' time keeping meant data matching was impractical. As a result, data collection at Portsmouth required a different method to validate the barcode data.

Piloting the various patient dependency and co-morbidity tools at Totton was useful. The patient dependency-scoring tool showed promise in terms of face validity, high completion rates and preliminary evidence of construct or empirical validity through correlation with the KPS. Piloting identified refinements to improve and clarify the tool's coverage, and increase the consistency of scoring at Portsmouth. The KPS had a nearly 100% completion rate and no identified scoring problems. The audit of available data showed that accurate co-morbidity scoring using the Wright/Khan and Charlson indices was unfeasible. It remains unclear whether co-morbidity scores are useful in assessing nursing time and hence costs for routine HD.

Overall, piloting was worthwhile. In summary, it established barcode scanning but ruled out work sampling as a feasible method to collect the nursing time per patient. It also ruled out co-morbidity scoring to classify patient heterogeneity, but found the KPS useful and identified improvements to the dependency-scoring tool. The next chapter presents the revised methods and analyses used at Portsmouth.

Chapter 8 Methods for main data collection

Data collection methods for the main study were similar to those used in piloting, subject to the changes already discussed. Consequently, this chapter describes only the relevant additional information about the methods. This includes the setting, validity and reliability checks for barcode scanning and the dependency-scoring tool, the dataset structure, statistical analyses, and assessment of economic consequences.

8.1 The Portsmouth MRU

The main data collection was at the main renal unit (MRU) in Portsmouth, a 25-bay unit providing outpatient chronic HD to 145 patients, with on-site medical support. The unit opened 6:30-24:00 Sunday to Friday and 6:30-14:00 on Saturdays. It had 40 nurses, a mixture of RGNs and HCAs (including dialysis assistants).

The Portsmouth unit faced many similar challenges to the Totton unit in staffing and financial pressures, plus additional pressures due to a major reorganisation. This involved closure of an unfunded ward (G8) where outlier 'sick' patients were dialysed, relocation of these patients to the MRU, and transfer of MRU patients to a newly opened RSU. Other aspects planned but not realised were an increase in nursing levels and introduction of new HD machines. This upheaval delayed data collection at Portsmouth by more than 6-months.

8.2 Barcode scanning

The methods used for barcode scanning were the same as those in the pilot phase (see sections 6.1.1 to 6.1.1.3). Additional validity checks were undertaken beyond the simple overview of nurse recruitment and data completeness. These comprised checks of 'missing' and outlier data, nurses' informal feedback, and 'spot check' observations to assess the accuracy of nurses' recordings.

The 'spot checks' were akin to work sampling observations, but only twice a day (and at different times each day). Appendix 19 shows the form used to record the observation times (from a radio-controlled watch) alongside the identities of nurses involved in direct care or preparing dressings trolleys (an indirect care activity). All other indirect care and general activities were ignored because it was difficult for an observer to differentiate between them. Nurse recordings were checked to see if they matched observer recordings within 60-seconds (to allow for slippage in the barcode scanners' time keeping).

8.3 Classification of patient heterogeneity (KPS and dependency-scoring)

The simplest classification of patient heterogeneity was patients' designated eligibility for RSU care. The ability for this to differentiate between patients was assessed using the Karnofsky Performance Scale (KPS, described in section 6.2.2) and dependency-scoring tool (described in section 6.2.1). In addition, the dependency-scoring tool was used to assess the relationship between nursing time and dependency.

The KPS was rated twice (once in each data collection block) by whichever nurse was allocated to put the patient on to HD.

Based on the findings from piloting, the dependency-scoring tool was revised. This also involved discussion with the senior nurses from the local HD units and nurses at Portsmouth. Revisions included changes to the layout to aid arithmetic, addition of instructions (as previously there were none), and re-wording to clarify items or make them more objective. A few extra items that were considered important by nurses were also included (Appendix 13 shows the final version).

At Portsmouth, the nurse who put the patient on to HD rated the patient's dependency and so both RGNs and HCAs undertook dependency scoring (at Totton only RGNs did so). Validity of the patient dependency-scoring tool was assessed by overall data completeness. Evidence was also sought for construct or empirical validity; hypotheses tested were that dependency scores would be:

1. higher for patients ineligible for RSU care than for patients eligible for RSU care,
2. higher for patients allocated a RGN rather than HCA,
3. negatively correlated with the KPS, since a lower KPS score indicated poorer functional activity.

Reliability was difficult to assess. It was impractical to check test-retest reliability as dependency was expected to vary for individual patients (within-subject) and across patients, and so changes could be due to the patient's status or to the rater. It was also difficult to assess inter-rater reliability - the consistency of nurses' scoring the same patient. As shown in Appendix 20, some of the tool's items were relatively fixed patient attributes or easily verified. Other items were more subjective, particularly as the nurse's experience might influence the rating (e.g. difficulty needling the patient). Within the study, it was impractical for a second HD nurse to observe each nurse one-to-one for whole shifts. Instead, the plan was for the Modern Matron to independently rate a limited number of items (e.g. ease of vascular access); however, even this was not possible due to staffing constraints. As an alternative, the researcher audited the data quality by comparing entries on the dependency form with those on patients' HD charts (nursing notes). As far as possible, the audit included each nurse and each patient at least once, mostly from 10 to 12 November 2006.

8.4 Data management and statistical analyses

Data management followed the same procedures as described in section 6.1.1.2. Since analyses were more complex than those in piloting were, it is helpful to understand the dataset structure before presenting the statistical analyses undertaken.

8.4.1 Dataset structure

This section outlines the complexities of the dataset as regards patients, barcode scanning data of nursing times, and overall data considerations.

8.4.1.1 Patient attendances

The population of patients on HD was not static. It comprised a core of 'permanent' patients in addition to others who started or ended treatment during the data collection period (see Table 8.1). Patients typically attended for three HD treatments a week, although a few attended twice a week. Overall, the dataset was uneven and comprised repeated measures.

Table 8.1 Population of patients at the MRU

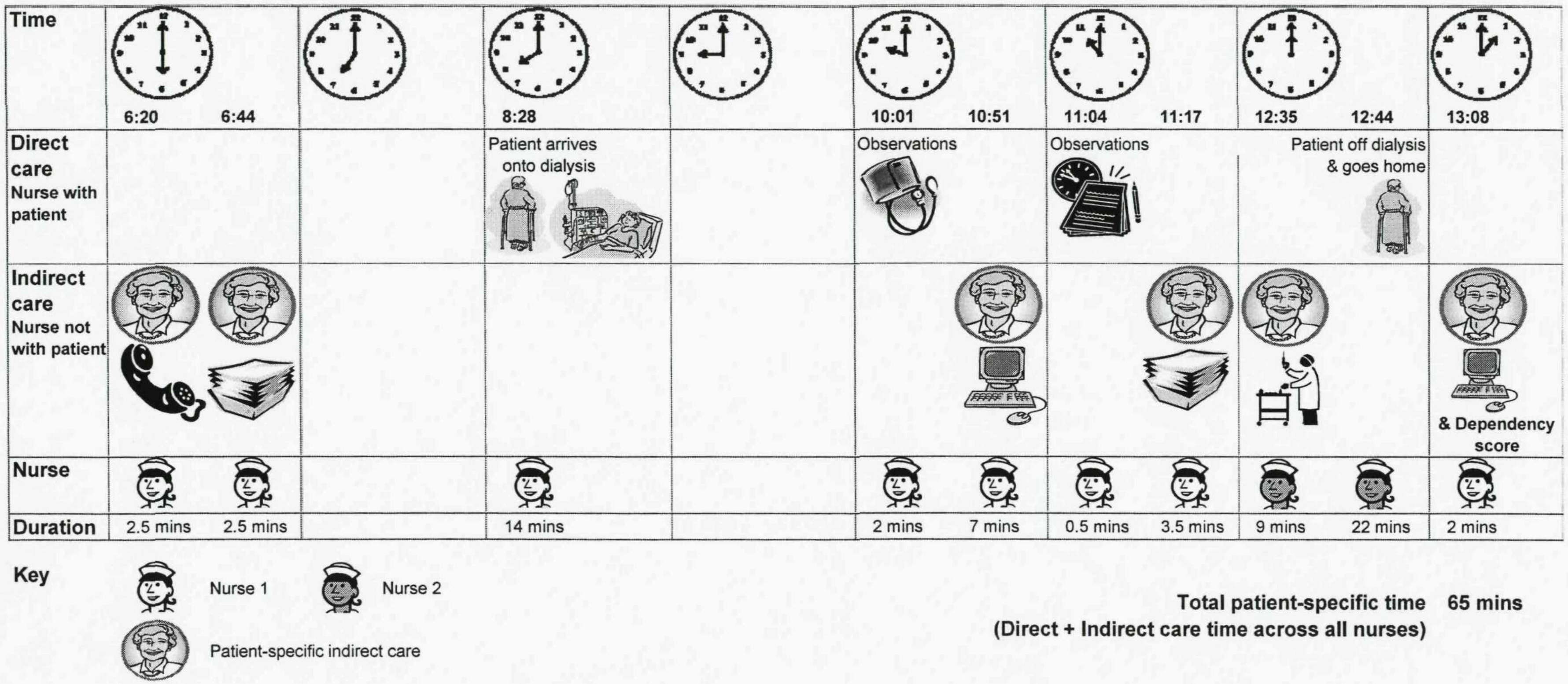
Temporary patients	'Permanent' patients	
	Entrants	Exits
• Short-term unsuitability for RSU, e.g. due to vascular access problem • New patients starting HD (before designation of RSU eligibility)	• Starting MRU HD • No longer suitable for RSU HD (long-term)	• Death • Moved to RSU • Change modality (transplant / peritoneal dialysis)

A patient could be temporarily absent due to hospitalisation, being away (e.g. on holiday) or refusal to attend.

8.4.1.2 Barcode scanning - raw nursing time data

Figure 8.1 illustrates the structure of the raw data (barcode scans) for one patient's HD session. Direct care was for nurses' face-to-face contacts with patients, such as putting the patient on to dialysis, observations (e.g. taking blood pressure) and taking the patient off dialysis. Indirect care was for a specific patient, but not face-to-face (e.g. telephone calls, paperwork or computer work and drawing up drugs for the particular patient). The figure shows a face associated with all these tasks to differentiate them from similar general activities that did not relate to a specific patient. The figure also shows that two different nurses looked after the patient; one put the patient on to dialysis and performed the routine observations, whilst the second nurse took the patient off dialysis. In addition, the nurse who put the patient on to dialysis rated the patient's dependency and entered data on computer, which explains the indirect care time after the patient had gone home. The times shown are the start of the nurse's barcode scanning for that task, with the duration shown in the last row.

Figure 8.1 Example of barcode scanning time recorded for one patient



8.4.1.3 Barcode scanning - aggregated nursing time data

To analyse the data, a new dataset was generated by aggregation of individual barcode scans as illustrated in Table 8.2. This gave the overall times across all nurses per HD session for each patient's direct care, indirect care and patient-specific time (i.e. direct and indirect care combined, the main outcome). General activity time was irrelevant for patients' time analyses, but was needed to calculate the proportions of nursing time by activity to derive the unit cost of nursing time (see section 8.4.3).

Table 8.2 Example of a patient's nursing time and dependency data aggregated by HD session

Patient	Date	Study nurse	Time recorded (mins)			Dependency
			Direct care	Indirect care	Patient-specific (Dir + Ind care)	
P1	25/9/06	N1	16.5	17.5	34	4
		N2	22	9	31	N/A
		Total	38.5	26.5	65	N/A
P1	27/9/06	N1	1	1	2	N/A
		N3	35	0	35	N/A
		N4	31	0	31	3
...		Total	67	1	68	N/A

Notes: P = Patient N = Nurse
N/A = not applicable, as dependency score linked to the nurse rater

8.4.1.4 Overall data considerations

Having illustrated the complexities of the dataset, Table 8.3 summarises important considerations for the statistical analyses.

Table 8.3 Data considerations for statistical analyses

Issue	Comment
Repeated measures	Data (time and dependency) clustered at patient-level. Ignoring data clustering can lead to standard errors (SE) that are too small (and hence p-values too small and confidence intervals too narrow).
Uneven dataset	Due to the rolling population of patients and attendance patterns.
Missing barcode scanning data and unpaired time and dependency data	*Missing completely or partially, for example due to:* • Non-study nurse providing full or part of a patient's care for a HD session. • Study nurses who stopped scanning. Dependency scoring was routine practice and not reliant on the study nurses. The nurse-patient allocation sheets (for putting on and taking off dialysis) should have facilitated handling missing data. It transpired that these sheets were statements of intent, not updated during the shift, and so did not always reflect practice.
Reliability of time measurement	*Unable to assess inter-rater reliability because patient-specific* time for each HD session comprised direct care and indirect care times aggregated across multiple nurses.
Reliability of dependency-scoring	Potential inter-rater issues discussed in section 8.3.
Stability of dependency scores	Dependency scores were expected to reflect dependency at each visit and therefore vary from session to session.
Possible confounding *in analyses of time* and/or dependency	1. Variability in number of patients per nurse on each shift. 2. Interaction between type of nurse and patient dependency. E.g. if RGNs cared for the more dependent patients and more data were missing for RGNs than HCAs. 3. Interaction between nurses' experience and either dependency ratings or nursing times (e.g. with greater experienced a nurse might needle a patient quickly and easily). Unfeasible to assess the interaction because ratings did not indicate who had delivered the care and multiple nurses could be involved.

8.4.2 Statistical analyses

The analyses were based on the patient-specific nursing time per HD session delivered. The latter is the output used for HD reimbursement. Statistical analyses followed a sequence of increasing sophistication, as described below. Descriptive statistics, validity tests and analyses using summary statistics were performed in SPSS 14 (2005). Multiple linear regressions and GEE modelling were performed in Stata 9 (2005).

1. Descriptive statistics and validity tests

Standard summary or descriptive statistics and graphs were used to describe the baseline data and perform validity checks. As before, data were presented as means rather than medians. Comparisons of means between different groups (e.g. by eligibility for RSU care) were made using the two-sample t-test and using the results for unequal variance if the test for equality of variances was significant. The Mann-Whitney U test was used to compare ordinal data (KPS) by eligibility for RSU care. Two tailed p-values are quoted. As before, tests of associations between time or dependency and other variables used Spearman's rho correlation coefficients. These were interpreted using Altman's (1991) guidance on strength of agreement, as were kappa values for the tests of agreement between observer and nurse-recorded data assessed.

2. Analyses using summary statistics

Using the statistical tests above, analyses were performed to examine differences between patient-specific time by eligibility for RSU care and by dependency using two summary measures, namely:

- each patient's mean patient-specific nursing time per HD session, and
- each patient's mean dependency per HD session.

The summary measures excluded individual patient sessions where either time or dependency was completely missing. They were not weighted to account for the number times a patient attended, as other sophisticated analyses were planned. These analyses took account of the repeated data (Matthews et al 1990). They were useful, but limited, as they could not control for other factors such as the nursing hours available or an individual patient's (within-subject) variation over time.

3. Multiple linear regression analyses

A series of ordinary least squares (OLS) regressions were performed with patient-specific time per HD session (in minutes) as the predicted variable. Two explanatory variables were of interest and were entered singly or together: i) eligibility for RSU care and ii) patient dependency score. Since the former variable was categorical and the latter continuous, there was not expected to be a problem with multi-collinearity (i.e. correlation between the two variables that leads to the estimated regression coefficients becoming unstable and having inflated standard errors).

In addition, all models included two further explanatory variables that were shift-specific control variables:

1. The number of patients per nurse on the shift to take account of constraints on nursing hours. A negative relationship was expected so that if there were few patients per nurse, more nursing hours would be available and patient-specific time would be greater, and *vice versa*.
2. The percentage study nurses on the shift. A positive relationship was expected with proportionally more time recorded as the number of study nurses increased. This variable offered a way to control for data partially missing due to non-study nurses.

In addition to the usual OLS model-based standard errors (SE), the regression analyses were repeated using alternative methods to estimate the standard errors. This is important as failures to meet OLS assumptions can lead to biased estimates of coefficients and standard errors. The following alternatives were used:

- robust standard errors that accounted for minor failures to meet OLS assumptions (e.g. normality and outlier or influential observations);
- clustered by patient ID (a multivariate technique), which allowed for observations not being independent of each other (contrary to OLS). However, this method could not take account of potential within-subject correlation over time (i.e. a patient's observations that were closer together in time were more likely to be similar than those further apart).

Additional models examined included the following:

- Inclusion of an interaction term between patient dependency and eligibility for RSU care.
- Models that dropped potentially influential values (identified from diagnostics), which might substantially change the estimated coefficients. Such values included outlier observations with large residuals, those with a high leverage (i.e. an explanatory variable's observation with a large deviation from its mean), or influence.
- Robust regression that automatically and iteratively re-weighted least squares to give lower weighting to or exclude influential data points.
- Using data transformed to produce a more normal distribution.

3. General estimating equations (GEE) modelling

GEE modelling is a multivariate technique that is an extension of general linear models to analyse longitudinal or repeated data (Liang and Zeger 1986). It is a population-averaged model. The coefficients are the average effects for the population rather than the estimates for a particular individual within that population (such as those from a random effects multi-level model). Therefore, GEE can control for patient-level clustering that is of no specific interest (Kirkwood and Sterne 2003). The technique employs various possible correlation structures for within-subject correlations; however, GEE perform well even if the working correlation structure is incorrectly specified (Horton and Lipsitz 1999, Zorn 2001).

It was likely that a patient's observations closer in time would be more similar than observations that were further apart. Therefore, the GEE analyses specified the within-subject correlation structure as auto regressive (with lag one), an exponential correlation structure. The time variable was the patient's consecutively numbered visits for which both time and dependency data were available. The use of visit number was necessary because whilst the autoregressive (exponential) correlation structure allowed the data to be unbalanced (i.e. it did not require each patient to have the same number of observations), it did not allow unequal spacing or gaps. This would have occurred

1. if the date were used, as visits were unevenly spaced throughout the week,
2. on occasions when a patient missed his / her usual HD session, or
3. when time (or dependency) data were unavailable, chiefly due to care from non-study nurses.

In addition, the GEE analyses were re-run using an alternative, exchangeable (within-subject), correlation structure that has the same correlation between any two variables and does not place any restrictions on the spacing of the observations. The models used robust standard errors, as recommended by Dupont (2002), as these produce unbiased coefficients even if the correlation structure is specified inaccurately.

8.4.3 Economic consequences

Nursing time was valued to estimate the following:

- unit costs of nursing time: per hour paid, per hour worked, and per hour of patient-specific time
- the average nursing cost per HD session
- the extra cost per HD session of nursing a patient ineligible for RSU care
- the extra cost per year of nursing a patient ineligible for RSU care.

Appendix 21 details the methods and assumptions (the cost year was 2006). In summary, resource use (nursing time) from the study was valued using the additional sources of information shown in Table 8.4. Costs included payment for basic hours (both unproductive and worked, as appropriate), and additional payments for unsocial hours. (Other salary enhancements were not relevant in the current study: enhancements for specialities (e.g. psychiatric or geriatric nursing) and locality payments (e.g. London or other location allowance to attract staff.) All relevant salary oncosts (National Insurance and superannuation) were added.

Table 8.4 Additional sources of information to estimate costs

Cost aspect	Source
Salaries: National pay scales	NHS Employers (2006)
Salary oncosts (employer's contribution):	
National Insurance (NI)	HM Revenue and Customs (2005)
Superannuation	Department of Health (2007d)
Additional payments for unsocial hours	NHS Whitley Council (2004)
Expected working hours	Curtis (2007)

Section 3.4.1.1 and Table 3.7 illustrated theoretical implications of using bottom-up and top-down costing approaches. The empirical work offered an opportunity to examine these in practice and so costs were estimated using both approaches. The top-down approach estimated pay expenditure using salaries for each nurse. It differed from the unit's actual pay expenditure[43] as it used midpoint and highest salaries rather than the actual pay point for each nurse. In addition, it was assumed that all pay was at standard rates, as although a few shifts were worked as overtime these were not identified on the duty rota.

8.5 Summary

This chapter has described the setting for the main data collection, outlined where methods were similar to those in piloting, and provided additional information where the methods differed. In summary, nurses used barcode scanners to record their time use and so measure the nursing time per patient. They also used the KPS and patient dependency-scoring tool to classify patient heterogeneity. The chapter

[43] The data collection did not coincide with calendar or accounting months and so the HD unit's expenditure on salaries would not have covered exactly the same period. This alone could contribute to differences between top-down and bottom-up estimates.

outlined the validity and reliability checks for both barcode scanning and the dependency-scoring tool. It has also explained the dataset structure, data issues, and statistical analyses and valuation of economic consequences. The next chapter presents the findings.

Chapter 9 Results (Portsmouth)

The findings from Portsmouth comprise an overview of data collection, descriptive statistics and validity checks for both the barcode scanning and the dependency data, and results of the statistical analyses.

9.1 Overview of data collection and patient characteristics

This section gives an overview of the data collection (see Table 9.1), and patient and unit-level information.

Table 9.1 Barcode scanning data collection

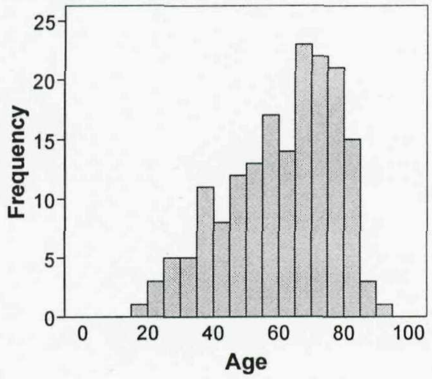
Unit operation	6:30-24:00 Sun-Fri (17.5 hours per day) and 6:30-14:00 Sat (7.5 hours)
'Pre-piloting'	All nurses, 18-23 Sept 2006 (6 days)
Data collection dates	Continuous data collection (76 nursing shifts) 25 Sept to 8 Oct 2006 and 1-14 Nov 2006 i.e. 28 days, with 23-day gap between blocks

Patient population 174 patients scheduled for HD at the MRU (nature of population outlined in Table 8.1).

Mean age 60 years (SD 16, median 64, range 18-91), 49% were aged 65 years or over (see Figure 9.1).

Most patients (64%) were eligible for RSU care[44] (see Table 9.2).

Figure 9.1 Age distribution of patients (n = 174)



[44] Designated by the Charge nurse following review of patients' eligibility status held on computer (Proton) on 14/11/06.

Table 9.2 Eligibility for RSU care (N=169)

Eligible for RSU care	N	%
No	44	26
Yes	108	64
Unknown (new patient)	17	10

HD capacity utilisation The MRU delivered 1641 of 1740 HD sessions available (94% utilisation). Of the unused sessions, 2% were for patients who were away or refused to attend on the day, 3% were for inpatients and the remainder (1%) were 'spare' sessions.

Patients underwent HD 1-13 times, although 85% of patients had six or more HD sessions.

Staffing levels The number of patients per nurse (across all grades), varied from 2.5 to 4.0 (mean 3.2, SD 0.3, median 3.1). On Saturdays (only an early shift), the number was consistently less (2.5 to 3.0 patients per nurse, mean 2.9, SD 0.3, median 3.0).

The proportion of study nurses on each shift ranged from 40-100% (mean 84%, SD 15).

Overall data completeness Partial or complete time and dependency data were available for 169 patients. Five other patients were excluded from analyses; four were inpatients on renal wards, and one scheduled but temporary patient did not attend.

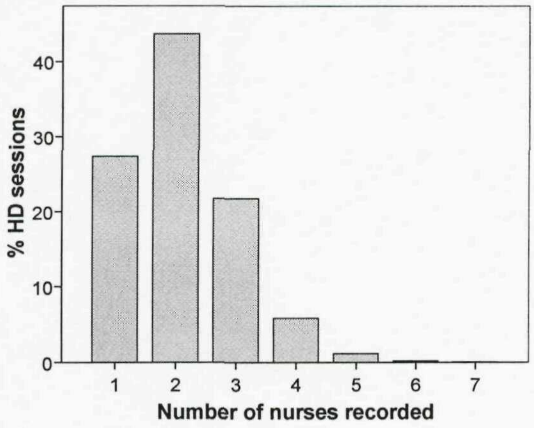
Barcode scanning data set Raw data comprised over 18,500 barcode scans, of which 55% were for direct or indirect care.

The number of scans per nurse per shift ranged from 7 to 133, corresponding to 1 to 19 scans per patient.

Within each HD session, an individual patient usually received care from two or three study nurses (range 1 to 7, see Figure 9.2), and may have received care from non-study nurses.

The main outcome of interest was patient-specific time, the sum of each patient's direct and indirect care time across all nurses (dataset structure outlined in section 8.4.1).

Figure 9.2 Number of nurses recorded per individual patient's HD session (1624 HD sessions)



9.2 Barcode scanning data - validity issues

Validity checks of barcode scanning encompassed the completeness of data, 'missing' and outlier values, nurses' feedback, and assessments of the accuracy of nurses' recordings, as described below.

9.2.1 Recruitment of nurses and completeness of data

Figure 9.3 and Table 9.3 summarise the recruitment of nurses to barcode scanning, their completion of data collection, and nurses' grade mix. The majority of nurses (87%) entered the study. The high recruitment rate was a major achievement. Data collection had been delayed until after a service reorganisation (see section 8.1), only to find it coincided with a review of the unit's hours and working practices by hospital managers and a consultancy. It took a lot of effort to reassure nurses (and some patients) of the independence of the study and that results would not identify individuals.

Figure 9.3 Recruitment of nurses

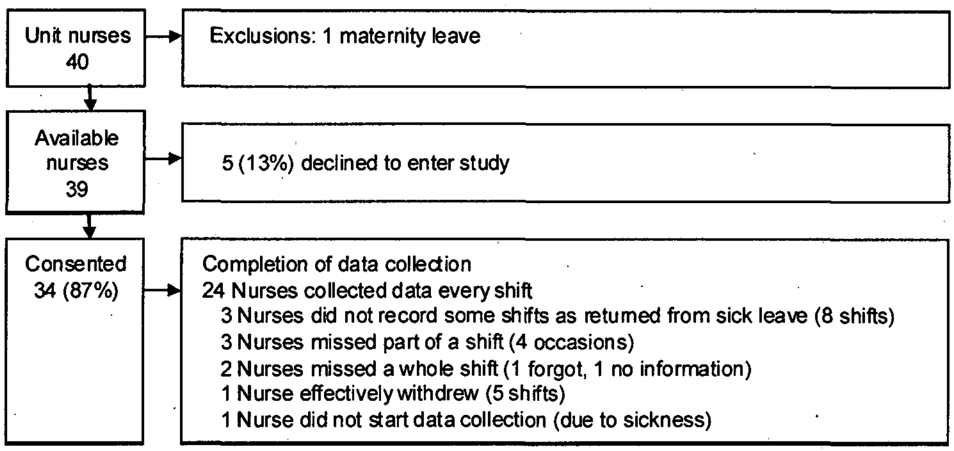


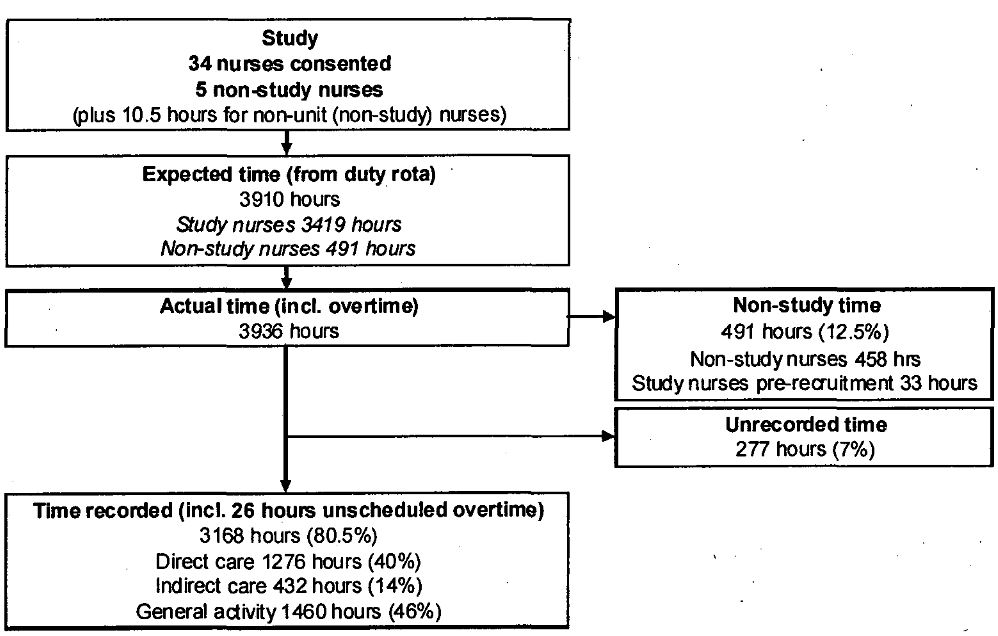
Table 9.3 Grade mix of nurses

Nurse	Grade (Agenda for Change)	Nurses		
		WTE*	Available	Study (%)
RGNs	D to G (5 to 7)	24.8	26	23 (88%)
HCAs	A, B, and HD assistants (2 to 4)	9.3	13	11 (85%)
Total		34.1	39	34 (87%)

Note: * Average based on paid hours as proportion of total paid hours divided by 150 (i.e. 37.5 hours per week x 4 weeks of study equivalent to 1 WTE).

Figure 9.4 summarises the nurses' barcode scanning; study nurses recorded 3168 hours, a high proportion (92%) of their actual hours, and 80% of overall hours.

Figure 9.4 Barcode scanning time recorded



RGNs and HCAs recorded a similar proportion of their nursing hours, as shown in Table 9.4. Overall, the small amount of 'voluntary' unrecorded time at the start or end of shifts was primarily general activity when study nurses forgot to start scanning or finished their shift early. It was unimportant for the main analyses as it represented 'overhead' time. The main reasons why nurses stopped scanning have been shown in Figure 9.3. 'Inevitable' missing data (12%) were for nurses who were ineligible or refused to enter the study, or before consent was given. As expected from piloting, RGNs spent more time than HCAs on patient-specific tasks.

Table 9.4 Time by type of nurse (HCAs, RGNs and overall)

	HCA/student		RGN		Overall	
Actual hours	1001		2935		3936	
Hours recorded	825		2343		3168	
Missing data - hours (% actual)	176 (18%)*		592 (20%)		768 (20%)	
'Voluntary' at start/end of shift		4%		4%		151 (4%)
'Voluntary' stopped scanning		3%		4%		118 (3%)
Inevitable (non-study nurses)		10%		13%		491 (12%)
Activities - hours and % of recorded						
Patient-specific	50%		55%		1708 (54%)	
Direct care		37%		41%		1276 (40%)
Indirect care		13%		14%		432 (14%)
General activity	50%		45%		1460 (46%)	

Notes: * Included 8 hours for scanner that would not download (0.2% of all actual hours).
Missing data percentages differ from sub-totals due to rounding.
General activity included all study nurses' meal breaks, as these were paid breaks.

9.2.2 Examination of possible outliers

The raw data (18,500 barcode scans) and aggregated data across 1641 HD sessions were complex (outlined in sections 8.4.1.2 and 8.4.1.3). The need to switch between the datasets was a major challenge to data cleaning. The following sections describe the two datasets and findings about outlier values (i.e. unexpected or extreme values).

9.2.2.1 Single barcode scans

Figure 9.5 shows the distributions of single direct and indirect care scans. Both were positively skewed, which is typical for resource use data. The top graphs show all values. The bottom graphs show the tails of the distributions using an arbitrary cut-off of 45 minutes for direct care and 35 minutes for indirect care as some values appeared excessive. The descriptive statistics in Table 9.5 show that individual activity scans ranged from a few seconds to nearly two hours.

Figure 9.5 Durations of single direct and indirect care scans

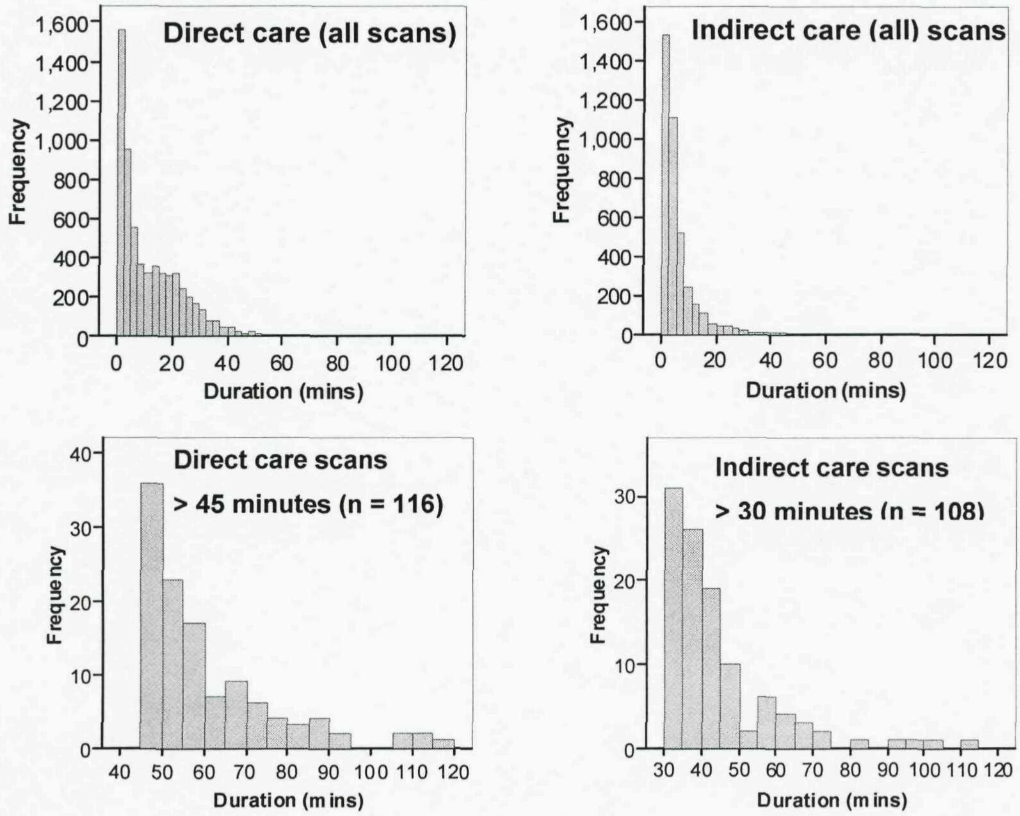


Table 9.5 Descriptive statistics for single direct and indirect care scans

		Direct care	Indirect care
Number of scans		6233	3994
Duration per scan (mins)	Mean (SD)	12.3 (12.5)	6.5 (8.7)
	Median (IQR)	8.0 (16.9)	3.8 (5.3)
	Min* - Max	0 - 116	0 - 113

Note: * Minimum of zero = more than one second

Based on the nurses' feedback and researcher's observations (see sections 9.2.3 and 9.2.4), it was decided not to remove or adjust outliers. Some scans of long duration were accurate measurements and not errors. Other scans were due to coding mistakes, particularly when nurses collected equipment for a patient (indirect care) and then forgot to scan when they started the patient's direct care. Such mistakes did not affect overall patient-specific time, but suggested that analyses at the level of direct and indirect care would be less accurate.

9.2.2.2 Aggregated time per HD session

This section examines outlier values per HD session, those that were high, and those that were unexpectedly low due to partially missing data. Times for direct care, indirect care, and overall patient-specific time varied considerably both within-subjects and across patients. (Scatter plots illustrate this for individual patients in Appendix 22; however, no patients had consistently very high or very low patient-specific times per HD session.) Like the raw data, Figure 9.6 shows the aggregated data included many outliers. The boxes represent the median and inter-quartile range (IQR), circles represent values 1.5 to 3 times the IQR and asterisks represent values more than 3 times the IQR.

Figure 9.6 Box plots of direct, indirect and patient-specific care time per HD session

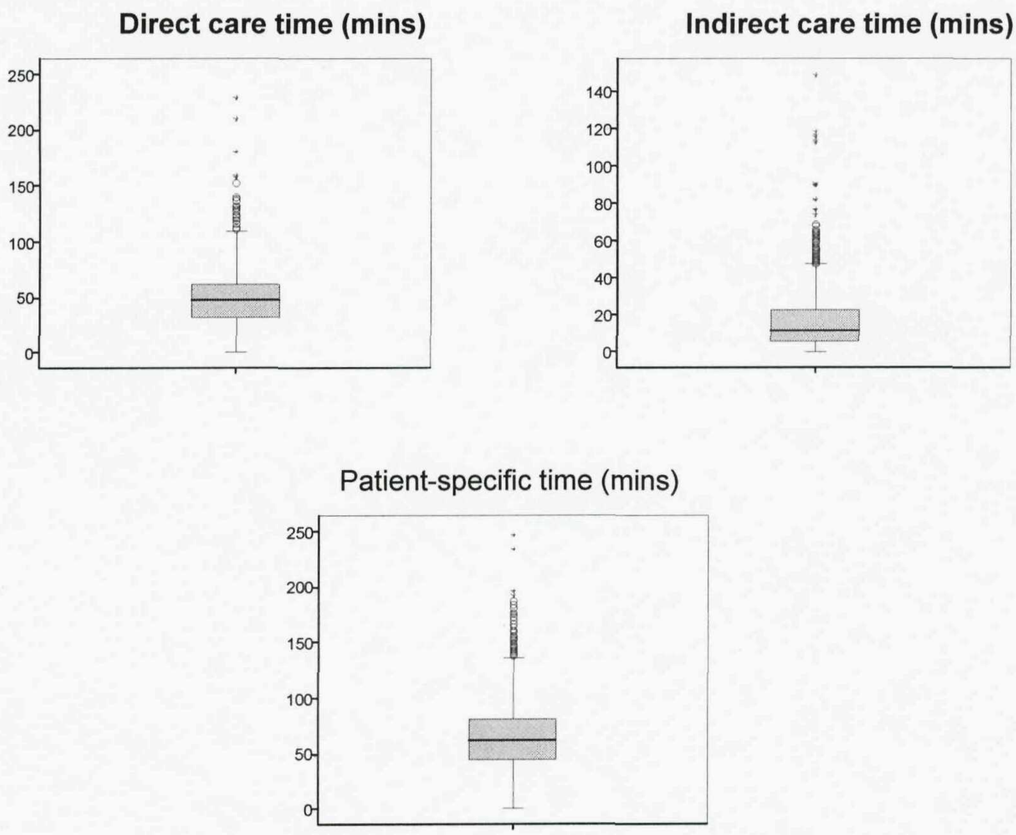


Table 9.6 shows the descriptive statistics for nursing time per HD session. Again, high values were not removed as use of arbitrary cut-offs risked discarding data with long but correct durations, although later analyses examined the effect of excluding outliers. All patients should have had some direct and indirect care time recorded; however, the table shows that for some patients, data were missing for one or other

activity. Moreover, data for some patients were incomplete shown by the times close to zero. Although patient-specific times were available for 96% of patients' HD sessions, this concealed a complicated pattern of complete or partial missing data. Hence, the mean times per HD session were slightly underestimated.

Table 9.6 Descriptive statistics - aggregated times per HD session

	No: sessions (% of 1641)	Time per HD session (mins)		
		Mean (SD)	Median (IQR)	Min*-Max
Direct care	1550 (95%)	49 (25)	47 (31)	0-229
Indirect care	1478 (90%)	17 (16)	12 (17)	0-149
Patient-specific care	1574 (96%)	65 (31)	63 (37)	0-246

Note: * Minimum of zero = more than one second

Nurses routinely entered the start and end time for each patient's HD session on computer (Proton). Therefore, to identify incomplete data, scans in a 30-minute window around these times were examined with the expectation of finding some indirect[45] and direct care activity. Table 9.7 shows the results. Interpretation is complicated because care for an individual patient was often from multiple nurses. Data could be 'missing' due to provision of care by a non-study nurse, a study nurse's accidental or conscious omission or miscoded activity. From earlier (Table 9.4), 12% of missing hours were due to non-study nurses and only 3% of missing hours were due to study nurses who voluntarily stopped scanning. Therefore, on balance, where both indirect and direct recordings were missing at the start or end of HD, it was most likely because the non-study nurse provided the care. Given that 15% of hours were missing overall[46], one would have expected about the same amount of missing data at the start or end of HD. This was broadly true for direct care (i.e. 17% and 21% of sessions), but proportionally more indirect care was missing (i.e. 25% and 34% of sessions).

[45] As a minimum, indirect care should have included preparation of a dressing trolley before both putting on to and taking off HD.
[46] This ignores the 4% 'voluntary' missing time that was primarily general activity.

Table 9.7 Missing nursing data at the start or end of patients' HD sessions

Type of care	No activity recorded for patient	
	No: HD sessions	% (of 1641)
Putting on to HD		
Direct	351	21%
Indirect	551	34%
Both direct & indirect*	272	17%
Taking off HD		
Direct	286	17%
Indirect	410	25%
Both direct & indirect*	145	9%

Notes: * i.e. neither direct or indirect care recorded

Table 9.8 shows the nursing times per patient at the start and end of each HD session. These varied widely and as before, minimum times close to zero indicated that data capture was incomplete for some patients. Although slightly underestimated, each patient typically had about 25 minutes of direct care and 10 minutes of indirect care at both the start and the end of HD. Comparison with overall times in Table 9.6 confirmed that the start and end of HD required the majority of nursing time, with minimal time for monitoring during HD.

Table 9.8 Nursing time at the start and end of each patient's HD sessions

Type of care	Time per HD session (mins)		
	Min*-Max	Mean (SD)	Median (IQR)
Putting on to HD			
Direct	0-116	22 (12)	20 (13)
Indirect	0-113	9 (11)	5 (7)
Taking off HD			
Direct	0-120	26 (14)	24 (14)
Indirect	0-107	10 (11)	6 (9)

Note: * Minimum of zero = more than one second

9.2.3 Nurses' feedback about their individual data

The researcher sought informal feedback from the nurses about their perceived accuracy of recording time data. Nurses who seemed to have any overly long indirect care times were asked to estimate typical times for a patient's indirect care at the start and end of HD. These estimates (shown in Table 9.9) were comparable with recorded times (in Table 9.8) for indirect care at the start of HD. Nurses' estimates of indirect care at the end of HD encompassed both the median and mean times recorded, although their typical estimate was closer to the median than mean. This provided some reassurance that despite coding mistakes and the presence of outliers, the mean times recorded were valid.

Table 9.9 Estimates of indirect care times per patient (from 10 nurses)

	Indirect care time (mins)	
	Typical estimate	**Min - Max**
Putting on to HD	10	6 - 20
Taking off HD	5	2 - 10
Total	15	8 - 30

Individually, nurses were given graphs of their data: direct and indirect care by *patient and total activity (direct care, indirect care and general activity)* by shift, with 'excessive' times flagged. Feedback involved 20 of the 34 nurses and covered 101 nursing shifts. Table 9.10 shows the findings. From this, the commonest mistake seemed to be inflated indirect care times. Most nurses thought they must have collected equipment for a patient (indirect care) and then forgotten to scan when they started the patient's direct care.

Table 9.10 Possible scanning mistakes identified from nurses' feedback

Possible scanning mistake	No: patient's sessions
'Excessive' overall patient care duration - probably forgot to scan next activity	7
'Excessive' duration of direct care - reason unknown	3
'Excessive' duration of indirect care	
Probably forgot to scan between indirect and direct care	33
Confirmed 'accurate' (approx. 50 mins & 75 mins)	2
Reason unknown	2
Missing direct care for patient	1
Scanned wrong patient	2
Total	50

The original plan was to get nurses to check data for all their shifts and to use the information to clean the data. In practice, this was not possible because feedback sessions were limited. Technical problems caused delays in manipulating the barcode scanning data into a suitable format. Often it was hard to find convenient times to meet the nurses and so feedback was some days later when nurses found it difficult to recall events and their estimates were less reliable. Feedback was not comprehensive; it did not cover all study nurses or shifts. Moreover, the unit's working practices were more complicated than originally expected as a single nurse rarely provided the sole care for a patient's HD session[47]. The nurse-patient allocation sheet listed the nurses to put a patient on to HD and take a patient off HD, but was a statement of intent rather than actual care. Hence, it was of little help in resolving data discrepancies. Some nurses may have been more accurate than others were; however, since no nurse consistently or solely cared for the same patient and the dataset was large, it was assumed that mistakes would not bias the overall results.

9.2.4 'Spot check' observations

The researcher did 196 'spot checks' to assess the accuracy of nurses' barcode scanning. The observations across 21 of the 28 days of data collection involved 31 of the 34 study nurses, each observed 1-14 times (mean 6.3). Table 9.11 shows the results. Even through 151 (77%) records matched, the kappa statistic (0.40) indicated only a fair level of agreement. For 12 checks (6%), the observer recorded direct care but nurses recorded indirect care. In the previous section, this was the commonest mistake identified from nurses' feedback. Nurses only twice (1%) scanned direct care when the observer recorded indirect care. The checks assumed nurses scanned the correct patients and so mistakes between direct and indirect care did not affect the analyses of patient-specific time. Nurses recorded general activity in 31 checks (16%) for which the observer recorded direct care (13%) or indirect care (3%). Discrepancies could arise for various reasons listed in section 6.1.1.3. On the other hand, the checks only verified recordings within 60-seconds before or after the observed time - nurses may have remembered to scan later during an activity. Indeed, 73% of nurses at the pilot sites acknowledged they sometimes scanned during the activity.

[47] During piloting, the RSU nurses pointed out differences between their 'batch-orientated' approach and the MRU, where nurses were allocated individual patients. The researcher (incorrectly) assumed this meant a single nurse undertook the patient's care for the whole HD session.

Table 9.11 Agreement between nurse-recorded and observer-recorded activity

Nurse recorded	Observer recorded			
	Direct care	Indirect care	General activity	Total
Direct care	**130**	2	n/a	132
Indirect care	12	**21**	n/a	33
General activity	26	5	**n/a**	31
Total	168	28	n/a	196

Notes: n/a not applicable
Observer-nurse agreement on the diagonal (in bold type).

9.3 Classification of patient heterogeneity

A key purpose of the classifications of patient heterogeneity was to asses whether patients eligible for RSU care differed from those ineligible for RSU care. These results are presented, followed by findings about the validity of the patient dependency-scoring tool.

9.3.1 Heterogeneity of patients on HD (KPS and dependency)

This section reports the assessments of patient heterogeneity using the Karnofsky Performance Scale (KPS) and patient dependency-scoring tool (tools described in sections 6.2.1, 6.2.2 and 8.3).

The KPS assessment of functional activity was rated once in both data collection blocks. Overall, KPS scores ranged from 40 to 90 where lower scores indicate worse functional activity. Of the 169 patients, 126 patients had both ratings, 32 patients had one rating, and 11 patients were missed. Figure 9.7 displays the difference in scores between the first and second data collection blocks and shows that scores were fairly consistent over time. For 70% of the patients, scores were the same or only differed by one category despite different raters.

Figure 9.7 Difference in KPS scores between data collection blocks (n = 126)

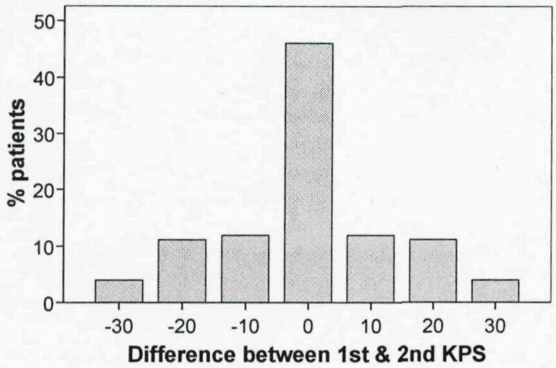


Figure 9.8 shows the distributions of patients' mean KPS scores by eligibility for RSU care and Table 9.12 shows the summary statistics. As expected, patients who were ineligible for RSU care tended to have lower scores (i.e. worse functional activity) than those eligible for RSU care did (Mann-Whitney U test p<0.001). By comparison with actual RSU patients at Totton, patients at Portsmouth who were eligible for RSU care had slightly better KPS scores. Whilst none of these patients at Portsmouth scored less than 60, six patients (11%) scored 40-50 at Totton.

Figure 9.8 Distribution of patients' mean KPS by eligibility for RSU care (n=147)

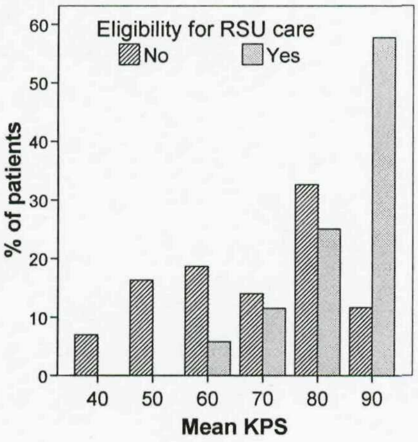


Table 9.12 Descriptive statistics for patient summary KPS scores

	Patient summary mean KPS		
	Portsmouth eligibility for RSU care		*Totton RSU patients*
	No (N=43)	**Yes (N=104)**	***(N=54)***
Mean (SD)	67 (15)	82 (9)	*75 (14)*
Min-Max	40 - 90	60 - 90	*40 - 90*

Dependency scoring was routine practice for each patient's HD session and virtually no scores were missing (nine (0.5%) missing of the 1641 HD sessions). Figure 9.9 shows the number of dependency ratings per patient, which ranged from 1 to 13 (mean 10, median 12). Most patients (85%) had at least six dependency ratings (i.e. 2-weeks data), and 63% of patients had the full 4-weeks data. As expected from piloting, dependency varied both within-subjects and across patients (see for Appendix 23 for profiles).

Figure 9.9 Number of dependency ratings per patient (1632 HD ratings)

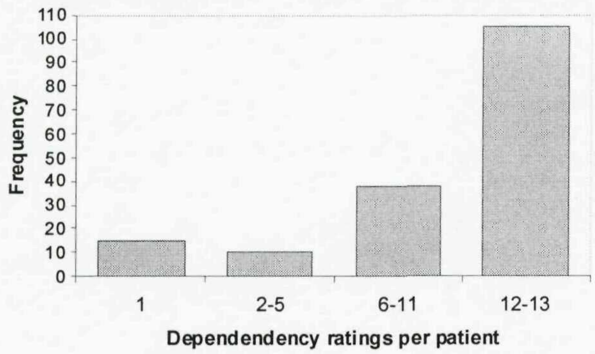


Figure 9.10 shows the distributions of dependency scores by eligibility for RSU care; both were positively skewed as few patients had high scores. Those eligible for RSU care tended to have lower scores (i.e. were less dependent) than those ineligible.

Figure 9.10 Distribution of dependency scores (all data) by eligibility for RSU care (1579 ratings)

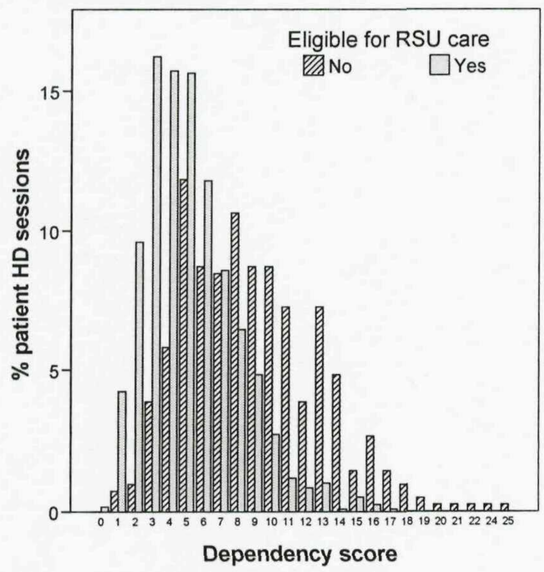


Table 9.13 shows descriptive statistics for the summary dependency scores that accounted for repeated data (i.e. each patient's mean, minimum and maximum), and Appendix 24 shows the graphs. Dependency scores for patients at Portsmouth eligible for RSU care were comparable with actual RSU patients at Totton[48]. Within the Portsmouth MRU, dependency scores had a narrower range and lower mean for patients eligible for RSU care and the mean difference was 3.9 points (95% CI 3.0-4.9 p<0.001). From these data and nurses' interpretation of scores, those eligible for RSU care were low-dependency whilst those ineligible for RSU care were mid-dependency.

Table 9.13 Descriptive statistics for patient dependency scores

	Eligible for RSU care	
	No (N = 43)	**Yes (N = 108)**
Min-Max	1-25	0-17
Summary mean (SD)	9.1 (2.8)	5.1 (1.7)
Interpretation*	Mid-dependency	Low-dependency
Nurse to patient ratio	1 to 2.5	1 to 3

Notes: * Based on the nurses' interpretation of scores from Table 6.2

The study did not include a few routine patients who dialysed outside the MRU's HD unit (i.e. as 'outliers') on the renal day ward to prevent blocking HD bays. These were mostly patients who needed stretcher transport by ambulance and were likely to be of higher dependency. Minority outliers are common; for example, the RSU study found about half the MRUs had outliers, although of unknown dependency status (Roderick et al 2005).

9.3.2 Validity of the patient dependency-scoring tool

It was important to assess the validity of the dependency-scoring tool. As noted above, virtually no scores were missing. Each nurse rated between 2 and 72 patients (mean 37, median 33), although the nurse's identity was missing for 13% of ratings. This section presents findings about the tool's construct or empirical validity and the quality of data entered on dependency forms.

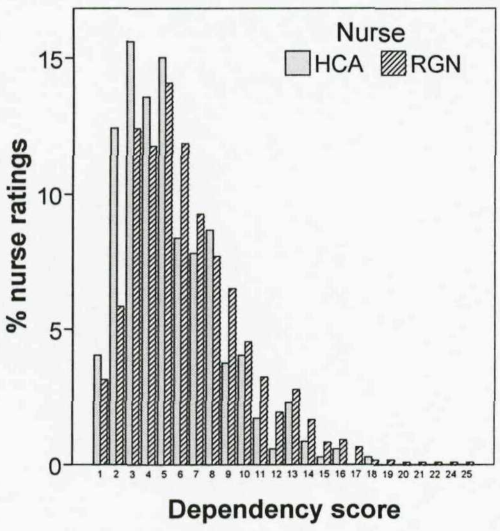
[48] Summary mean dependency scores: 4.6 (SD 1.8) for Totton patients and 5.1 (SD 1.7) for Portsmouth patients eligible for RSU care (slightly greater scores due to extra items in the later versions of the tool).

9.3.2.1 Construct or empirical validity of dependency-scoring tool

Three hypotheses were tested to assess the dependency-scoring tool's construct or empirical validity. The previous section showed evidence for the first hypothesis, that patients ineligible for RSU care would have larger dependency scores than those eligible.

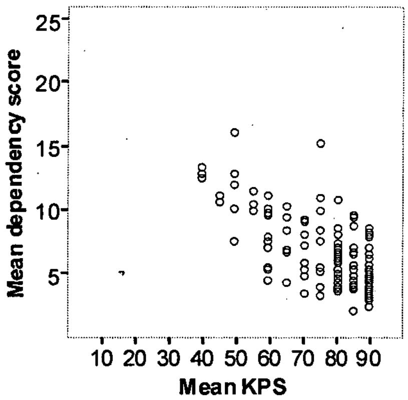
The second hypothesis was that the more dependent patients would be allocated a RGN to put them on to HD and rate their dependency score. The distributions of scores by RGN or HCA appeared to support this (see Figure 9.11) and RGNs rated a maximum score of 25 compared HCAs maximum of 18. Although the mean difference in scores between patients allocated a RGN and HCA was small (one point, 95% CI 0.6 to 1.4), it suggested the tool could detect expected differences.

Figure 9.11 Distribution of dependency scores (all data) by RGN or HCA rater (1426 ratings, 38 nurses including non-study nurses)



The third hypothesis was that there would be a negative association between patients' dependency and KPS. This is demonstrated in Figure 9.12 and by a moderate correlation (Spearman's rho -0.6, p<0.001). Overall, the three tests provided evidence of the tool's construct or empirical validity.

Figure 9.12 Patients' mean dependency and mean KPS (n = 158)



9.3.2.2 Audit of dependency data quality and inter-rater issues

The audit of data quality compared information recorded on the dependency forms with that recorded on the HD charts (nursing notes). It covered 163 dependency ratings for 149 patients (none more than twice), and all study nurses (34) with 1-10 ratings each. Non-study nurses were included as dependency rating was routine practice. Appendix 25 shows the full results, which are summarised here.

Items, such as total score, temperature, and drugs administered agreed between the dependency-scoring sheet and HD chart. Other items were recorded less accurately. This resulted in under-scoring for blood pressure and blood glucose, and over-scoring of weight deviation. In addition, ratings between nurses varied on relatively fixed patient characteristics - mobility, deafness and blindness.

The tool was complex and the findings showed that despite attempts to address concerns raised by piloting, the tool needed further developmental work. Anecdotal evidence suggested some reasons for the scoring problems. Although dependency scoring had been routine practice for approximately a year, the nurses did not feel the data were used. Many nurses were therefore uninterested in the scoring and showed some irritation towards doing it. The audit also found that some nurses still did not understand all the tool's items and so improvements to scoring might be achievable through further training. Local working practices also played a part. The nurse assigned to put the patient on to dialysis did the dependency rating, but it was rare for the same nurse to take the patient off dialysis. Although nurses said they consulted each other, it is unclear how well they captured each patient's whole HD session. Moreover, the large number of nurses meant that inter-rater differences were inevitable. Merely limiting the ratings to fewer nurses would compound the problem of rating unobserved activity. There do not appear to be simple solutions to these problems. Radical options include revising the tool to eliminate items that are more subjective or assigning one nurse for each patient's entire HD session.

9.4 Statistical analyses of nursing time

Statistical analyses addressed the following research questions:

- Is the nursing time per patient i) statistically or ii) economically different between patients by eligibility for RSU care?
- Is the nursing time per patient statistically different between patients of different dependency?

Section 8.4.2 outlined the analysis plan to assess the patient-specific nursing time per HD session delivered. (The outcome excluded 22 hours (1%) recorded for patients on non-dialysis days, for example when nurses dealt with missed sessions (planned or unexpected), and care for patients who were too ill to complete a HD session.)

9.4.1 Analyses using summary measures

These preliminary, simple analyses for repeated measures used the following summary measures:

- each patient's mean patient-specific nursing time per HD session, and
- each patient's mean dependency per HD session.

Data for HD sessions where time or dependency data were missing were excluded.

9.4.1.1 Nursing time and eligibility for RSU care

Figure 9.13 shows patients' mean patient-specific time per HD session by eligibility for RSU care. The boxes represent the median and inter-quartile range (IQR). A minority of patients had outlier values shown by the circles 1.5 to 3 times the IQR and asterisks more than 3 times the IQR, from the box edges.

Figure 9.13 Box plot of summary (mean) patient-specific time by eligibility for RSU care

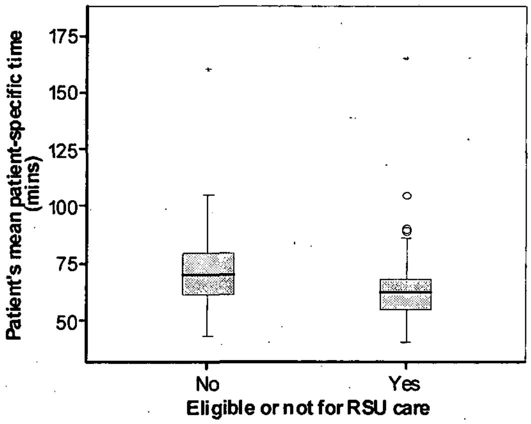


Table 9.14 shows the descriptive statistics for the summary measures. Patients ineligible for RSU care received 10 minutes additional nursing time compared with those eligible (95% CI 4-16 minutes p=0.001). The table also shows the overall average across all patients, which is equivalent to the type of estimate used in most of the economic evaluations described in section 4.5.3.2. Here, the overall average (68 minutes) was close to mean for patients eligible for RSU care (64 minutes) partly because the latter were the major patient group.

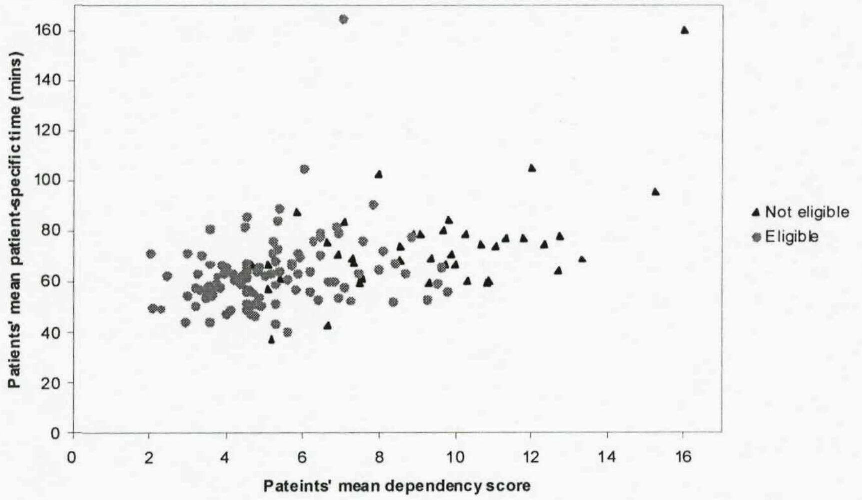
Table 9.14 Descriptive statistics for summary (mean) patient-specific time

Eligible for RSU care	Patient summary (patient-specific) time		
	N	Mean (mins)	SD
No	42	74	19
Yes	107	64	15
Overall (eligibility known and unknown)	165	68	19

9.4.1.2 Nursing time and patient dependency

Figure 9.14 illustrates a weak correlation between the summary measures (mean per session) for patient-specific time and dependency (Spearman's rho 0.44 $p<0.001$). The markers illustrate the predicted trend for patient dependency to be greater for patients ineligible for RSU care. Analyses did not control for the number repeat attendances and the two outliers (with large patient-specific care times) were patients who only attended once[49].

Figure 9.14 Scatter plot of summary measures of patient-specific nursing time and dependency by eligibility for RSU care (149 patients, 1514 HD sessions)

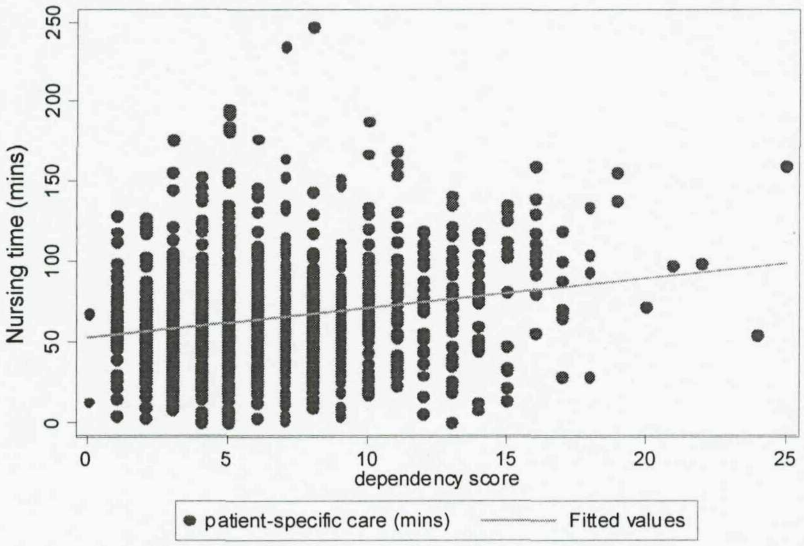


9.4.2 Multiple linear regression analyses

The multiple linear regression analyses used data for individual patient HD sessions (n = 1566). Figure 9.15 shows the scatter plot for patient-specific time and dependency score. The regression line intercepts the y-axis at 53 minutes, the average nursing time for patients of zero dependency. The slope of the regression line is 1.9, indicating that patient-specific time per patient increased approximately two minutes for each dependency point. The figure also illustrates the wide variation in patient-specific times at each dependency score.

[49] Exclusion of these two outliers reduced the standard deviations but had minimal impact on the previous results. I.e. patients who were ineligible for RSU care received 9 minutes (instead of 10 minutes) additional nursing time compared with those who were eligible (95% CI 5-13 minutes p<0.001).

Figure 9.15 Scatter plot of patient-specific nursing time and dependency per HD session (n = 1566)



The following tables present the results of a series of ordinary least squares (OLS) regressions performed with patient-specific time as the predicted variable. Two explanatory variables were of interest and were entered singly or together - eligibility for RSU care and patient dependency. Two shift-specific control variables were also included; i) the number of patients per nurse on the shift, and ii) the percentage study nurses on the shift.

Table 9.15 shows models of the separate effects on patient-specific time of eligibility for RSU care and dependency. From the regression coefficients, compared with patients eligible for RSU care, those ineligible took an average of 8 minutes extra time (similar to the results for summary measures). Alternatively, nursing time per patient increased two minutes for each dependency point. As expected, the patient-specific time decreased as the number of patients per nurse increased, although the coefficient was not statistically significant in the dependency model. Since it was an important control variable, and even though the constrained range (2.5 to 4) contributed to its lack of statistical significance, all models included the number of patients per nurse. As predicted, there was a positive relationship between the percentage of study nurses on shift and nursing time recorded, and this was statistically significant in both models. Based on R^2 (the amount of variation explained), the dependency score was a better explanatory variable than eligibility for RSU care. This was unsurprising given that dependency was measured each HD session whereas eligibility was relatively fixed.

Table 9.15 Estimated coefficients from preliminary regression models

Patient-specific time (mins)	Model with eligibility		Model with dependency	
	Coefficient	SE	Coefficient	SE
Eligible for RSU care (95% CI)	-7.67 (-11.11 to -4.23) [a]	1.75 **	n/a	
Dependency score (95% CI)	n/a		1.76 (1.34 to 2.18)	0.21 **
Patients per nurse	-5.14	2.50 *	-4.21	2.46
% study nurses on shift	0.59	0.06 **	0.57	0.06 **
Constant	35.74	8.60 **	18.91	8.68 *
Number of observations	1520		1566	
R^2	0.079		0.104	

Key / notes: * p<0.05 ** p<0.001
[a] Mean -7.67 minutes (95% CI -11.43 to -3.91) using SE clustered by patient

Neither dependency nor eligibility for RSU care accounted for more than 10% of the variation in nursing time, so 90% was due to other unknown factors or random variability. Other influences could be nurse factors, such as interactions between nursing time and grade or skill level. Some nurses might work quickly, provide higher quality care, or consciously or unconsciously choose to give more (or less) time to certain patients. In addition, nurses were told to code all face-to-face interactions with patients as direct care even if simply chatting. This avoided the nurse or observer having to try to distinguish between 'therapy' and 'passing time'.

Table 9.16 displays the results of regression models with both explanatory variables of interest (eligibility for RSU care and dependency). Although there was an association between the two variables, inclusion in the same model did not pose a multi-collinearity problem because one variable was categorical whilst the other was continuous. In the full model, patient-care time was slightly less for patients eligible for RSU care than those ineligible, but no longer significant having controlled for dependency. As before, the number of patients per nurse was insignificant, and there was a significant positive relationship between the percentage study nurses on shift and nursing time.

Table 9.16 Estimated coefficients from full regression model

Patient-specific time (mins)	Coefficient	SE	Robust SE	Clustered SE (by patient)
Eligible for RSU care	-1.72	1.96	2.06	2.09
Dependency score	1.61	0.24 **	0.27 **	0.25 **
Patients per nurse	-3.39	2.49	2.30	2.27
% study nurses on shift	0.58	0.06 **	0.06 **	0.06 **
Constant	17.44	8.94	8.37 *	7.96 *
Number of observations	1514			
R^2	0.105			
Key: * p<0.05 ** p<0.001				

Table 9.16 also shows the results of alternative methods to estimate the standard errors (SE), although these did not affect the coefficient point estimates. The first set was the usual (model-based) SE. As expected, these SE for both eligibility for RSU care and dependency were smaller than either:

- robust SE (that accounted for minor failures to meet OLS assumptions, e.g. normality and outlier/influential observations), or
- clustered SE (that accounted for repeated observations on the same patient).

In contrast, the robust and clustered standard errors were smaller than the model-based SE for the control variables - the number of patients per nurse and percentage study nurses on shift - but these were shift, not patient-level variables. Overall, the alternative methods to estimate standard errors did not change the statistical significance of the coefficients except for the constant, which was not significant using model-based SE. The R^2 for the full model was only fractionally better than the model without eligibility for RSU care (i.e. it explained 0.1% more variability in patient-specific time).

From Table 9.16, the corresponding regression equation was:

Patient-specific time (minutes) = 17.44 (constant)

- 1.72 if eligible for RSU care or zero if ineligible

+ 1.61 x dependency score

- 3.39 x patients per nurse

+ 0.58 x percentage study nurses on shift.

Using this equation, Table 9.17 shows the predicted patient-specific time for patients eligible and ineligible for RSU care based on the following:

- the mean summary dependency scores from earlier (i.e. 5.1 and 9.1)

- each nurse cared for 3.2 patients (the mean for the study)
- 100% study nurses.

Similarly, the overall (average) patient-specific time per HD session was predicted using the coefficient from a further multiple regression (clustered by patient, but only including the two shift-specific control variables). From this, the eight-minute difference between the two patient groups represented 11% of the average patient-specific time per HD session. These multiple regressions controlled for missing data to some degree, and each estimate was about five minutes greater than corresponding estimate from analyses using summary measures (in Table 9.14).

Table 9.17 Patient-specific nursing time per HD session

Eligible for RSU care	Patient-specific time per HD session (mins)
No	79
Yes	71
Overall (eligibility known and unknown)	73

Other multiple regression models examined included the following:

- inclusion of an interaction term between patient dependency and eligibility for RSU care (not statistically significant);
- robust regression (which iteratively re-weighted least squares to give lower weighting to or exclude influential data points);
- using log-transformed patient-specific time (since this was positively skewed);
- using square root time (since the distribution was closer to normal);
- using both log-transformed patient-specific time and dependency.

In these models, the significance of coefficients did not change although the values for R^2 varied slightly, but without a dramatic improvement. Additional models dropped potentially influential values identified by regression diagnostics. Whilst R^2 improved, changes to the regression coefficients and standard errors were minimal (although in some cases the number of patients per nurse became statistically significant). Since all the coefficients were small (i.e. a couple of minutes), the overall results were similar to the original full model.

9.4.3 General Estimating Equations (GEE) analyses

Whilst the regression analyses clustered by patient ID took account of the repeated data, GEE offered a further improvement because it takes account of within-subject correlations over time (explained in section 8.4.2). Table 9.18 shows the GEE analyses of the separate effects on patient-specific time of eligibility for RSU care and dependency; the coefficients and standard errors were almost identical to those from the multiple linear regression (Table 9.15).

Table 9.18 Estimated coefficients from preliminary GEE models

Patient-specific time (mins)	Model with eligibility		Model with dependency	
	Coefficient	Robust SE	Coefficient	Robust SE
Eligible for RSU care (95% CI)	-7.66 (-11.39 to -3.93)	1.90 **	n/a	
Dependency score (95% CI)	n/a		1.74 (1.32 to 2.17)	0.22 **
Patients per nurse	-4.79	2.32 *	-4.05	2.20
% study nurses on shift	0.59	0.06 **	0.56	0.06 **
Constant	34.49	7.90 **	18.34	7.81 *
Number of observations and patients	1508 observations 143 patients		1554 observations 153 patients	

Notes: * p<0.05 ** p<0.001
Correlation structure: Autoregressive (lag one) using visit number for each patient

Table 9.19 displays the results of the full GEE analyses with both explanatory variables of interest. Again the results were almost identical to those of the previous multiple linear regression model (Table 9.16). Eligibility for RSU care had no significant effect on patient-specific care time having controlled for dependency and the other two shift-specific variables.

Table 9.19 Estimated coefficients from full GEE model

Patient-specific time (mins)	Coefficient	Robust SE	95% CI
Eligible for RSU care	-1.83	2.08	-5.90 to 2.24
Dependency score	1.56	0.25 **	1.07 to 2.06
Patients per nurse	-3.25	2.27	-7.70 to 1.19
% study nurses on shift	0.58	0.06 **	0.46 to 0.69
Constant	17.16	7.93 *	1.58 to 32.74
Number of observations and patients	1508 observations, 143 patients		

Notes: * p<0.05 ** p<0.001
Correlation structure: Autoregressive (lag one) using visit number for each patient

The results of other analyses produced similar results. There was no significant interaction for eligibility and dependency. The coefficients and standard errors were very similar when the GEE analyses were re-run using an exchangeable (within-subject) correlation structure (i.e. with the same correlation between any two variables). This confirmed that GEE performs well even if the working correlation structure is incorrect.

The data appeared to warrant use of GEE to take account of the repeated data and possible correlation structure. Yet results of all three types of analyses - mean summary measures, multiple linear regressions and GEE - broadly concurred. This arose because much variation in nursing times both within-subjects and across patient groups was not due to the variables measured, and within-subject correlation was small.

9.4.4 Economic consequences

This section examines the economic consequences of the study's findings about nursing time for HD. First, it considers the resource use implications of differences in nursing time between patients by eligibility for RSU care. Then it presents the associated nursing costs - per hour, per HD session, and for the difference between patients by eligibility for RSU care.

9.4.4.1 Resource use implications of differences in nursing time

The study found differences between patients' of differing dependency and their nursing time needs. From Table 6.2, nurses interpreted dependency scores in terms of three levels with separate nurse to patient ratios. By applying the proportion of nurses' patient-specific activity to each ratio, it was possible to compare them with the study's findings (see Table 9.20). The ratios predicted greater patient-specific time at each dependency level than the study data did. This suggested divergence between 'ideal' and actual staffing levels.

Table 9.20 Comparison of patient-specific time per session from nurse to patient ratios and multiple regression

Dependency level	Nurse to patient ratio*	Patient-specific time per HD session (mins)	
		From ratio**	From multiple regression***
Low	1 to 3	81	68
Mid	1 to 2.5	97	75
High	1 to 2	122	88

Notes: * From Table 6.2 ** Assuming 54% patient-specific time per shift from study

Whilst the study found a statistically significant difference in patient-specific time between patients by eligibility for RSU care, the interpretation of whether this difference matters is complex. The absolute difference in nursing time appeared relatively small (mean 8 minutes, 95% CI of 4 to 11 minutes). This could be due to the nature of the HD nursing. Patients required an element of 'fixed' time (to put them on to HD, take them off HD, and routine observations during HD), and a further element of variable time according to a variety factors including patient dependency.

Given the wide variation in nursing times within-subjects and between patients, eligibility for RSU care seemed more about risk management than nursing workload. On observation, the day-to-day operations of the RSU and MRU were more alike than originally expected. Nurses ran the HD sessions and phoned for medical help when needed; doctors did not routinely visit. Differences only arose because patients at the MRU could see a doctor on-site, whereas patients at the RSU needed a nurse to arrange transport to the MRU. Hence, whilst it would be logical to keep patients at the MRU if they were likely to need frequent or urgent medical intervention, their usual nursing needs might vary little from other patients. To put this into context, in the RSU study (Roderick et al 2005), nurses reported 544 adverse events over 6 weeks for 368 patients at RSUs. Most events (57%) were due to vascular access and cardiovascular problems. A minority of events (4%), affecting 14 different patients, resulted in transfer of the patient from the RSU to MRU, but there was no consistent reason, age group or Wright/Khan co-morbidity score.

The effect of differences in resource use depends on how the information is used. As the basis for reimbursement, costs need to include all aspects of resource use to recover expenditure. Hence all nursing time needs to be attributed to the unit cost per patient. Furthermore, even small cost differences could be important for a HD unit because of the frequency of HD (i.e. patients attend three times per week).

Yet, from an accounting viewpoint, if the extra time for patients is available within existing resources, there is no need for extra nurses, no financial outlay and so the estimated cost difference is irrelevant. Unless fully stretched, nurses should cope with patients' variable care needs by scheduling their workload. A large proportion (45-50%) of nurses' work time was general activity. Although it involved many tasks needed to keep the unit running, it also included 'slack' time (though not separately quantified). In addition, direct care probably included time when nurses simply chatted to patients. Moreover, about 4% of nurses' hours were lost at the start or end of their shifts due in part to phasing of patients HD sessions.

On the other hand, whilst the unit aimed to operate with an average of 3 patients per nurse, the study found the average was 3.2 patients per nurse. Over the study period (28 days), unproductive time was 0.7 days more per nurse than expected from the "Unit costs of health and social care" (Curtis 2007). Of this, 0.6 days per nurse was due to paid maternity leave, which Curtis does not explicitly allow for. Table 9.21 shows the proportion of hours paid at different rates. It shows that unproductive time (akin to an overhead) was approximately a quarter of paid hours. This needed to be included when deriving the average unit cost per hour of nursing time.

Table 9.21 Nurses' hours by activity

	Unproductive*	**Worked hours**		
Pay rate	**Basic**	**Basic**	**30% extra**	**60% extra**
% of hours paid	24%	44%	21%	11%
% of hours worked	n/a	58%	27%	15%

Notes: n/a not applicable
* Hours paid but for annual leave, study leave, sick leave, etc.
Data presented for nurses overall, but separate figures for RGNs and HCAs were very similar.

From an economics viewpoint, there was an opportunity cost of the extra time because nurses were not available for other tasks. However, it could be argued that as 'slack' time was available (and not being used 'productively' for patient care), simply applying salary rates overvalues the time differences. As discussed in section 3.4.1.1, salaries can be converted into different average unit costs by attributing nursing time in various ways; it is unclear which one should be used to value the opportunity cost. In relation to economic evaluations, Drummond et al (2005) noted that in practice, marginal differences are context specific. Anecdotal evidence from the study suggested that the extra nursing time for some patients did not need to impinge on the care of others. Indeed, it could be argued that nurses' time was better spent on patient care than general activity, and so the opportunity cost was close to zero. Notwithstanding these reservations, the next section examines the costs of the differences in resource use between patients.

9.4.4.2 Nursing costs

Section 8.4.3 and Appendix 21 describe the methods to estimate the following nursing costs:

	Bottom-up	Top-down
Average unit costs of nursing time:		
per hour paid	✓	✓
per hour worked	✓	✓
per hour of patient-specific time	✓	n/a
Average nursing cost per HD session	✓	✓
Extra nursing cost for a patient ineligible for RSU care - per HD session and per year	✓	n/a

The top-down costs used estimated nursing pay expenditure. Bottom-up estimates were derived by first using expected working hours from the "Unit cost of health and social care" (Curtis 2007), which is the approach typically used by health economists. Second, bottom-up estimates were derived using actual hours worked. All costs (in 2006) included payments for unsocial hours and salary oncosts - the employer's contribution for National Insurance and superannuation.

Table 9.22 illustrates the steps taken to calculate the average unit cost of nursing time per hour paid, per hour worked, and per hour of patient-specific time. These were preliminary steps necessary to derive the overall bottom-up costs per HD session or for difference between patients. For comparison, the steps to estimate the equivalent top-down average unit costs per hour are also shown. The top-down approach in the study estimated nursing pay expenditure and so differed from a 'true' top-down approach that uses actual expenditure. However, this meant that the costs from both the top-down and bottom-up approaches covered the same period[50], whereas top-down costs are usually for the financial year or a month. (Full calculations of both approaches are described in Appendix 21.)

[50] Two 14-day blocks across three calendar months.

Table 9.22 Summary - illustrating estimation of average unit costs per hour of nursing using bottom-up and top-down approaches

Steps or data required		Bottom-up	Top-down
1a) External data required			
Cost data		← Salary scales, pay rates for unsocial hours and salary oncosts (employers National Insurance and superannuation contributions) →	
Nursing hours	Paid	← 1955 hours per nurse per year[1] →	
	Worked (expected)	1560 hours (Curtis 2007)	n/a
1b) Data required from study			
Nursing hours:	Paid	(1955 hours per year, as above)	5111 hours (during study)
	Worked (actual)	1495 hours per nurse per year (extrapolated)	3900 hours (during study)
Percentage patient-specific time worked		55% for RGNs	-
Percentage hours paid at each grade		E.g. 62% for Band 5 nurses	-
Proportion of hours worked at unsocial hours rates		Average across RGNs: 21% at Saturday or twilight rate (30% extra) and 12% at Sunday rate (60% extra).	Used nurse-level data.
2) Calculation steps		1. Annual salary = E.g. £21,646 for midpoint Band 5 nurse. 2. Annual salary incl. unsocial hours and oncosts = £29,854. 3. Unit cost per hour at each grade = annual salary (incl. unsocial hours and oncosts) divided by nursing hours per year: a. per hour paid b. per hour worked (expected and actual) c. per patient-specific hour (expected and actual hours). 4. Average unit cost of nursing per hour (paid, worked or patient-specific incl. unsocial hours and oncosts) = unit costs per hour at each grade x % hours paid at grade.	1. Basic cost per hour paid = annual salary divided by hours paid per year. E.g. for Band 5 nurse = £21,646 / 1955 = £11.07 per hour paid (excl. unsocial hours and oncosts). 2. Estimated nursing pay expenditure for each nurse = (basic cost per hour paid x hours paid) + additions for unsocial hours and oncosts. 3. Total estimated nursing pay expenditure = sum of estimated expenditure for each nurse (£73,122 using midpoint salaries)[2]. 4. Average unit cost of nursing per hour = total estimated nursing pay expenditure divided by nursing hours in study (paid or worked).
3) Results			
Average nursing cost per hour:	Paid	£14.31 per hour (cf. top-down, underestimated by <0.1%)	£14.31 per hour
	Worked: Expected	£17.94 per hour (cf. top-down, underestimated by 4%)	£18.75 per hour
	Actual	£18.72 per hour (cf. top-down, underestimated by <0.2%).	
	Patient-specific hour: Expected	£33.12 per hour (cf. actual hours, underestimated by 4%)	Not applicable as would require weighting by data from bottom-up data collection.
	Actual	£34.56 per hour	

Notes: Cost year 2006, for data sources and assumptions see section 8.4.3 and Appendix 21. (cf. = compared with)

[1] I.e. 365 / 7 weeks per year x 37.5 hours per week

[2] For a 'true' top-down approach, actual expenditure data would have been obtained from the hospital finance department. Such payroll expenditure is based on individual salary points for each nurse (not midpoint or highest salaries as here) and is based on monthly pay (i.e. annual salary divided by 12) and then adjusted for the proportion of a whole time equivalent. Like the study, the basic cost per hour paid is used by payroll to calculate the enhancement for unsocial hours. The basic cost per hour paid is also used in calculation of overtime payments, which were not included in the study top-down estimate as the few overtime shifts were not identified on the duty rotas. Therefore, it was assumed all pay was for usual working hours. Payroll calculates salary oncosts similarly to the study. Additional pay expenditure, though not relevant here, includes payments for agency staff and other allowances (e.g. locality or speciality payments etc.). 'True' top-down costs would not have been directly comparable with the bottom-up estimates as the study covered 28 days (not a full calendar month) over a 3-month period.

Comparison of the bottom-up and top-down average unit costs per hour found the costs per hour paid (both £14.31) and costs per hour worked using actual hours (£18.72 and £18.75) were virtually identical. However, the bottom-up cost using expected working hours was underestimated by 4% relative to the top-down cost (£17.94 and £18.75). This was chiefly due to the effect of maternity leave. The relative difference between the bottom-up estimates using expected and actual working hours remained the same (4%) for the costs per hour of patient-specific activity (£33.12 and £34.56), but the absolute difference widened. As shown later, for short duration treatments such as HD, this had minimal effect. However, for longer-term treatments the absolute difference might be large.

Before presenting the nursing costs per HD session, Table 9.23 examines the overall nursing time per HD session. It shows the top-down average time partitioned into nurses' working and unproductive time. To some degree, the mean patient-specific time per session took account of missing data through the shift-level control variables. The revised figure was an estimate of the nursing time per session with complete data (using unit-level top-down and nurse-level bottom-up information). It suggested the mean patient-specific time was underestimated by 4 minutes (5%) per session, but it is unclear whether the missing data affected the time difference between patients.

Table 9.23 Average nursing time per HD session

Nursing time	Minutes per session	Source or assumption
Worked: 58% basic, 27% Saturday or twilight rate, 15% Sunday rate	143	Top-down time: Total nursing hours from duty rota (across all nurses) divided 1641 (HD sessions delivered).
Unproductive (i.e. annual leave, study leave, sickness, etc.)	44	
Mean patient-specific time	73	From multiple regression of nurse-recorded times (Table 9.17).
Revised mean patient-specific time	77	Percentage of patient-specific time (54% from Table 9.4) multiplied by top-down time (143 minutes from above) to estimate effect of missing data.

Table 9.24 illustrates the steps taken to calculate the average nursing cost per HD session using both the bottom-up and top-down approaches. (Full calculations of both approaches are described in Appendix 21.) The bottom-up estimate, using expected working hours, deviated from the top-down cost by 10%. Switching to using actual working hours accounted for 4% of the discrepancy. The remaining

difference appeared to be due to the missing time data because when the revised time estimate was used, the costs were almost identical.

Table 9.24 Steps to estimating average cost per HD session using bottom-up and top-down approaches

Bottom-up	Top-down
Average nursing cost per HD session = Mean nursing time per patient[1] multiplied by average unit costs per patient-specific hour[2].	Total estimated nursing pay expenditure[3] divided by HD sessions delivered in study.
Average nursing cost per HD session (incl. unsocial hours and oncosts)	
Using midpoint salaries: • £40.27 (expected hours: cf. top-down, underestimated by 10%). • £42.03 (actual hours: cf. top-down underestimated by 6%) • £44.39 (revised actual hours[4]: cf. top-down underestimated by <0.5%).	cf. £73,122 / 1641 = £44.56
Using highest salaries: • £45.87 (expected hours) • £47.87 (actual hours) • £50.56 (revised actual hours)[4] Differences cf. top-down estimate as for midpoint salaries.	cf. £83,347 / 1641 = £50.79

Key / notes: cf. = compared with
Cost year 2006. Full data sources and assumptions shown in Appendix 21.
[1] From Table 9.23: Mean patient-specific nursing time per HD session (73 mins or 1.22 hrs).
[2] From Table 5 in Appendix 21.
[3] Table 2 in Appendix 21.
[4] From Table 9.23: Revised actual working hours from Table 9.23 - estimate tried to account for additional missing patient-specific time.

Having examined the overall average cost per HD session, Table 9.25 shows the effect on nursing costs of eligibility for RSU care. This was only calculated using the bottom-up approach (i.e. from patient-level data collected). The difference in time between the two patient groups was small, and so the discrepancy between the two average unit costs for patient-specific time[51] had negligible effect. The 'correct' unit cost to apply to the difference in patient-specific time appeared to be the average cost per hour of patient-specific time. However, alternative average unit costs could be applied and it is unclear which one should be used to value the opportunity cost (discussed earlier in sections 3.4.1.1 and 9.4.4.1).

[51] I.e. expected and actual hours.

Using the average cost per hour of patient-specific time (and actual working hours), the extra nursing cost for a patient ineligible for RSU care was £2.30 to £7.22 per HD session, or £340 to £1,130 per year. The figures were best-worst case estimates based on 95% confidence intervals of the time difference and using nurses' midpoint or highest salaries. This extra cost represented 5% to 14% of the average nursing cost per HD session and 1% to 5% of the overall cost of HD[52].

Table 9.25 shows the effect of using alternative average unit costs for nursing time. In particular, if the time difference between patients was valued using the average unit cost per hour paid (equivalent to the unit cost per hour used in a top-down approach) instead of patient-specific time, the cost was less than half. Importantly, the average unit cost per hour paid attributes all nursing costs equally across patients, whereas the average cost per hour of patient-specific time attributes the 'overhead' elements of nursing costs in proportion to patient-specific time.

[52] Using the reference costs or national tariff shown in Table 4.1. Alternatively, it was 1 to 4% of the national tariff multiplied by Portsmouth Hospitals NHS Trust's market forces factor (Department of Health 2006b) to adjust for variation in 'case mix', regional wages and other costs.

Table 9.25 Steps to estimating mean extra nursing cost per patient ineligible for RSU care (using bottom-up approach)

	Mean extra cost (incl. unsocial hours and oncosts)	
	Midpoint salary	**Highest salary**
Resource use: Extra nursing time per HD session per patient ineligible for RSU care	8 mins (95% CI 4 to 11 mins) or 0.133 hours (95% CI 0.067 to 0.183 hours)	
Cost calculation: Extra cost per HD session per patient ineligible for RSU care	Extra nursing time per patient x average unit cost per hour of patient-specific time (E.g. 0.133 x £33.12 = £4.42 for midpoint salary)	
Estimates using unit costs per patient-specific hour		
Using average cost per patient-specific hour (expected working hours)		
Extra per year (95% CI)	£690 (£340 to £950)	£790 (£390 to £1,080)
	5 to 14% nursing or 1 to 4% overall cost per HD session	
Using average cost per patient-specific hour (actual)		
Extra per HD session (95% CI)	£4.61 (£2.30 to £6.34)	£5.25 (£2.62 to £7.22)
Extra per year (95% CI)*	£720 (£360 to £990)	£820 (£410 to £1,130)
	*5 to 14% nursing or 1 to 5% overall cost per HD session***	
Alternative estimates		
Using average cost per hour worked (actual)		
Extra per year (95% CI)	£390 (£200 to £540)	£440 (£220 to £610)
	3 to 8% nursing or 1 to 2% overall cost per HD session	
Using average cost per hour paid (unit cost per hour 'same' as top-down approach)		
Extra per year (95% CI)	£300 (£150 to £410)	£340 (£170 to £470)
	2 to 6% nursing or 0.5 to 2% overall cost per HD session	

Key / notes: Cost year 2006, based on average nursing costs per hour from Table 5 in Appendix 21.
* Extra per year = extra per session x 3 x 52 (i.e. assuming patient attends 3 sessions per week).
** Overall cost per HD session was HD reference cost, national or local tariff from Table 4.1.

Table 9.26 summarises the key differences between the bottom-up and top-down costing in the study. These issues are discussed in detail in section 10.1.

Table 9.26 Summary of key differences between bottom-up and top-down costing in study

BOTTOM-UP	TOP-DOWN	COMMENT
RESOURCE USE: Nurse-level activity Extrapolated from study to days worked (actual) per year. Used activity averaged for RGNs and HCAs: % unsocial hours, % time spent on patient-specific activity and % hours paid at each grade. Also used expected working days per year.	Used each nurse's total hours (unproductive, worked and unsocial).	Using expected working hours underestimated the average unit cost per hour worked and the average cost per HD session by 4% compared with the top-down estimate.
RESOURCE USE: Patient-level data - nursing time Used mean nursing time per patient, which required decisions about data cleaning including how to handle outliers.	Nursing time per patient not required and so no issue about missing data.	Missing patient-level data led to average cost per HD session being underestimated by about 6% cf. top-down (impact unknown on difference between patient groups).
UNIT COSTS Used national salaries - midpoint (typical for approach) and highest. Included payments for unsocial hours and salary oncosts worked out at annual equivalents.	Used midpoint or highest national salaries. Included payments for unsocial hours and salary oncosts applied to each nurse's total hours.	'True' top-down costs use actual expenditure data (at each nurse's salary point) and include pay rates for overtime and agency staff etc. Hence, the study top-down costs would differ from actual expenditure. However, 'true' top-down costs are often the average for the financial year or monthly, whereas in the study they covered the same period as the bottom-up costs.
Unit cost per hour Started with annual salaries adjusted for % unsocial hours and oncosts, then divided by nursing hours per year (paid, and both expected or actual hours for worked and patient-specific time).	Derived basic hourly rate paid i.e. annual salary divided by hours paid per year, which automatically included paid unproductive time.	Easy to overlook unproductive time in bottom-up estimate (which would underestimate the unit cost).
Average nursing cost per hour Average nursing cost per hour = unit cost per hour weighted according % hours paid at each grade.	Total estimated nursing pay expenditure = sum of each nurse's pay (including unsocial hours and oncosts). Average cost per hour = total estimated pay expenditure divided by nursing hours in study (paid or worked).	
Nursing cost per HD session = mean patient-specific time per patient x average nursing cost per patient-specific hour.	Nursing cost per HD session = total estimated nursing pay expenditure divided by HD sessions delivered.	Differences arose between bottom-up and top-down cost - see comments above about resource use.
Extra cost per patient ineligible for RSU care = mean extra patient-specific time x average nursing cost per patient-specific hour.	Extra cost per patient ineligible for RSU care not calculated as required judgement about how to weight the overall average cost per HD session.	Valuation of extra nursing time for patient ineligible for RSU care using average nursing cost per patient-specific hour was about double that cf. using average cost per hour paid (equivalent to a top-down approach) because it attributed 'overhead' time in proportion to patient-specific time rather than equally across patients.

Notes: cf. = compared with. The full cost per HD session (reference cost or tariff) includes other cost elements: other staff (doctors, clerical staff, technicians etc.), capital (land, buildings, equipment and maintenance), drugs, consumables (e.g. dialysis lines and fluids), overheads, and patient transport.

9.5 Conclusions

Despite unease amongst staff due to impending reorganisation, it was feasible to implement barcode scanning. Study nurses recorded 92% of their hours, which represented 80% of all nursing hours. Regression and GEE analyses partly controlled for missing data by inclusion of the percentage non-study nurses on duty.

Barcode scanning generated two datasets - individual scans and aggregated data. It was not straightforward to identify and handle outlier values, missing data and scanning mistakes, both because of the need to switch between the datasets and because multiple nurses provided care for each patient.

All patients should have had some direct and indirect care at both the start and end of HD. Since 15% of relevant data were missing overall, one would expect approximately the same amount to be missing at the start or end of HD. This was broadly true for direct care, but proportionally more indirect care was missing (i.e. for 25-34% of sessions).

There was some evidence of scanning inaccuracies as agreement between the observer and nurses was only fair (kappa 0.40). These checks could not detect if a nurse remembered to scan later during the activity, which 73% of nurses at the pilot sites said they sometimes did. From nurses' feedback, the commonest mistake appeared to be when nurses collected equipment for a patient (indirect care) and then forgot to scan when they started the patient's direct care. This contributed to the low kappa value, but should not have affected analyses of patient-specific time. Nevertheless, nurses' estimates of indirect care at the start and end of HD were broadly comparable with the average times recorded, which suggested that despite coding mistakes and the presence of outliers, the mean times recorded were valid.

Patients eligible for RSU care at Portsmouth were comparable with actual RSU patients at Totton by dependency scores, but had slightly better functional activity (KPS). Within the Portsmouth MRU, patients ineligible for RSU care were the minority (26%) and there were striking differences between the two patient groups. Patients ineligible for RSU care had significantly poorer ratings on KPS and dependency than those eligible for RSU care. There was approximately 4 (95% CI 3-5) points difference in mean dependency between the two groups. By nurses' interpretation of the scores, patients eligible and ineligible for RSU care could be categorised as low and mid dependency respectively, with associated nurse to patient ratio of 1 to 3 and 1 to 2½.

Analyses examined whether the nursing time per patient varied between patients eligible and ineligible for RSU care and between patients of different dependency. Mean summary measures, multiple linear regressions and GEE results broadly concurred. The overall patient-specific time per HD session was 73 minutes (predicted from regression clustered by patient ID). On average, patients ineligible for RSU care required an extra 8 minutes of patient-specific time per HD session (95% CI 4 to 11 minutes), which represented 11% of the overall patient-specific time per HD session. In addition, the nursing time per patient increased approximately two minutes for each dependency point.

For costs in 2006 including salary oncosts, the best-worst estimates of the mean extra nursing costs per patient for those ineligible for RSU care were £2.30 to £7.22 per HD session, or £340 to £1,130 per year. These represented 5% to 14% of the nursing costs per HD session and 1% to 5% of the HD national reference costs, and indicative national and local tariff.

The extra cost was based on the difference in resource use (i.e. mean extra time nursing per patient ineligible for RSU care) multiplied by an average unit cost of nursing time. However, the bottom-up costing approach calculated three different average unit costs of nursing time - per hour paid, worked and patient-specific. From an economics viewpoint, it was unclear which of these best reflected the opportunity cost for differences in resource use between patients. The base case used the average cost per patient-specific time, which attributed the 'overhead' elements of nursing costs in proportion to patient-specific time. However, the cost of the same resource was less than half if valued using the average unit cost per hour paid. This was equivalent to the unit cost per hour used in a top-down approach, which attributed all nursing costs equally across patients.

The interpretation of the differences in nursing time and costs between patients by eligibility for RSU care is complicated. It depends largely on how the information is used (e.g. costing as a basis for reimbursement and management information at the HD unit). Moreover, it could be argued that when 'slack' time is available that is not used 'productively' for patient care; simply applying a salary rate overvalues nursing time.

The study found differences between estimates of the average cost per session using the two costing approaches. The top-down method used estimated rather than actual nursing pay expenditure divided by activity (HD sessions delivered). For the bottom-up approach, the mean patient-specific time was multiplied by the average nursing cost per patient-specific hour. When that latter was based on nurses' expected rather than actual working hours, the cost per hour and per session were underestimated by 4%. In addition, missing patient-level data caused a 6% underestimate in the cost per session using the bottom-up compared with the top-down approach. Overall, these discrepancies between the costing approaches had negligible effect for short-term care such as HD, but the absolute differences would be greater for longer-term care episodes.

The study provided the first assessment of validity of the dependency-scoring tool. Scores were rarely missing because the tool was in routine use throughout the local HD units. The tool showed evidence of construct or empirical validity as it differentiated between patients on three tests (i.e. that scores would be i) higher for patients ineligible for RSU care than for those eligible, ii) higher for patients allocated a RGN rather than HCA, and iii) negatively correlated with the KPS). Dependency was a better explanatory variable for nursing time than simple eligibility for RSU care. The research revealed a number of concerns about the consistency of scoring (both within-rater and between raters) that would need to be addressed before wider adoption of the tool beyond the HD unit could be advocated.

Chapter 10 Discussion and conclusions

This chapter begins with an in-depth discussion of the findings from the empirical work and literature reviews: costing methods, measurement of staff time, and patient heterogeneity in HD. It then discusses the strengths and limitations of the empirical research, implications for researchers and future research, and presents overall conclusions and outputs from the PhD.

10.1 Costing methods

The first objective of the thesis was to give an overview of costing and review costing approaches for economic purposes. The thesis has argued that costing in health services is complex because health care has multiple purposes and outputs. It is often difficult to define final patient-outputs and to link or attribute resource use to them, not least because there is variation between both patients and staff, aside from variability in other aspects (e.g. technology, settings, etc.). The literature review found the scope of the costing exercise (i.e. purpose, perspective and timeframe) central to the choice of appropriate methods. It also showed there were significant gaps in practical guidance and research on methods to collect resource use data, estimate unit costs and, in particular, to attribute staff time that is shared across patients (Johnston et al 1999, Adam et al 2003a).

Sources of guidance on costing for economic purposes acknowledged the importance of valid and precise estimates, feasibility, and research costs as important influences in the choice of approach. Key publications advocated bottom-up costing as the overall preferred approach (Gold et al 1996, Mogyorosy and Smith 2005). Other publications recommended bottom-up costing in specific circumstances (Brouwer et al 2001, Swindle et al 1999, Drummond et al 2005, Gruen and Howarth 2005, Luce et al 1996, Wordsworth et al 2005). A crucial finding from the literature review was the minimal evidence examining the effect of choosing one approach over the other. The only paper, coincidentally a study of HD and peritoneal dialysis (by Wordsworth et al 2005), showed differences between the approaches but these were not consistent between cost categories, centres or treatments.

In fact, the thesis has argued that classifying approaches as so-called top-down or bottom-up is not clear-cut. In theory, the two approaches are located at opposite ends of a spectrum of precision and specificity (Drummond et al 2005, Brouwer et al 2001, Luce et al 1996). In practice, it is more complex. The approaches may be used to measure resource use, to estimate unit costs, or both, and may even combine bottom-up resource use with top-down unit costs. In addition, the bottom-up method can involve data collection at either the unit-level, or the patient-level, or both. Guidance suggests that the bottom-up method generally provides more precise, accurate and reliable estimates, and is less reliant on assumptions (Brouwer et al 2001, Drummond et al 2005, Mogyorosy and Smith 2005).

Based on the advice above, the empirical work adopted a bottom-up approach. Specifically haemodialysis had significant staff inputs (a major cost driver) and shared staff between patient groups. Wordsworth et al (2005) explicitly recommended bottom-up costing for HD, and it appeared the best option because top-down estimates were not available by eligibility for RSU care.

The empirical research provided evidence challenging the view that bottom-up costing is better than top-down. The belief that bottom-up costing is less reliant on assumptions was not borne out; every stage - data collection, data cleaning, data management and analyses - required choices between equally 'correct' ways to handle the data. The top-down approach used estimated nursing pay expenditure (rather than actual expenditure) and whereas top-down costs often include other cost elements such as a share of overheads, the study's estimates were purely for nursing salaries. Compared with the top-down cost, the bottom-up method underestimated the nursing costs per session by up to 10%. Of this, 6% appeared to be accounted for by missing patient-level resource use data. The remaining 4% was accounted for by simply switching one assumption used to derive the unit costs per hour - from expected to actual working hours. This affected the costs per session, even though the underlying resource use was unchanged.

Wordsworth et al (2005) found larger differences between the two approaches for staff costs (not just nurses). Bottom-up staff costs for HD were 21% less than the top-down ones at a Scottish hospital, but 12% more at an Estonian hospital. For peritoneal dialysis, bottom-up costs were both less (59% and 78%) than top-down costs at the respective hospitals. The authors provided little insight into why the differences occurred. One reason could have been that whereas the top-down costs were derived from annual accounts and activity, the bottom-up costs were based on typical resource use for either a session or week. In contrast, the current study of top-down and bottom-up estimates referred to exactly the same period. Moreover, here the top-down method removed variability between individual nurses' pay points by using either midpoint or highest salaries, which helped clarify what was driving differences between the two approaches.

In the current study, particular complications arose in bottom-up costing due to the patient-level data. In theory, it was feasible to undertake full patient-level costing. The barcode scanning data included care times by nursing grades for each patient. Therefore, it should have been possible to apply each relevant unit cost of nursing time and then total the component costs per patient. Such patient-level costing was not undertaken for several reasons. Missing resource use data meant patients' costs would have been underestimated. The combination of patient-level and nurse-level data would have been extremely complicated[53]. It also implied that the contribution at each nursing grade was an active decision rather than chance. Given the multiple inputs into an individual patient's care, it would have been challenging to assess any substitution or interaction effects between grade / skill and time. Indeed, Carr-Hill and Jenkins-Clarke (2003) found that on average (across 30 hospitals) there was little difference in tasks preformed by different grades of staff, and suggested that this meant staff were not being used efficiently.

In reality, patient-level costing in HD would have produced perverse results. For pragmatic reasons, lower dependency patients tended to dialyse in the evening (and at weekends). Yet these nursing shifts were more expensive due to payments for unsocial hours. A bottom-up approach that ignored these would have greatly underestimated costs, as such payments applied to 42% of working hours in HD. These findings about unsocial hours payments may be atypical of other evaluations, although they could apply to initiatives to carry out extra work after office hours (e.g. for waiting list initiatives and outpatient consultations). This could occur if catering for the 'worried well' or due to the potential need for other input (medial care, diagnostic tests, NHS patient transport, etc.).

[53] Calculations of salary oncosts are complex. National Insurance (NI) rates are applicable to total pay and are tiered (with figures for annual, monthly and weekly pay, not hourly pay). This means that when NI is applied the sum of cost components ≠ total cost and that working with elements of hourly pay is problematic.

For all the reasons above, the bottom-up method used an hourly rate weighted by the proportion of hours paid at each grade and, like the top-down method made no assumptions about who delivered the care. So, whilst an advantage of self-recording had been its ability to collect information by nursing grade, in reality these data were obsolete, although the data by RGN or HCA were used[54].

Whilst the thesis focussed on resource use measurement, it has also highlighted complexities in the estimation of unit costs of nursing time. In fact, the findings raised questions about the practicalities of costing following economic principles. Mogyorosy and Smith (2005) noted divergent views about opportunity costs. These may be context specific, time consuming and expensive to calculate, and routine accountancy costs may be reasonable proxies. Alternatively, they are an approach to decision making to ensure evaluations consider all feasible options. Drummond et al (2005) advocated the adjustment of prices to reflect opportunity costs if it were necessary to avoid substantial bias and if it could be done in an objective way. Yet Adam et al (2003a) noted a lack of practical guidance on how to make such adjustments. Netten et al (1998) argued that sometimes (e.g. role substitution), salary costs alone do not reflect the true costs of making professionals available due to the costs invested in education. A more fundamental problem in the current study was how to value nursing time differences between patients using economic principles. Whilst Hughes (1991) noted that salaries are used to represent economic cost of staff time, there does not appear to be guidance about how to use such data.

In the current study, the top-down cost simply attributed all estimated nursing expenditure equally across patients. The bottom-up approach faced the dilemma of which hourly rate to use - per hour paid, hour worked or patient-specific hour. Each had different implications for how staff time was attributed to patients, as shown in Table 10.1. The cost per hour paid (also used by the top-down approach) attributed all costs equally. The hourly rates progressively increase as they load more time (and salary costs) on to a decreasing proportion of nursing time. Hence, the patient-specific hour was approximately double the cost per hour worked. If the approach used in the "Unit costs of health and social care" (Curtis 2007) were adopted, the cost per hour for face-to-face contacts would load 69% of nursing time (and costs) on to the 31% of direct care nursing time.

[54] Information on working hours (by grade) and whole time equivalents (WTE) were obtained at the unit-level from nursing duty rotas.

Table 10.1 Implications of methods to attribute nursing time to the unit costs

Unit cost (per hour)	Paid nursing hour	
	%	**% of paid hours to be attributed**
Top-down or bottom-up		
Paid	100%	None, attributes time equally across patients.
Bottom-up		
Worked	76%	24% for unproductive time
Patient-specific	41%	59% for unproductive time and general activity
Direct care (i.e. for face-to-face contacts)	31%	69% for unproductive time, general activity and indirect care. Equivalent to "patient-related" time in the "Unit costs of health and social care".

Note: Percentages of nursing time from study data.

Some elements of nursing time, in particular 'unproductive' hours and general activity, are shared costs akin to 'overheads'. The thesis argued that these might be more appropriately attributed equally across patients, although it complicates costing on an hourly basis. This problem is similar to the dilemma about whether and how to attribute other overheads, and the issue about use of average or marginal costs. Drummond et al (2005) argued that overheads should be included unless common to all the options, though there was no single way to allocate them to patient care. Although technically correct, the patient-specific hour appeared to over value the time differences between patients (i.e. increase the marginal difference). A key conclusion is that practice could be improved by greater practical guidance on how to value the opportunity costs of staff.

The bottom-up approach did give in-depth insight into working practices and time use. Patient-level data picked up nurses' unscheduled overtime and additional help that was often poorly recorded on the ward and unit (although it left the dilemma about whether to cost this time, since it is usually unpaid). It also quantified the amount of unused time when nurses' shifts ended early. Together, these provided an indication of how the unit coped with existing resources. Alternatively, top-down costing (using actual expenditure) would capture the higher rates paid for agency staff or overtime when applicable. Nevertheless, for staff costs, divergence between top-down and bottom-up estimates will arise through averaging at different points to handle nursing grades, salary points, unsocial hours and salary oncosts. For these reasons, a further conclusion is that actual expenditure based top-down estimates and bottom-up estimates cannot reconcile exactly, although in the current study the difference (10%) was relatively small.

In terms of feasibility, bottom-up resource measurement using patient-level data collection incurs the greatest research costs. Barcode scanning required considerable research effort, in addition to more than £6000 for equipment alone. Costing guidance acknowledged that a study might combine top-down methods for some cost components and bottom-up methods for others (Luce et al 1996, Brouwer et al 2001, and Mogyorosy and Smith 2005). Yet even where a resource item seems important, in-depth measurement as here would be impractical for some or all elements in many studies, especially if multi-centre. Moreover, it is not always possible to undertake one approach or the other. Wordsworth et al (2005) found that in three[55] of ten centres information was not suitable to undertake both costing approaches. Ensuring comparability across centres is challenging. Whilst it may be difficult to unravel what is included in top-down costs, the bottom-up approach often requires assumptions about how to make costs comparable across centres.

The generalisability of the findings about top-down and bottom-up costing are limited; the study focussed on nurses and only examined one setting. However, the results reinforced the importance of reporting resource use separately from costs (Drummond and Jefferson 1996). Cost differences occurred both due to missing resource use from the patient-level data collection and due to how the unit cost per hour of nursing time was valued, even when resource use was unchanged. The findings also supported the assertion by Netten (1999) that there is no single cost for a service as the value depends on the scope and circumstances of the costing exercise. Yet, whilst there is no single 'true' cost for an item, a costing exercise should minimise bias. Given potential differences between costing methods, the results lend support to the call by Swindle et al (1999) for data validation to be built into studies when the cost is a significant decision aspect.

Whether differences between costing methods matter depends on the use of the cost data. If the costs provide the basis for tariffs (such as through "Payment by Results"), they need to cover all expenditure otherwise some production costs will not be recovered. Coles (1989) proposed that costs derived using top-down methods may be adequate for strategic level decision making, but inappropriate for resource management decisions at a local (clinical) level, especially within a DRG. Indeed, Glick et al (2007) argued there was limited evidence about whether the

[55] Albania and Russia - limited information.
The Netherlands - "finance data were too complex to collect information independently for the approaches".

accuracy of tariffs affected conclusions of economic evaluations. However, it would be premature to assume this was the case based on so few comparisons and more appropriate to heed Altman and Bland (1995) - "absence of evidence is not evidence of absence".

Routine costing in the NHS is improving, which may reduce the need for formal costing studies. The Audit Commission (2008) found that the move to "Payment by Results" has raised the profile of data quality for both NHS activity and costs, although major improvements are still required. Where appropriate, NHS Trusts are being encouraged to implement patient-level information and costing systems (Department of Health 2007b). Some cost categories will be relatively easy to trace to patients (e.g. the number of tests, X-rays, drugs, consumables). However, the systems are likely to have to adopt top-down averaging approaches to attribute staff time shared across patients. Furthermore, Ellwood (1996, as cited by Mogyorosy and Smith 2005) cautioned that complex costing methods might be as flawed as simple systems, because some overheads may not behave linearly with regard to a cost driver. Nevertheless, the Audit Commission (2008) suggested further improvements to NHS costing that would provide interesting data for investigation. Three aspects were particularly relevant, i) sampling cost data from accredited providers, ii) introduction of some normative tariffs (i.e. based on costs that an efficient provider with good quality service might expect to incur, rather than average costs), and iii) separate funding for capital (and hence costs) and quality.

10.2 Measurement of staff time

The second objective of the thesis was to evaluate methods to measure staff time per patient. Chapter 3 concluded that in choosing a method to measure time, the pros and cons needed to be traded-off because there is no 'gold standard' technique. Based on the literature review, nurses' self-recording of time by barcode scanning and observer work sampling appeared the most suitable methods for testing in the HD case study. It had been easy to be critical of the lack of information on the validity, reliability and practicalities of applying different techniques in published papers. Yet, in practice, it was difficult to design robust assessments due to the nature of both the time measurement and health care itself. For example, test-retest reliability appears largely irrelevant for time studies in health care, as one can rarely replicate exactly the same conditions.

10.2.1 Barcode scanning

In the empirical study, the first research question addressed was what could be learned about the feasibility and data quality of barcode scanning to self-record nursing time per patient? Barcodes have a wide range of uses - identification, inventory, tracking, and point of sales - in the food industry, retail, manufacturing, warehousing, distribution and shipping. Increasingly barcodes are being used in health care and the Department of Health (2007a) is promoting their use to improve patient safety and efficiency.

With so many applications, it appeared that barcode scanning offered great promise to harness technology for data collection. A main reason for choosing the method had been 'content validity', in particular the ability to capture both direct and indirect nursing care at the patient-level and by nursing grade. Chapter 3 showed barcode scanning had been used successfully in a variety of research contexts such as patient classification (Martin 1990, Blount 1999, Connell et al 1996 and 1997, Eastham 2006), comparisons across settings (Holmes 1997a), and examination of staff workload (Macfarlane and Lees 1997, Taylor et al 1998). Despite unease amongst staff due to impending reorganisation, the study found the method feasible to implement.

Yet the empirical study found implementation of barcode scanning to measure time presented numerous unexpected challenges. Whilst Taylor et al (1998) had noted it was time consuming to develop activity codes, the current study used far fewer codes. Even so, barcode scanning was still more costly and more time consuming to setup and sustain than expected. The scanning technology was disappointing, although the problems had no major effect on the nurses' time data per se. The technology did not seem to have improved over 10 years, as earlier studies by Connell et al (1997) and Taylor et al (1998) also reported technical problems. In particular, the scanners' time keeping was poor; they often stopped recording seconds and could not recognise the difference between 12 mid-day and midnight. The scanners could not be considered a long-term purchase since the manufacturer's warranty was short (3-months) and technical support was virtually non-existent. The equipment and software cost over £6000 and seemed poor value in view of the scanner failures and short 'shelf-life'.

A specific issue about using barcode scanners is that unlike paper recording, the user cannot retrospectively complete or 'correct' entries. This was flagged as an advantage by Connell et al (1997) following their initial trial of self-recording using paper time sheets. Conversely, it could present a barrier to implementation because barcode scanning reveals very sensitive information about how and when individuals work. At Portsmouth, largely due to unease about impending reorganisation, five nurses refused to enter the study, though two (HCAs) were on short-term placements. In contrast, at SUHT and Totton, no nurses refused to enter the study; those who did not consent were either unavailable or returning from long-term sick leave. Concern about the implications of findings may have accounted for the reluctance to record data by some appointments staff reported by Macfarlane and Lees (1997), and under-recording by the nurses on one unit reported by Walsh et al (2003).

After using barcode scanning in piloting, responses to the acceptability questionnaire showed the majority of nurses found the technique acceptable and experienced few problems. Taylor et al (1998) found similar results for primary care staff, although GPs would only record direct contacts after the consultation finished. Connell et al (1997) was very positive about the use of barcode scanning for local management and clinical purposes in the NHS, and their work was important as it contributed to the resource measurement for the first neurosurgery HRGs. They recognised that the success was heavily influenced by the enthusiasm and internal support from senior clinical staff who were instrumental in setting up the whole series of projects co-ordinated from SUHT. The situation was very different in the current study, a stand-alone research project driven from outside the NHS.

Turning to data quality, although acceptable, evidence of miscoding suggested that combined patient-specific time was more accurate than the separate direct and indirect care times. Waller (1999) came to the same conclusion in a study of GP practice nurses. This might partly explain why, in the current study, both RGN and HCA haemodialysis nurses appeared to spend less time in face-to-face contact with patients than other hospital nurses cited in the "Unit costs of health and social care" (Curtis 2007) - see Table 10.2. Alternatively, the HD nurses actually measured their time, whereas the other data came from consultation. Then again, the nature of HD nursing may have contributed too; it involves peaks of direct care activity to put patients on to, and take them off HD machines. Otherwise, apart from occasional monitoring, nurses had periods without patient contact unless a machine or patient problem arose.

Table 10.2 Comparison of direct (face-to-face) care by hospital nurses

Nurse	Direct care % of working time	
	Curtis (2007)[a]	Study
Team manager or team leader (i.e. sister or senior staff nurse)	45%	
Day ward	55%	36 - 41% [b]
24-hour ward	50%	48% [c]
Clinical support worker	60%	31 - 37% [b]

Notes: a. from consultation with NHS Trusts
b. Totton - Portsmouth data c. SUHT (RGN and HCA combined)

Data were representative of nurses' work, however, 20% of nursing hours were unrecorded and the nursing cost per session appeared to be underestimated by about 6% due to this missing data. To some degree, the missing data were controlled for through the shift-level control variable (percentage study nurses on shift). Imputation of missing data could be investigated in future analyses if expert opinion data were collected on times to put patients on to HD, take off HD, and monitor during HD.

It is difficult to compare the findings on data validity with the other studies from the literature review. Most of the studies did not appear to perform validity checks (Martin 1990, Macfarlane and Lees 1997, Freund et al 1998, Blount 1999). Whilst Connell et al (1996 and 1997) and Holmes et al (1997a) imputed missing values, they did not report the extent of missing data. Walsh et al (2003) noted that due to under-recording at one site, data could only be used descriptively. Otherwise, only Taylor et al (1998) gave any indication of the extent of missing data. They reported that staff under-reported their time by 2 to 11% compared with the observer, who in turn under-reported available time by up to 12%. Nonetheless, it is unclear whether the older scanners' inability to record activities of less then one minute contributed to this.

Haemodialysis was a useful case study to examine methods to attribute staff time across heterogeneous patient groups, although, as mentioned in the previous section, generalisability of the findings is limited. Indeed, in many ways routine HD is atypical of other areas of health care; it is a highly technical nurse driven service. It also more closely resembles a production line than most other areas. Like the other studies from the literature review, barcode scanning was successful for a cohort of patients on a single ward and unit.

Other economic evaluations where staff time and patient heterogeneity are important include role substitution, care in different settings, and care by specialist compared with non-specialist staff or units. Furthermore, health care is often provided by a combination of professionals besides nurses, including doctors, physiotherapist, occupational therapists, radiography, and other support and diagnostic services etc. Other researchers showed barcode scanning was feasible across a range of staff besides nurses (Macfarlane and Lees 1997, Connell et al 1996 and 1997, and Taylor et al 1998). Nevertheless, the challenges to co-ordinate and validate barcode scanning mean it would be far more difficult to implement for geographically dispersed subjects - either patients or staff. In fact, the experience suggested the need for more researcher time to manage the data and for 'daily' data checks or feedback to the nurses. In particular, barcode scanning is likely to be more difficult or even unfeasible in a multi-centre study.

10.2.2 Work sampling

The second research question addressed by the empirical work concerned the feasibility of using work sampling to measure nursing time per patient. Work sampling had been selected because as an observation technique, it did not disrupt staff's work and it offered a more efficient use of observer time than time and motion study. Conversely, compared with staff self-recording, work sampling could not estimate time for indirect care and was unable to produce information by nursing grade.

The findings about patient-level work sampling were discussed in detail in Chapter 7. In summary, work sampling was unfeasible in the current study because of difficulties in accurately linking data to patients. It remains theoretically feasible at the patient-level, though under very restrictive conditions (e.g. if patients are immobile or if the observer knows the patients). In practice, it may still be very challenging to implement to measure staff input for individual patients.

Nevertheless, work sampling is still potentially useful as a means to measure staff time, but not at the patient-level. Indeed, it has many potential advantages over other methods of time measurement. By sampling rather than continuous observation, it minimises the impact of workload fluctuations. It only requires unit-level not individual consent and the training of observers is likely to be less onerous than for staff self-recording. Moreover, in multi-centre studies it might be feasible to undertake work sampling using a limited number of observers.

The activity forming the smallest proportion of overall activity drives the sample size of observations required. Like all sampling methods, estimates will only be valid if the sample's coverage is representative and does not under or over sample activities occurring at regular or specific times. However, work sampling appears ideally suited to recording activities where coding is limited to a small number of the broad key categories such as direct and 'other' care that make up staff workload. It could be used as a preliminary means to determine the proportion of direct care undertaken, or to check staff opinions of such time. Using this information in 'pre-study modelling' may reveal whether more intensive data collection (e.g. using staff self-recording) is worthwhile. For example, if direct care is only 30% rather than 60% of working hours, the scope for variability between patients is much reduced and may have little effect on the overall results. In this case, it might not be worth investing in further in-depth study. Further research would help clarify the feasibility and usefulness of work sampling as a means to collect such data for costing purposes.

10.2.3 Time measurement in general

Overall, the thesis has demonstrated that staff inputs are complicated to measure, analyse and interpret. Many patient, staff and organisational factors may interact and influence the time inputs to patient care and influence outcomes. Moreover, time measurement is based on the actual rather than an optimum staffing structure.

Figure 10.1 shows a flow diagram of issues to consider in choosing whether and how to measure staff time in a health care setting. This draws together key points from Chapter 3 and the empirical work. Trade-offs between the methods were shown in Table 3.3. However, all methods should be considered fully, revisiting earlier steps in the figure if necessary before arriving at a final decision. This is necessary to avoid simply selecting a method based on 'content validity'. In the current study, self-recording had appeared the obvious choice because it captured indirect care time at the patient-level. Yet 5% (22 hours) of the indirect care time was recorded on non-dialysis days, which due to the output used (HD sessions delivered) could not be attributed to the individual patients anyway. Indirect care was 25% of the patient-specific time or 10% of paid nursing hours. Therefore, on reflection, it is unclear whether simply measuring direct care using another technique would have offered a good enough proxy for patient-specific time.

The appropriateness of a time measurement method also depends on a trade-off between in-depth and broad coverage. Finkler et al (1993) argued that time and motion study offered depth and work sampling offered breadth. The empirical work found that barcode scanning offered in-depth study at one HD unit, but required too much research effort to incorporate into most multi-centre studies. In this context, a question is whether the data were representative of the population, as the RSU study had found great variation in grade mix within MRU-RSU pairs and across units (Roderick et al 2005). Chapter 3 suggested the information on staff time use in the "Unit costs of health and social care" (Curtis 2007) was based on weak evidence from consultations and surveys (self-report or expert opinion), but on reflection, these may have been the best available.

Figure 10.1 Issues for consideration in choosing whether and how to measure staff time in a health care setting

Purpose of quantifying staff time

Local or national use.
Tasks, activity categories (e.g. direct care, indirect care and general activity), etc.
Use of data - absolute or relative time.

↓

Activity to be quantified

Components (fixed and variable time), frequency (regular / sporadic), proportion of overall activity.
Estimated variability in duration (min, max, average and likelihood of the extreme values).
By type of staff ± grade.

↓

Options

Time and motion, work sampling, self-recording, self-reporting and expert opinion, or alternative proxy methods (e.g. dependency and workload tools).
Possible compromises for data collection.
Unit-level or patient-level data collection (if applicable).

↓

Barriers to implementation

Staff involvement required (working conditions, individual or unit consent, training).
Acceptability of method to staff ± patients.
Practicalities in study context (stability of population of interest, geographical dispersal for tracking, who to target for required data, number of centres - generalisability issues e.g. coding of activity).
Research effort and costs.

↓

Implications for data and outcome

Validity and reliability of data (and how to assess).
Burden of data management.
Likelihood of missing data ± outliers - how to identify and handle.
Statistics required.
Other data required (e.g. unit costs or information to estimate costs).

↓

Final decision - choice of method

The empirical study demonstrated that 'unproductive' time and general activity, which were shared costs akin to overheads, diminished the effect of patient-specific time. Even within patient-specific time, HD appeared to comprise ‘fixed’ elements because patients required ‘core’ time to be put on to HD, taken off HD, and for routine monitoring. Additional marginal elements included time related to nurse factors, and patient factors such as illness, dependency, and ‘popularity’. Therefore, early on in deciding whether or how to measure staff time, it is important to estimate the level of variation expected and its effect on the amount of data required (i.e. similar to a sample size calculation). The thesis reasoned that in-depth measurement of staff time for patients attending fixed appointments was not worthwhile. If data were required, it might be more helpful for staff to record or report the types of patients at the extremes (i.e. those 'quick' to treat and those who exceed the allocated times), and by approximately how much and how often this occurs.

The choice of technique and data collection at the patient-level or nurse-level, rather than simple unit-level management information, have a number of implications. The need for consent, ethics and research governance approval at each site will extend timescales and, due to the effort involved, may preclude data collection at multiple sites if part of a large study. Whereas work sampling only required unit-level consent, the other methods require individual consent. This adds to research effort and hinders data collection where temporary staff are a significant part of the workforce. It was illustrated at SUHT where study nurses recorded 86% of their hours but only 62% of overall hours due to temporary staff.

Again, at the patient or nurse-level, a further barrier to data collection is the need to find an acceptable and stable period, a major hurdle given the frequent reorganisations in the NHS. As in the current study, these factors may delay a study and require considerable effort and reassurance to gain the co-operation of staff. At worst, it may not be possible to undertake the study or, as experienced by Walsh (2003), result in data that are unsuitable for statistical analyses. As mentioned earlier, using nurses to self-record their time use is a sensitive issue. It may be unfeasible if staff acceptability is low, they refuse to enter the study or do not comply with the data collection methods.

Coding activity is an inherent limitation of time measurement and affects the comparability of data between centres. There are potential 'boundary' issues since nurse-patient interactions are not like a manufacturing process for which time measurement techniques originated. Coding decisions have important implications because i) in practice they may be less clear-cut, ii) codes chosen affect the extent to which it is possible to validate the data using another method and to generalise results to other settings, and iii) differences in working practices across settings hamper the use of comparable codes or decision rules. For example, the proportion of general activity was lower for HCAs at Totton compared with Portsmouth (37% cf. 50%) due to both differences in coding and true practice variations. Likewise, only direct care could be compared with nurses' schema in the "Unit costs of health and social care" (Curtis 2007) because "other clinical" and "non-clinical" time were not comparable with indirect care or general activity in the current study.

Research effort and costs are overarching feasibility issues, especially because measurement of staff time may be just one component of a costing exercise. In particular, patient or nurse-level data collection may generate large volumes of data at a disaggregate level that require considerable research time to clean and analyse. An important consideration for researchers is the statistical analyses and support required for the type of data. The current study found sophisticated analyses (GEE) unnecessary. More straightforward multiple linear regression clustered by patient was sufficient to account for the repeated data. This occurred because a lot of the variation in patient-specific time was not associated with the variables used. Similarly, Holmes (1997a) found that the dependency score explained less than 20% of variance in staff time. It is unclear whether this would be true for other datasets.

A number of issues were beyond the scope of the thesis. One aspect for investigation is the effect of using staff time data in absolute terms (i.e. actual time linked to unit costs - a bottom-up approach) or in relative terms (e.g. to weight available top-down cost data)[56]. Linked to this is the question about the role of patient classification systems (discussed in section 10.3) to help attribute staff time as a possible alternative to direct measurement, although clearly there are validity issues. Indeed, just because a top-down cost is unavailable does not mean one needs to undertake full bottom-up resource use measurement and costing. It may be possible to weight available top-down costs and investigate the effect through sensitivity analysis.

A further issue concerns the method to control for variation in nurse to patient ratios - both within and across units. Staff time is constrained and workload measurement is based on the premise that staff prioritise use of their available time. In the study, multiple regression and GEE analyses controlled for differences in nurse to patient ratios on different shifts, although the variable was not statistically significant in

[56] Or to measure staff time on a relative basis using self-report or expert opinion.

many analyses. Alternatively, Brand (1991) proposed converting time into standardised units. They advocated calculating the average time for a reference group of patients with the lightest care needs at each unit, weighted by the ratio of salaries for different nurses[57]. Then the average time ('unit of service') could be used to calculate multiples for other types of patient. This issue warrants further investigation.

Finally, Netten (2002) acknowledged that the "Unit costs of health and social care" needed up-to-date information about staff time use. She noted that there was a wealth of time use data routinely collected by health professionals as part of clinical practice, audits or management information systems, but not publicly available or collated nationally. With the move to patient-level costing (discussed in section 10.1) it remains to be seen whether systems will link these data.

10.3 Patient heterogeneity in chronic haemodialysis

The final objective of the thesis was to assess the impact of patient heterogeneity on nursing costs for chronic haemodialysis. From the literature review, many economic evaluations of HD did not appear to consider patient heterogeneity or poorly reported the methods, and they ignored the effects on routine HD nursing costs.

In the empirical work, the final questions addressed were whether in chronic HD the nursing time per patient was statistically or economically different between patients i) who were eligible and ineligible for care at a renal satellite unit (RSU), and ii) of different dependency. In addition, the analyses specifically examined the impact of using an overall average rather than an estimate for the relevant sub-group. The latter related to the challenge of like-for-like comparisons when there is patient heterogeneity in at least one patient group.

As a preliminary issue, the study confirmed there was evidence of heterogeneity between the two patient groups through differences in functional activity (KPS) and dependency scores by eligibility for RSU care. These supported findings by Plough et al (1984) that MRUs had a higher proportion of patients in the higher-severity groups than RSUs did, albeit in an old study from the US.

[57] I.e. weighted time in minutes = aide time + (ratio of salaries x nurse time).

10.3.1 Eligibility for RSU care

The empirical work found that despite much within-group variation, compared with patients eligible for RSU care, those ineligible required about 8 minutes (11%) extra patient-specific nursing time per session. The implications of the differences in resource use and cost depend on how the data are used. At Portsmouth, the cost difference (£2.30 to £7.22, in 2006) represented 5% to 14% of the mean nursing cost per HD session, or 1% to 5% of an overall mean cost per HD session. However, as discussed in section 9.4.4.1 and section 10.1, the monetary value may have over-valued the effect since nurses coped with the differences in workload within existing resources. Even if every patient at the MRU switched to being ineligible for RSU care, the nurses might be able to cope with the extra work, though it would have implications for skill mix (and hence costs). (HCAs were a minority, but did not provide all aspects of care that RGNs did.)

Analyses also examined the impact of using an overall average rather than the sub-group estimates. It found the MRU-average (73 minutes) was almost identical to that for patients eligible for RSU care (71 minutes), despite the 'case mix' differences shown above.

In practice, the overall effect of variation between patients depends on both the absolute difference between them and the relative proportions in each patient group. In the study, the small amount of extra nursing time required only applied to the minority (26%) of patients who were ineligible for RSU care. For the aggregate cost per HD session, the impact diminished further because nursing was just one component of the total cost. From a policy perspective, in this case, the use of average rather than the sub-group costs would not have overturned the conclusions of previous economic evaluations in HD. However, practice could be improved if researchers acknowledged whether confounding due to patient heterogeneity is likely.

The findings illustrated the problems in identifying cost drivers. As Whynes and Walker (1995) found, *a priori* prediction of high cost aspects was difficult. Nursing was a major element of overall costs in HD, and nurse opinion suggested that inter-patient differences were important. Nonetheless, the research suggested that in terms of driving cost differences, nursing was less important than expected.

The results raise the question about whether 'pre-study' modelling, as advocated by Chilcott et al (2003), would have predicted a similar outcome. On reflection, it would have been helpful to do pre-study sensitivity analyses by applying weights to average costs or guesstimating time differences and assessing the impact for different proportions of ineligible patients. Alternatively, it might have been possible to measure costs for sub-sets of patients. On the other hand, in the RSU study, nurses found it difficult to quantify the implications of patient or care characteristics (Roderick et al 2005). Without additional evidence, it is also unclear whether opinions are for average patients or unduly swayed by exceptions, and whether estimates are for actual or desirable resource use.

10.3.1.1 Implications of findings for costing outputs and reimbursement

Analyses assessed the nursing time per HD session delivered, which was easy to measure; however, as mentioned in section 10.2.3, some patient care did not accrue to completed HD sessions. Examples included arrangement of holiday HD (involving negotiation with another unit, extra paperwork and blood tests), follow-up for patients who refused to attend HD, and care for patients who attended but could not complete their session (e.g. due to vascular access problems or illness). Although this had implications for costing, in the study the unattributed time was less than 1% of recorded hours and therefore had negligible impact. Nevertheless, the study unexpectedly revealed that, despite managers' expectations, nurses were poor at recording undelivered sessions for which the hospital hoped to negotiate a reduced level of payment.

The cost per HD session delivered is also the output used for reimbursement. The MRUs co-ordinate the provision of HD through sessions at the MRU, or NHS RSUs, or subcontracting to private RSUs. The output is simple to administer, especially for cross-payments to the private units that increasingly contribute to RSU provision. Whilst alternative patient output measures, such as the cost per week or per year, could have attributed all nursing time to patients, these were not necessarily ideal for costing or reimbursement. Not all patients attend three times per week; patients change status (e.g. from RSU eligible to ineligible), they change modalities (PD, HD, or kidney transplant), and it may be difficult to ensure continuity of data collection if patients move between units.

Despite the differences found between patients by eligibility for RSU care, current HD commissioning does not disadvantage service providers on such grounds. Indeed service level agreements have used a flat payment rate (analogous to an average tariff), which enabled recovery of the unit's fixed costs and a reduced ('marginal') rate for additional activity. Since administration is by the MRUs, flat rate payment helps avoid perverse incentives across settings. Therefore, for HD, the change in HRG codes from version 3.5 to 4, which are independent of setting, was logical even though the associated 'Payment by Results' (PbR) tariff is only indicative.

It is unclear whether HD will be included in the PbR mandatory tariff in the future. If it is implemented, reimbursement of over-activity at the standard tariff may lead to inefficient overpayment for extra activity. On the other hand, the tariff is set nationally and despite direct payments from the Department of Health to take account of the local 'market forces factor', service providers may find their costs exceed the reimbursement rate. In contrast to other hospital services, PbR is unlikely to produce much incentive for HD units to try to attract patients. There is no real competition between providers, they lack spare capacity, and there is minimal scope for patient-choice between units.

Conversely, in the US there is competition between units in the large cities to attract patients (Hirth 2007), and the old 'composite-rate' (flat rate[58]) produced disincentives to care for costly patients. Medicare has changed to a new case mix adjusted payment (Wheeler et al 2006), but the impact has yet to be evaluated. It remains to be seen whether a case mix adjusted rate will be necessary in England and Wales. The situation would be very different if each unit was directly reimbursed and service provision ceased to be co-ordinated by the MRUs. Then PbR might lead to further expansion of private units, which though positive in some respects could also lead to unused capacity.

10.3.2 Patient dependency

The empirical work also found the nursing time per patient increased by about two minutes for each dependency point. The average nursing time per session across the three dependency levels ranged from 68 to 88 minutes. This was a narrower

[58] With minimal adjustment for wage rate differences, MRUs, paediatric patients and geographical isolation.

range than the 61 to 97 minutes (across five acuity levels) reported by Freund et al (1998) in the US. Conversely, at Portsmouth the nurse to patient ratios associated with each dependency level predicted an even wider range, from 81 to 122 minutes. Given that some patient-level data were missing, the study's range of times may have been slightly restricted.

The empirical work also contributed to knowledge through validation work on Portsmouth's patient dependency-scoring tool. The tool showed evidence of construct validity as it differentiated between patients on three tests: by eligibility for RSU care, by the type of nurse, and by functional activity (KPS). Yet comparison of the dependency forms and HD notes revealed inaccuracies in nurses' ratings and inter-rater inconsistencies. Options to address these problems include further training, revision of the tool to eliminate the more subjective items, or assigning one nurse for each patient's entire HD session.

Given the concerns about the tool - its complexity and time to complete - another option would be to investigate a simpler classification method, such as, overall low, mid and high dependency groups. Alternatively, one might categorise patients as standard care, acutely ill (or having the potential to deteriorate) or increased dependency (needing more than baseline nursing input). These are akin to levels zero, 1a and 1b in the "Acuity and dependency measurement tool" developed from critical care standards by Harrison (2004). Otherwise, it might be possible to identify patient groupings based on the type of nurse who could deliver all or the majority of the care, as some patient characteristics required RGN input (e.g. necklines).

Furthermore, it might be possible to improve the tool through examination of dependency items' scaling levels and discriminatory power. The nurses wanted the tool to reflect all patient care they delivered. Consequently, some items did not discriminate between patients. For example, "refreshments" was redundant as almost all patients had the same rating. To simplify the tool, some rarely used items could be merged or excluded. A more radical option would be to try to identify key items, for example by statistical techniques such as factor analysis. This would require careful handling because the tool's 19 items comprised a mixture of ordinal scales (of varying length) and unconnected elements. Alternatively, use of the tool to categorise patients by their need or use of nursing time might be improved by using parts of the tool instead of the overall score. It was not designed specifically for this purpose, although some items had actual time estimates. Examples were control of bleeding when taking the patient off dialysis and duration of dressing changes. Other items related to patient risk (e.g. blood pressure) where the link with nursing time was not clear.

A fundamental purpose of the systems to categorise patient heterogeneity (KPS and dependency-scoring tool) was to help assess comparability of patients across centres by better defining the patient outputs. A further issue, beyond the scope of the current study, is the role that patient classifications designed for a variety of purposes might have in attributing staff time to different patients (e.g. by weighting top-down costs). Importantly, for costing purposes, the categorisation needs to discern groups of patients (statistically and meaningfully) who use similar amounts of resources. Staff time might be attributed through a variety of output measures at different levels of aggregation, such as actual staff time, or length of stay, etc. Given the minimal difference between the overall average and that of the sub-group for patients eligible for RSU care, the unit or ward itself might serve as case mix tool.

10.4 Strengths and limitations of the empirical study

Table 10.3 highlights the key strengths and limitations of the empirical study.

Table 10.3 Key strengths and limitations of the empirical research

Strengths	Limitations
Impact of costing methods	
Assessed the effect of i) top-down and bottom-up costing approaches, ii) eligibility for RSU care on nursing costs in chronic HD and iii) using costs averaged across patients instead of the relevant sub-group.	Top-down estimates used single salary points rather than financial data from the Trust (although this helped uncover differences between the two costing approaches). Data collection at a single MRU limits the generalisability of the results. Research limited to nurses, rather the overall costs of a HD session, and did not assess possible role substitution (nurses-clerical worker, e.g. for the clerk's days off when nurses undertake clerical tasks).
Time measurement	
Evaluated three key aspects for two time measurement techniques: ability to capture patient-specific time, feasibility of implementation and validity. Collected 4-weeks continuous data with 80% completion.	On reflection, overly ambitious for a solo researcher to run the techniques in parallel. With additional resources, it would have been possible to give more timely reports to nurses to check and rectify scanning mistakes. Generalisability of feasibility to other settings limited (although piloting on a ward provided valuable insight).
Analyses	
Used analyses appropriate to the repeated data, although the more sophisticated analyses (GEE) did not change the overall conclusions. (Some published time studies appeared to ignore the repeated data.) All models tried to control for constraints on nursing hours (by the number of patients per nurse) and missing data (by the % of study nurses on the shift).	Handling of missing data was limited (controlled for % of study nurses on shift). Imputation of missing data was beyond the scope of the current study. Statistical analyses did not control for the nurse rating the dependency score and could not control for multiple nurses recording a patient's nursing time.
Patient heterogeneity in chronic HD	
Assessed heterogeneity amongst patients on chronic HD and assessed the validity of a new patient dependency-scoring tool for outpatient HD.	The choice of methods to categorise patients was pragmatic and heavily influenced by the opinions of the HD nurses. Further developmental or validation work is required for the dependency-scoring tool and only basic reliability tests were performed.

10.5 Implications for researchers and future research

Based on the preceding findings and discussions, this section provides general suggestions to help researchers plan costing exercises and outlines topics for future research.

10.5.1 Suggestions for costing in economic evaluations

Important lessons of relevance to future costing exercises were learned from the empirical work. First, bottom-up costing required many assumptions and the estimates did not reconcile with the top-down expenditure estimates, although differences were small (up to 10%). Second, despite patient heterogeneity, the overall top-down average cost was almost identical to the bottom-up estimate for the required patient sub-group and so would not have affected the conclusions of the previous economic evaluation.

The definition of a costing approach as top-down or bottom-up is complicated. The top-down approach disaggregates expenditure to activity. In contrast, the bottom-up approach involves aggregation of resource use elements multiplied by their respective unit costs, that in turn may have been derived using expenditure data (i.e. top-down). Moreover, the overall cost for each option in an economic evaluation comprises combinations of labour, materials and capital that could each be valued using either costing approach. Cost estimates need to be fit for purpose and the decisive test is whether the methods used are likely to change the results or decisions.

It would be helpful to undertake pre-study spreadsheet modelling using crude estimates[59] to help plan a costing exercise. One aim would be to get a preliminary estimate of the overall results and thereby identify the potential importance of costing to the evaluation. For example, cost estimates will be particularly important if the cost per QALY is close to a decision threshold, as illustrated in Table 10.4. Such situations will be more complicated to recognise if multiple outcomes are used rather than QALYs.

[59] E.g. from publications, qualitative work, expert opinion, or preliminary work sampling, etc.

Table 10.4 Illustration of implications of a decision threshold on costing

Example result	Cost or	Effect	Need for greater costing 'precision' (decision threshold £30,000 per QALY)
£1,000 per QALY	Low	Large	Unlikely to change decision
£30,500 per QALY			Important as close to decision threshold (examine uncertainty in costs ± effects)
>£100,000 per QALY	High	Small	Unlikely to change decision

A second aim of pre-study modelling would be to identify the main cost drivers - both the major contributors to the overall cost and for probable differences between options - and to estimate their likely impact. This would enable researchers to focus on the cost categories expected to have the biggest impact on the overall results.

A third aim of pre-study modelling would be to investigate whether available costs are suitable for the study context, and to examine trade-offs in choosing between the two costing approaches. Usually it will be quicker and cheaper to use 'routine' unit costs (typically derived using top-down methods) rather than undertaking bottom-up costing. However, the aggregate top-down costs may conceal important differences in case mix or service mix that would bias the study results.

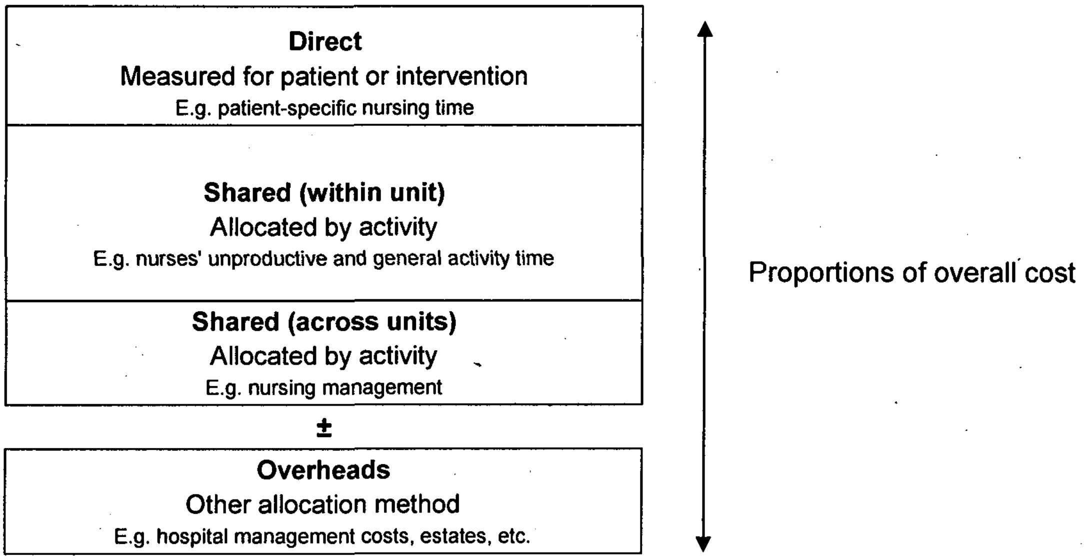
If routine costs are unavailable or appear unsuitable (even with some adjustment), modelling may help show whether bottom-up costing is actually warranted (i.e. is likely to influence the results). Likewise modelling may be useful in determining at what level costs need to be measured since bottom-up costing can be undertaken at the patient-level or unit-level (and single or multiple centres). Costs comprise different components as shown in Table 10.5 and the proportions of these components may vary (see Figure 10.2), with implications for whether measurement is likely to be worthwhile. For example, if direct costs are a relatively large proportion of an overall cost, between-patient variability may influence the results and so in-depth measurement may be useful. Conversely, where shared costs and overhead components that are allocated by activity or other methods form large components, the effect of between-patient variability for direct costs will diminish. Disentangling these components may help in deciding between new data collection and using or adapting existing information.

Table 10.5 Cost components

Cost element	Accounting term*	Typically attributed
Directly attributable to services or patients	Direct cost	Directly to services or patients.
Shared across cost centres or patients	Indirect cost	Based on activity data (e.g. sessions delivered, patient days, etc.).
Overheads for shared support services typically not involved in face-to-face patient contact	Overheads	Based on a 'fair share' not activity data (e.g. by building volumes, or as a last resort by the gross cost of patient treatment services).

Key: * Term used in NHS Costing Manual (Department of Health 2008b)

Figure 10.2 Cost components (for measurement or disaggregation)



It is important to include some degree of data validation for cost categories that are likely to affect decisions. This should include checking for internal consistency to make sure that cost information broadly reconciles, especially as omissions are a major threat to the validity of bottom-up costing. Moreover, when comparing top-down and bottom-up costs, it is important to check whether overheads or other shared costs have already been included in the top-down figure.

Drawing together the points above, Figure 10.3 provides a suggested framework to help researchers choose between costing approaches for economic evaluations. This includes considerations about how to assess whether available top-down costs are appropriate and whether a detailed bottom-up approach is likely to influence the final decision. Finally, to determine the value of pre-study modelling it would be necessary for researchers to monitor where more complicated bottom-up costing affects results and overall decisions.

Figure 10.3 Suggested approach to choosing costing approach

Purpose of costing exercise (economic evaluation ± budget impact)
Likely decision-making criteria such as cost per QALY or cost-effectiveness thresholds.

↓

External validity or generalisability issues
How generalisable do data need to be?

- National or strategic level decisions (e.g. for NICE) need representative costs. Top-down costs are typically more generalisable[60].
- Local decisions need local costs (context specific). Bottom-up costs typically reflect resource use over the short-term (e.g. month, week, shift etc.), but information may be useful to understand activity-cost relationships.

↓

Pre-study 'modelling' - preliminary steps
For the 'intervention' and outcomes of each option under consideration, identify resource use in categories (e.g. labour, materials and capital).
Complete a spreadsheet (example illustrated below) using whatever estimates are available. The comments column can be used to note whether a cost item is the same for all options[61], whether adjustment is possible to make unit cost more appropriate, etc.

A	B	C	D	E	F	G_1	G_2	H
Resource	Resource use	% patients applicable	Unit cost	Unit cost valid for study?	Cost per patient (B x C x D)	% of cost category	% of total cost	Comments
Option 1								
E.g. Staff								
Nursing	1 session	100%	£40 per session	Yes / No / 'Guessed'	£40	80%	20%	Same for all options
Sub-total								
Grand total								

↓

Pre-study 'modelling' - Identify cost drivers
Identify which costs (items or categories) appear to be the major cost drivers both overall and for differences between the options considered.
Do sensitivity analyses using different plausible estimates of unit costs and resource use.

↓

Internal validity checks of unit costs in relation to study context
For resource use for each option and at each centre:

- Do unit costs appear to reflect actual costs (i.e. production costs not charges or tariffs, and do they broadly reconcile with other data)?
- Are costs sufficiently disaggregated for the decision context (e.g. if costing within HRG)?
- Are there possible biases through differences in case mix or service mix? E.g. in using top-down costs, do the aggregated (average) expenditure or activity data conceal differences that would introduce significant bias for the study context?
- Is costing similar at different sites (e.g. like-for-like comparison, similar costing rules applied, same costs included or omitted)? Assessment may be more difficult for aggregated data, especially across multiple centres.
- Are there site selection issues? E.g. are costs at a single site representative of the 'standard' average or most efficient site? For multiple centres, are centres similar (i.e. likely to adopt similar costing procedures)? If not, what can be done to ensure comparability (e.g. valuing resource use at a more disaggregated level)?[62].

↓

[60] Top-down costs are based on actual expenditure averaged over financial year or month etc. which smoothes fluctuations in resource use ± unit costs.
[61] Budget impact analysis requires all cost items to be valued. For economic evaluations, resource use that is identical across options may be noted but not valued.
[62] E.g. compared with the UK, in the US there are more types of health care provider with potentially different costing or accounting rules.

↓

Pre-study 'modelling' - Choice between costing approaches

- Identify which unit cost estimates appear unsuitable for study context (e.g. actual figure unavailable[63], not valid without adjustment, or insufficiently disaggregated[64]).
- Examine whether bottom-up costing might be appropriate, to what degree and for what costs. The key question is how sensitive is the cost (item, category or overall) to changes, for example to correct possible biases? Aspects to consider:
  - What effect does the cost have in relation to decision-making criteria (see Table 10.4)? Broad top-down estimates are suitable if they do not affect the decision. Investigation using detailed bottom-up costing may be necessary if the cost affects the decision.
  - What is the likely variability in the cost (minimum, maximum, average)?
  - How will the following cost components affect the usefulness of measuring the cost::
    - proportions of direct, shared and overheads (see Figure 10.2)
    - fixed or variable costs.

↓

Feasibility issues

- What are the likely barriers to top-down or bottom-up costing? Consider practicalities of data collection for study including any research governance and ethical issues.
- What research resources are available (for data collection, management and analysis)?
- Will more complicated costing offer value for money? Top-down is usually relatively quick and cheap to undertake. However, the level of data aggregation may make it difficult to ensure comparability between centres. Bottom-up costing is usually more time consuming and costly (i.e. data collection at the patient or unit-level for single or multiple centres). This affects the feasibility within multi-centre studies and so bottom-up costing may need to be restricted to a limited number of cost categories.

↓

Final decision - choice of method and level of detail for each cost category

↓

Undertake costing exercise with ongoing data validation

Undertake validation checks for internal consistency:

- Do 'new' estimates appear valid (revisit internal validity checks)?
- Make sure that costs add up (i.e. broadly reconcile with other data). If not, check whether different costs included, different assumptions used, effect of missing data, and for omissions such as 'overheads'.

↓

Monitor effects of pre-study modelling

Evaluate how pre-study modelling affected costing decisions, results and overall decisions.

10.5.2 Recommendations for future research

Based on the preceding discussions, aspects identified for future research and guidance include the following:

1. **Regulatory and guideline forming bodies need to improve aspects of practical costing advice for economic purposes.**

   There needs to be guidance on which hourly rate for staff time represents the appropriate opportunity cost. Linked to this, it would be helpful to survey NHS

[63] E.g. routine costing systems do not cost to the specific intervention or patient group.
[64] E.g. where resources such as staff or facilities are shared across a heterogeneous patient group or multiple outputs.

providers to quantify typical unproductive time for different staff groups to ensure it is not underestimated (e.g. by exclusion of maternity leave). Although there have been a number of important systematic reviews of costing research and guidance, these have not been converted into definitive advice for health economics researchers.

2. **Investigation of the choice between top-down and bottom-up or hybrid costing approaches.**
   - Evaluation of the use of pre-study modelling - monitoring the effect on costing decisions, results and overall decisions:
     - to identify relevant cost drivers (overall and for cost differences) and estimate their likely impact,
     - to decide between top-down and bottom-up costing approaches for different cost categories,
     - to help target research resources and effort.

     This research would help identify whether the choice of costing approach makes a difference to the conclusions of economic evaluations. It should be collated to provide evidence to help formulate better guidance both for costing overall and for specific cost categories. It is important to include explicit assessment of feasibility, validity and, where possible, reliability issues.
   - It would be helpful to understand how and why health economists decide on which costing approach to adopt (both overall and for cost categories). This could be investigated by requesting new projects to include these aspects in their reporting and explored using qualitative methods.

3. **Further investigation of methods to measure staff inputs.**
   From the empirical research, the research effort and costs preclude advocating barcode scanning as a method to collect staff time data for most economic evaluations. Given the challenges and sensitivity of using nurses to self-record their time use, less threatening, cheaper and practical approaches are needed to measure staff inputs. Therefore, it would be advantageous to evaluate hybrid methods (or triangulation). Staff estimates obtained through interviews, questionnaires or formal consensus methods could be validated by limited data collection from time and motion or work sampling observation, as appropriate to the setting. When using two methods, it is important that data collection is

concurrent as this was a fundamental criticism of comparative studies in the literature review. Specific issues for investigation might include:

- Comparison of the effect on costs of measuring staff time on an absolute basis (i.e. minutes) with measurement on a relative basis. This could include investigation of whether patient classifications designed for a variety of purposes might help in attributing staff time to different patients.
- Investigation of recall periods for staff to estimate their time use for both typical patients and unusual workload patterns. This should include questions that can be validated using another method to check whether estimates are for actual or ideal staffing levels.
- Investigation of whether direct care is a good enough proxy for patient-specific care (direct and indirect care combined). In the current study, indirect care was 26% and direct care 74% of patient-specific time, but the latter was only 41% of paid nursing hours (i.e. including 'unproductive' time and general activity).

4. **Categorisation of patient heterogeneity.**

   Before widespread use of the dependency-scoring tool for outpatient haemodialysis could be advocated, further developmental and validation work is required (discussed in section 10.3.2).

10.6 Conclusions

The conclusions are presented in three sections covering costing, measurement of staff inputs, costing and patient heterogeneity in HD. Overall, the thesis has contributed to knowledge about: differences between estimates from top-down and bottom-up costing approaches, methods to collect resource use data on staff time, and methods to attribute time shared across patients. It has also contributed to knowledge about the effect on nursing time of patient heterogeneity in terms of eligibility for RSU care amongst patients on chronic HD. Lastly, it contributed to validation work on the dependency-scoring tool and highlighted further developmental work required.

10.6.1 Costing methods

Guidance on costing for economic purposes advocates a bottom-up rather than top-down approach both in general and for specific costs. Yet the thesis found minimal evidence to support such choices or to determine the consequences for decision-making, and the distinction between the two approaches is not clear-cut in practice.

The empirical research provided evidence challenging the view that bottom-up costing is better than top-down. It found the bottom-up estimates less accurate than the top-down one; in this study, the bottom-up method underestimated costs by up to 10% compared with the top-down cost: 6% due to missing patient-level resource use data, and 4% due to the unit cost using expected rather than actual working hours.

The empirical work suggested that for staff costs, expenditure based top-down estimates and bottom-up estimates cannot reconcile exactly. The two approaches differ by averaging at different points and attribution of nursing time and costs to patients. The bottom-up approach required numerous assumptions. Moreover, it transpired that full patient-level costing in HD using the combined nurse-level (i.e. by grade) and patient-level data would have produced perverse results. This was because, for pragmatic reasons, lower dependency patients tended to dialyse in the evening and at weekends, when nursing costs were higher due to payments for unsocial hours.

In terms of feasibility, patient-level data collection using barcode scanning required considerable research effort, in addition to more than £6,000 for equipment alone, which is simply impractical for many studies especially if multi-centre.

Overall, it was concluded that bottom-up costing cannot be considered a 'gold standard' approach. The thesis suggested a framework to help researchers choose between the two approaches. Issues included the purpose of exercise, generalisability issues, pre-study 'modelling' (to identify cost drivers and assist choice of method), internal validity checks, feasibility issues, data validation, and monitoring the effects of modelling. The findings reinforced the importance of reporting resource use separately from costs. Costing practices could be improved by greater practical guidance on how to value the opportunity cost of staff.

10.6.2 Measurement of staff inputs

The thesis has demonstrated that staff inputs are complicated to measure, analyse and interpret. Coding activity is an inherent limitation of time measurement and affects the comparability of data between centres. None of the techniques to measure staff time is a 'gold standard' and so the choice between methods must trade-off their advantages and disadvantages.

Published studies gave little information about the feasibility of implementing techniques to collect patient-level data. In addition, although difficult to undertake, most studies did little to assess validity or reliability.

The empirical study found that, despite unease amongst staff due to impending reorganisation, it was feasible for nurses to self-record their time using barcode scanners. The majority of nurses found barcode scanning acceptable and experienced few problems. Data were of acceptable validity, although evidence of miscoding suggested that combined patient-specific time was more accurate than the separate direct and indirect care times. Despite widespread use of barcode scanning for other purposes, implementation for time measurement was very challenging. It required more research effort and costs than expected and the scanners suffered numerous technical problems. Overall, barcode scanning requires too much effort to advocate it as a widely applicable method to measure patient-specific nursing time.

The empirical study found work sampling at the patient-level was unfeasible because of difficulties in accurately linking data to patients. For other types of activity data, work sampling still appears to offer advantages because it does not require individual consent and therefore may be feasible in multi-centre studies. Therefore, work sampling may be useful in measuring the overall proportion of time spent on broad activity categories such as direct and 'other' care either as to check staff opinion or to determine whether more intensive data collection might be worthwhile.

The thesis developed a flow diagram of issues to consider in choosing whether and how to measure staff time in a health care setting. Issues concern the purpose of quantifying staff time, activity to be quantified, measurement options available, barriers to implementation, and implications for data and outcomes. All methods should be considered fully, including possible compromises, before arriving at a final decision.

It remains to be seen whether costing staff will become easier as NHS Trusts, encouraged by the Department of Health (2007b), implement patient-level information and costing systems.

10.6.3 Costing and patient heterogeneity in HD

The literature review demonstrated that published economic evaluations of HD in different settings largely ignored the impact of patient heterogeneity on routine nursing costs. The empirical study confirmed there was evidence of heterogeneity between patients eligible and ineligible for RSU care through differences in functional activity (KPS) and dependency scores by eligibility for RSU care.

The study showed that despite much within-group variation, compared with patients eligible for RSU care, those ineligible required some additional nursing time. However, the patient-specific time per session using the average across all patients was similar to that for the sub-group of patients eligible for RSU care. The impact had been diminished because the differences were small in relation to the overall cost per session.

The implications of the differences in resource use and cost depend on how the data are used. MRU nurses coped with the workload within existing resources and might do so even if all patients switched to being ineligible for RSU care. In practice, the overall effect of variations between patients depends on the absolute difference between them, the relative proportions in each patient group and the contribution to the overall cost. From a policy perspective, the results did not overturn conclusions of previous economic evaluations. Moreover, despite the 'case mix' differences, current HD commissioning does not appear to disadvantage service providers. The findings illustrated the problems in identifying the drivers of cost differences. They also lent support to recommendation by Chilcott et al (2003) for 'pre-study modelling' and sensitivity analyses.

In addition, nursing time per patient increased by dependency score and dependency level. The study provided evidence of construct or empirical validity of the dependency-scoring tool, but showed that further developmental work is required. The systems to categorise patient heterogeneity (KPS and dependency-scoring tool) gave insight into the comparability of patients across centres by better defining the patient outputs.

Haemodialysis was a useful case study to examine methods to attribute staff time across heterogeneous patient groups. Other economic evaluations where staff time and patient heterogeneity pose challenges include role substitution, care in different settings, and care by specialist compared with non-specialist staff or units. However, the conclusions cannot be generalised to other settings without further research.

10.7 Outputs from PhD

Outputs from the PhD have been as follows.

Published paper

Nicholson T, Roderick P (2007) International Study of Health Care Organization and Financing of renal services in England and Wales. International Journal of Health Care Finance and Economics 7(4): 283-299.

Conference - contribution

Discussion paper: Nicholson T, Gerard K and Roderick P. (2005) Appropriate estimation of staffing costs in economic evaluations. Health Economists' Study Group, University of Oxford (5-7 January).

Conference - oral presentations

- Nicholson T. (2007) Appropriate estimation of nursing costs for economic evaluations: The challenge of shared resources in haemodialysis. University of Southampton, Faculty of Medicine, Health and Life Sciences Postgraduate Conference (5 June).
- Nicholson T and Roderick P. (2006) Universal coverage with aspects of rationing and performance: the case of England and Wales. 6th European conference on Health Economics, Budapest, Hungary (6-9 July).
- Nicholson T. (2006) Appropriate estimation of staffing costs in economic evaluations: Measurement of patient dependency and nursing time. University of Southampton, Faculty of Medicine, Health and Life Sciences Postgraduate Conference (6 June).

Conference - posters

- Nicholson T, Raftery J, Gerard G, Roderick P. (2007) Costing methods in economic evaluations: The challenge of shared staff resources. Health

Technology Assessment International (HTAi) conference, Barcelona (18-20 June).

- Nicholson T, Gerard K, Roderick P. (2006) Appropriate estimation of staffing costs in economic evaluations: Measurement of patient dependency and nursing time. 6th European conference on Health Economics, Budapest, Hungary (6-9 July).
- Nicholson T, Gerard K, Roderick P, Wolstenholme J. (2005) Appropriate estimation of staffing costs in economic evaluations: Measurement of patient dependency and nursing time. University of Southampton, Community Clinical Sciences Division Annual Conference (30 June).

Appendix 1 Barcodes and barcode scanners

This appendix describes technical details about barcode scanning technology. It synthesises information from a variety of sources listed in the 'References / resources used'.

Barcodes - what are they?
Barcodes encode numbers, letters or special characters typically using combinations of bars and spaces of varying widths, with a 'quiet zone' free from printing at each side. The narrowest bar or space in the array is called the 'X' dimension and it is measured in thousandths of an inch (mils). Larger X-dimensions produce lower density barcodes but require a greater area.

Barcodes do not contain descriptive information. In barcode technology, symbology is the language that enables scanners to read the barcodes and printers to turn the information into labels. Different industries and problems necessitated the development of different symbologies, so there is no standard barcode.

Printing barcodes
The production of printed barcodes requires special software. The printed barcodes are referred to as labels. Special barcode printers and stationery are available, although laser printers and good quality stationery produce acceptable results. Poor quality printing increases the likelihood of the scanners being unable to read or misreading the barcodes.

How barcodes are read - barcode scanners
Barcode scanners read the barcodes by sweeping a light across the label. The bars absorb the light, whilst the spaces reflect the light that the scanner then converts into an electrical signal. The scanner uses the quiet zone to calibrate and hence determine the label width and this zone usually needs to be at least 10 times the width of X-dimension. Varying the height of the bars facilitates keeping the sweep in the symbol area so that the barcode can be read. Typically, the larger the X-dimension the more lenient the barcode is for reading by the scanner. However, as the amount of information coded increases, so the length and height of the barcode label must increase. The scanner also comprises a decoder. This recognises the barcode symbology, checks the content of the scanned barcode and transmits data to a computer.

Types of scanner
There are three main types of barcode scanner. Fixed scanners connect to a host computer or terminal and transmit each data item as the barcode is scanned. Portable batch scanners are battery operated and require batch transfer of the data stored in memory to a host computer. Wireless portable scanners are similar to portable batch scanners except that data are transmitted immediately to the host computer (i.e. in real time).

Barcode symbologies
Symbologies have different characteristics. The main ones are:

- Type of characters used (alphanumeric, numerical, special characters)
- Kind of symbology:
    - Discrete - every barcode character can be interpreted individually without reference to rest of the barcode
    - Continuous - individual characters in the barcode cannot be interpreted individually.
- Length - fixed or variable

- Control (start and stop) characters at the left and right (or bottom and top) of the barcode. These allow barcodes to be read bi-directionally i.e. from left to right or *vice versa*.
- Use of check characters. Typically, these are stripped from the message by the scanner decoder.

The table below shows examples of different symbologies and their uses.

Symbology	**Use**	**Comment**
Universal Product Code (UPC)	Retail and food industry - point of sales	Contains producer and product identification (very compact).
European Article Number (EAN)	Retail and food industry - point of sales	Superset of UPC with extra digits for country identification. EAN-13 also known as WPC (World Product Code).
Code 128	E.g. shipping industry	Uses other characters, barcodes very compact and dense.
Code 3 of 9 (Code 39)	Typically non-food standard - identification, inventory, tracking e.g. in manufacturing	Most popular symbology. Codes alphabet and numbers. Barcodes relatively long.
Interleaved 2 of 5 (ITF)	E.g. shipping industry, distribution and warehousing	Barcodes very compact.
Postnet	Unique to United States Postal Service	Encodes zip codes. Fixed bar and space fixed width, so not strictly a barcode.
PDF417	E.g. on driver's licenses in some states in the US	2 dimensional, high density, non-linear. PDF417 is a portable data file (PDF) rather than a reference number and encodes ASCII or ISO characters.

Code (symbology) used in the thesis empirical study - Code 128

Computer Identics Corporation introduced Code 128 in 1981. Advantages include the following: easy ability to encode all 128 ASCII characters, efficient use of label space as X-dimension is small, and easily readable with high message integrity (due to its message check routines). Code 128 offers a choice of three start characters that determine how the following characters are encoded. Start Code A encodes into upper case alphanumeric and ASCII control characters, Code B encodes into upper and lower case alphanumeric characters. Code C encodes into pairs of numbers (00 to 99), i.e. double density. With a barcode of an even number of numeric characters, Code 128 allows the shortest label length. Code 128 allows short barcode expression on labels and currently is the recommended first option for people designing barcodes and was used in the thesis empirical study (8 numbers long Code C).

References / resources used

- Bar Coding For Beginners (Part No. 20077) 1999 Symbol Technologies, Inc. USA
- Bar Codes - an Overview. http://www.barcode.org.uk/barcode%20types.htm Accessed 21/07/2004 (Application Developments Ltd, Wokingham)
- Symbology. http://whatis.techtarget.com/definition/0,,sid9_gci860633,00.html Accessed 11/10/2004
- Answers to FAQs: What are the most popular symbologies I can use? What are the general symbology characteristics? http://www.mac-barcode.com/faqs/answ3.htm Accessed 11/10/2004 (The Mac-Barcode Company, Portsmouth)
- Code 128 Barcode Specification. http://www.barcodeman.com/info/c128.php3 Accessed 4/11/2004 (Barcodemill.com, Altek Instruments Ltd, Walton).

Appendix 2 Literature search for time measurement studies in health care

Results of literature searches for time measurement studies in health care

Time measurement	Additional barcode scanning studies	
529	61	Total bibliographic details downloaded
143	13	Duplicates excluded
386	48	Abstracts examined
51	-	Potential papers
6	5	Relevant papers for data extraction

Time measurement papers excluded (45)

No:	Areas	Description
5	Activity - Doctors	Activity e.g. proportions of direct care / other time across patients
7	Activity - Nurses	
3	Activity - Mixed staff/patient groups	
2	Activity - Patients	How patients spent their time
1	Barcode Patient ID (transfusion)	Barcodes to aid administration and minimise patient identification errors
2	Consultation duration	
1	Cost of illness study	
6	Drugs	Drug administration
11	Task / procedure	
4	Work arrangements	E.g. looking at interruptions, time waiting for patients and team arrangements
2	Workload measure	Time underestimated by workload measure c.f. working hours Use of nursing workload measure (Excelcare) for staffing in Australia. No info about recording of timings for tasks.
1	PhD thesis	Phase 1 was to include time study secondary data, but stopped due to data quality problems

Additional barcode scanning papers excluded (43)

No:	Areas	Description
13	Laboratory automation	Laboratory procedures and techniques where barcodes aided sample identification or stored the actual test data
6	Records	Inventory / records management
4	Blood transfusion	Barcodes could aid product administration and minimise patient identification errors
2	Drug dispensing / administration	
4	Data collection	Barcode scanners used to collect data from a variety of sources, including for hospital management (clinic waiting times), pharmacy information and nutritional data
3	Study techniques	Barcodes to facilitate study (e.g. to access self-teaching packages)
11	Miscellaneous	E.g. reviews, descriptive re use of barcode scanners

Data extraction table for time measurement studies in health care

1st author, setting	Measured	Method	Subjects	Quality of methods & practicality issues addressed	Purpose
Time & motion					
Oliver (2001) Study 1994-1995 US 84 Family Practices	Physician time: Outpatient visits by African-American c.f. white patients	Time and motion study (trained nurses N =?) 2-days observation at each practice Activity coded using Davis Observation Code (DOC)	134 physicians 89% (n=4454) consecutive patients (agreed) 527 African-American pts 3852 white pts	Validity / reliability / practicality issues of time measurement not discussed. Validity already assessed of DOC	To discover if racial differences in time spent by physicians
Zupanic (2002) Study 1998-1999 US Neonatal ICU	Staff time interactions with infants by patient characteristics	Time and motion study (students N =?) 8-12 hours per designated infant (daytime)	154 infants (1235 hours observation) unknown is on more than one day All staff attending designated infants	Validity / reliability / practicality issues of time measurement not discussed. Sample time discussed and fact that times 'normalised' to 24-hour period may not be appropriate. Presented median not means times by professional (no indication of variance). Unknown whether repeated measures data not taken into account / discussed Found a high correlation between actual and standard hours (0.742 p<0.001).	Characterisation and prediction of time inputs into NICU using infant characteristics (from chart review)
Larson-Lohr (2003) US Wound care and hyperbaric medicine centre	Time for tasks or procedures (38 for wound care and 22 for hyperbaric oxygen therapy) - each the average of 10 observations.	Observation One-month of data collection	For development Nurses N? Patients N?	Face validity checked by panel of nursing experts and informal feedback from 10 centres that tested the productivity/acuity tool. Tested on 708 patients over 65 days to compare predicted and actual staffing.	To develop a productivity/acuity tool for staffing / budget calculations.
Self-recording					
Carpenter et al (2003) Study 2001 England 4 nursing homes at 3 locations	Nursing time Validation of a needs assessment tool (MDS/RAI) and case-mix classification (RUG-III) in nursing home residents	Self-recording (paper based) 24-hours	193 nursing home residents 24 RGNs, 56 Care Assistants	RUG-III system previously tested for validity and reliability using self-recorded time sheets. No validity / reliability checks mentioned for current study. RUG-III groups explained 56% care time variance	To determine whether RUG-III system differentiated between residents receiving low, standard and enhanced RGN care time.
Cromwell et al (2004) Study conducted over 18 months 66 units within 27 hospital psychiatric facilities	Resource intensity (time multiplied by wage weighting) for all staff activity - patient specific and other. Multiple patient characteristics (diagnostic, behavioural, demographic and treatments)	Self-recording (paper based) 7-day data collection at each unit (excluded seeing patients off-unit). Non-patient specific time allocated across all patients (by shifts)	Incl. nurses, therapists, mental health specialists, consultants and non-unit staff whilst on unit Psychiatric patients (4149 Medicare patient days)	Site co-ordinators checked forms for completeness and accuracy. Imputation of one shift's worth of information (mostly night shift) on ~6% patient days (imputed total staff minutes 1.8% > than non-imputed days). Other validity / reliability / practicality issues of time measurement not discussed.	Examined patient characteristics of very high and very low staff intensity groups with purpose of reviewing casemix classification for claims

1st author, setting	Measured	Method	Subjects	Quality of methods & practicality issues addressed	Purpose
Barcode					
Walsh (2003) England Nurse led unit (NLU) c.f. acute ward	Nursing time (patient specific and other). Quality of nursing care using Qualpacs (using observer)	Self-recording using barcode scanners to collect time data 2-weeks	Unknown number of nurses and patients (16-bed NLU, 24-bed ward) 33507 minutes NLU, 11462 minutes acute ward	Lack of time data recording, particularly on ward. Inter-rater reliability assessed for Qualpacs. Validity issues discussed in relation to missing data, although amount of time expected not presented.	Examination of whether nurse-led unit was associated with increase in 'therapeutic' nursing activities
Extra barcode					
Martin (1990) (PhD thesis) US One home care setting	Recorded activities / time and compared with patient classification model	Self-recording using barcode scanners to collect time data 3-weeks	Fifteen nurses Convenience sample of 404 home visits over three weeks (excluded 497 duplicate visits by random selection - to maintain statistical independence)	Explained missing data - 55 visits by one nurse eliminated as unable to collect time data following several instruction sessions. Otherwise validity / reliability / practicality issues of time data not discussed Inter-rater reliability tested for patient classification system. Time data used to validate classification system - Regression analysis: classification system accounted for 42% of the variance in direct care. Discriminant analysis: 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.	To design and test the validity of a patient classification system based on resource consumption for home care
Macfarlane and Lees (1997) Three projects 1. Study 1993 Wessex Neurological Centre (2 neurological wards, ITU, and 2 theatres)	Nurse time for all patients specific and other activities Routine casemix (HRG) and demographic data	Self-recording using barcode scanners to collect time data 2-months	98 WTE nurses Unknown no: patients (96,000 activities recorded) (249 nursing activities)	Overall: Unclear about treatment of missing data. Some info on individual studies, but little information / discussion of validity / reliability issues of time data. Discussed need for openness about aims of study Nine panels of nurses determined absolute min and max timings for activities (in some cases confirmed by stopwatch timings) Start and finish recorded	Examined nurse deployment and grade-mix and casemix relationship. Used to improve efficiency (ward difficulties recruiting skilled nurses and increasing workload (nos: & complexity).
2. Study 1996 Medical records	Medical records staff time for tasks	Self-recording using barcode scanners to collect time data 7-days each site	32 medical records staff (596 hours) Plus further 27 and 42 staff at 2 other sites	One site, appointments staff reluctant to record time (only 30 hours by some individuals). Recommended more training time and explanation about study purpose.	To identify times for tasks and proportion of speciality specific work - with view to decentralising some work.

1st author, setting	Measured	Method	Subjects	Quality of methods & practicality issues addressed	Purpose
3. Consultants (medical and surgical)	Consultants time (excluded private patients) - across 19 activities i.e. not patient specific	2-weeks	6 consultants (322 hours over 42-staff days)	3 other consultants excluded - one on leave, one difficulty with barcode pen and one declined. Some gaps in data collection recorded manually by doctor (added to database if sufficiently specific)	To test barcode technology and support of staff for exercise (re workload)
Holmes (1997a) US 5 Special dementia care units (SCU) c.f. 5 nursing homes (non-SCU) (randomly selected)	Staff recorded time (10 staff groups incl. nurses, therapists etc) Staff self-report (? interview) on residents morning personal care provision (scored 0-3). Data on residents over 3-4 weeks (interviews with residents, questionnaire and notes review)	Self-recording using barcode scanners to record time and activities 7-days at each unit	336 residents: 97 SCU, 215 non-SCU Unknown number of staff Units matched to a degree through sampling (20 residents at units) and scoring of Mini Mental Status examination	'Internal' monitoring identified likely unrecorded actions (? how). Barcodes used at start and end of activity. Missing data imputed by system and report generated to use for supervision. System also monitors for multi-tasking and adjusts scores (?how) No validity studies available for this InfoAide system (but previously accounted for more service time than other approaches). Considered sufficiently valid to use for recalibration of New York States nursing home reimbursement system. Trainers available on-site over 7-days of data collection - with rapid feedback to staff re anomalies. Data for 97% residents in SCU and 94% in non-SCU Study to test convergent construct validity c.f. three hypotheses. Correlation between staff self-report and residents personal care time. Data not presented on extent of missing data etc. (Repeated data taken into account using ANCOVA)	To examine whether there were differences between service inputs to residents in SCUs and non-SCUs and to test data collection methods
Blount (1999) Study 1992-1994 (PhD thesis) US One hospital (33 hospital units)	Nurse recorded time c.f. 'standard time' from six Medicus Patient Classification System (PCS) categories (norm prorating)	Self-recording using barcode scanners to record time and activities 10-day study periods	Nurses (N=?) 3439 patient bed days (23,263 hours)	Validity / reliability / practicality issues of time data not discussed Nurses categorised patients prospectively, rather than on care received. Indirect care times deleted as not part of standard hours projections.	Examined correlation between actual and derived standard nursing care hours - and hence validity of PCS

Appendix 3 ISHCOF: Renal services in England and Wales

Nicholson T and Roderick P (2007)

International Study of Health Care Organization and Financing of renal services in England and Wales.

International Journal of Health Care Finance and Economics 7(4):282-299.

Available from http://www.springerlink.com/content/cl74105463u2816l

Abstract

In England and Wales, the quantity and quality of renal services have improved significantly in the last decade. While acceptance rates for renal replacement therapy appear low by international standards, they are now commensurate with many other northern European countries. The major growth in renal services has been in hemodialysis, especially at satellite units. Health care is predominantly publicly funded through a tax-based National Health Service, and such funding has increased in the last 10 years. Improvements in health outcomes in England and Wales are expected to continue due to the recent implementation of standards, initiatives, and monitoring mechanisms for renal transplantation, vascular access, and patient transport.

Keywords

Renal replacement therapy · Health expenditures · Financing, organized · Health services, needs and demand · Kidney failure, chronic · United Kingdom

JEL Classifications H51 · I10 · I11 · I18

Appendix 4 RSU study - Nursing time for MRU and RSU type patients

In the RSU study (Roderick et al 2005), the researcher (TN) undertook preliminary investigations to explore time variations across patients. She elicited the views of five key senior nurses at different units. Open-ended telephone discussions were used to identify characteristics that demarcated patients who needed more nursing time than the average during haemodialysis (HD). Then the nurses were sent a full list and asked to rate the importance of the characteristics, and if possible rank them or identify any extra ones. In addition, two of the nurses were asked to estimate the approximate amount of extra time required on each aspect both for qualified nurses (Registered General Nurses, RGNs) and for support staff (Health Care Assistants (HCAs) or the equivalent, e.g. health care support workers or nursing auxiliaries).

The table overleaf shows the results. The nurses identified and rated many indicators of need for extra nursing time. They considered some to have an indirect impact (e.g. communication was only a problem if the patient was new). It was not possible to pool the nurses' rankings, as there were inter-nurse variations in both factors and their relative importance to RGNs and HCAs. These variations were partly due to differing working practices across units. Some factors were less relevant at particular units (e.g. the proportion of non-English speaking patients). Whilst some aspects affected all nurses, a few predominantly affected the RGNs (e.g. handling adverse events, new patients and responding to the patients' blood results etc.).

Whilst it was relatively simple to identify the factors, it was hard to work out the time implications. Since patients typically have more than one characteristic, nurses found it difficult to estimate the times for single factors. Some factors overlapped, for example removal of excess fluid and counselling patients to adhere to their fluid restrictions. Time estimates varied from actual minutes to additional percentages or multiples of time. Overall, the data were difficult to interpret without information on typical times for standard tasks such as getting a patient on to HD.

Results: Opinions on factors demarcating patients on haemodialysis needing more than average nursing inputs

Factor	Important Overall	Important Ratings*	Important to RGNs	Important to HCAs	Approximate extra time**	Comments
Mobility (e.g. wheelchair)	Yes	Yes (5)	(Yes)	Yes	25%, may need 2 nurses	Also extra time arranging transport
Removal of excessive fluid	Yes	Yes (4)	Yes	Yes	30-60 mins on machine or 35% time	Due to patient non-compliance with fluid intake - especially for patient's first HD session of the week. Requires extra time on machine and patient monitoring
Vascular access problems	Yes	Yes (3)	Yes	Yes	x3 or 30-90 mins New fistula x2 or 30-90 mins Necklines x? (some units use 2 nurses)	60 mins if urokinase needed to improve blood flow
Co-morbidity e.g. diabetes	Yes	Yes (4) No (1)	Yes	Yes	Diabetes 25% Dressing 25-30%	Diabetes (monitoring, mobility issues if amputee etc), Infections, Dressings
Recent in-patient stay	Yes	Yes (4)	Yes		Unknown (depends on needs)	
Acute/recent event (e.g. fall)	Yes	Yes (3)	Yes	Yes	Unknown	
Adverse events (e.g. allergy to dialyser membrane)	Yes	Yes (3)	Yes		x1 or 15-20 mins	Very rare occurrence
New patient (to unit or HD)	Yes	Yes (3)	Yes		x2 for first week	Extra documentation, monitoring, and reassurance
Responding to pre- and post-HD blood results, blood pressure, weight changes etc.	Variable	Yes (2) No (1) Variable (2)	Yes		2-3 mins	Other factors more important e.g. removal of excessive fluid or other therapies
Communication problems e.g. non-English speaker, unable to speak (e.g. stroke), blind, deaf	Variable	Yes (1) No (2) Variable (2)	?	(Yes 1)	Unknown	Only really a problem if new patient
Living arrangements (e.g. alone, with others)	Variable	Yes (1) No (1) Variable (1)	Yes		Unknown	Via social problems
Elderly (e.g. ≥ 70 years)	No	Yes (1) No (2)			10%	Usually only via other factors
Male / female	No	No (3)			N/A	Only via other factors
Ethnicity	(No)	Yes (1) No (3) Variable (1)			N/A	Only via other factors - e.g. typically Asians have poorer vascular access due to narrower veins
Additional issues from nurses					Up to ?80 mins	Stopping bleeding from needle sites
			Yes		Unknown	Other therapies e.g. iron injections, blood transfusions Social / psychological problems and contacting other agencies
					Unknown	Arranging patient's holiday HD (documentation, blood tests, etc)

* By five senior nurses, but not all answered so figures do not always sum to five

** From two of the nurses

Appendix 5 Literature review - Data extraction table for costing or economic evaluations comparing of HD across settings

1st author & year	Base year	Study type	Treatments	Setting	Resource use data	Unit cost data & method	Variation (variance etc)	Costs adjusted for pt heterogeneity	Complications	Staff & apportionment	Transport	Comment
Agar (2005)	2003	Costs	Unknown duration Dx NHHD (6) RSU (low acuity, LC) Pts with 12-months uninterrupted HD	Australia (one area - one RSU)	Study Unknown whether patient-level: 10 HHD pts NK RSU pts	Finance (expenditure & receipts) Methods NK (? averaged)	No Comment only on nurse/pt ratios	Yes - in choice of RSU	No Hospitalisations mentioned but not comparable pt group	Nur Yes (? averaged) Drs No (no fee for Dx - work included part of salary) Tec Yes (NK alloc)	Pt No Staff NK	
Bjorvatn (2005)	? 2002	Costs	All RSU patients MRU, RSU	Norway 3 RSU (3 MRU)	Questionnaires / interviews (staff (? no) & pts) Unit-level data RSU 12 pts	Finance & Govt Method NK Not all explained	No, but incl. sensitivity analysis	No	No	Nur No Drs No Tec No	Pt Yes (& time) Staff Yes	
De Wit (1998)	1996	CUA (Markov model)	≥ 3 months Dx MRU HHD & LCHD PD (CAPD & CCPD) Tx	Netherlands 13-16 hospitals	Study & literature, Registry data MRU 46 pts HHD & LCHD 23 pts CAPD 59 pts CCPD 37 pts	Literature, Finance, Fee schedule, Govt	Sensitivity analysis	No (age adjustment for outcomes)	Yes	Nur Yes (alloc NK) Drs Yes (average for duration on treatment) Tec NK	Pt Yes Staff NK	Dr fees irrespective of location
Gonzalez-Perez (2005)	2001	CUA (Markov model)	(model) MRU RSU HHD (3 & 6)	UK	Systematic review	Literature	Sensitivity analysis	No (mentioned) (Yes for QALYs)	Yes (assumed MRU = RSU)	Alloc NK Nur Yes Drs Yes Tec NK	Pt No Staff No	
Jassal (1998)	1991	Costs & QoL	All patients MRU Specialised RSU	Canada 1 MRU 1 special RSU	Note review 37 pts	Finance & Govt Method NK	No	No	Yes	Nur Yes (NK alloc) Drs NK Tec NK	Pt Yes Staff NK	Comparator rehab / chronic care (in-pt) unit. Before & after (pts = own control, short follow-up)

Key: CAPD Continuous Ambulatory Peritoneal Dialysis CEA Cost-effectiveness analysis CUA Cost utility analysis Drs Doctors Dx Dialysis Govt Government
HDF* Hemodiafiltration or acetate free biofiltration HHD Home HD IPD Intermittent Peritoneal Dialysis LCHD Limited care HD MRU Main renal unit (i.e. hospital, HD unless otherwise specified)
NHHD (no:) Nocturnal home HD (no: nights per week) NK Not known Nur Nurse PD Peritoneal dialysis Pt Patient QoL Quality of life RSU Renal satellite unit (HD unless otherwise specified)
SCHD Self-care HD Tec Technicians Tx Transplant

1st author & year	Base year	Study type	Treatments	Setting	Resource use data	Unit cost data & method	Variation (variance etc)	Costs adjusted for pt heterogeneity	Complications	Staff & apportionment	Transport	Comment
Lee (2002)	2000	Costs	≥ 6 months Dx MRU RSU HHD SCHD PD	Canada (one area)	Study 166 of 332 poss pts: 88 MRU pts 31 RSU pts 8 HHD 1 SCHD 38 PD pts	12 months data for 124 / 166 pts Top down	Yes (95% CI)	Yes (Charlson index)	Yes	Nur Yes (workload measurement unit - Southern Alberta Renal Program (SARP) Database - no details) Drs Yes (per visit) Tec Yes (averaged)	Pt No Staff NK	Dealt with missing data
Lim (1999)	1996	CEA (life years saved) ? model	MRU HHD CAPD IPD	Malaysia	Ministry of Health costs NK how many patients ?? Primary data or only summary data used	Costing study - Ministry of Health costs except IPD - random sample (31 of 407 pts) in 1996	Sensitivity analysis	NK	Yes (hospital-isations), No for co-morbidities	Nur Yes NK Drs ? included Tec Yes NK	Pt No Staff NK	Life years saved from Registry data. MRU more cost-effective than HHD due to discounts on consumables.
Lindsay (2003, 2004) Kroeker (2003)	2001	CUA	≥ 6 months Dx HHD (5-6) NHHD (5-6) HD (HHD, SCHD, MRU, RSU, PD)	Canada (one area)	Note review HHD (10) NHHD (12) HD (22 matched controls (incl. modality as far as poss) ? location) Retrospective (12 month) + 6 months treatment. Each patient acted as own control	Finance, suppliers	Yes (95% CI for cost/QALY NK how derived) not for other costs	Yes (matched HD controls, but not at one location)	Yes	Nur Yes (workload measure -Ambulatory Resource Measurement System - no details) Drs Yes (per contact) Tec Yes (NK alloc)	Pt No Staff NK	Dr fees irrespective of location Poor study design
McFarlane (2002)	2000	Costs	≥ 3 months Dx NHHD (5-7) MRU	Canada (one area, 1 MRU)	Note review NHHD 33 pts MRU 23 pts (16 = SCHD)	Finance Govt fees schedule	Yes (? SD)	Matched cohort	Yes	Nur Yes (alloc NK) Drs Yes per contact Tec NK	Pt NK Staff NK	Drs weekly fee regardless of use. Included assessment of record accuracy (5%).
McFarlane (2003)	As above	CUA	All details as above	As above	NHHD 24 pts MRU 19 pts (13 = SCHD)	As above	Bootstrap 95% CI	As above	As above	As above	As above	As above

Key: CAPD Continuous Ambulatory Peritoneal Dialysis CEA Cost-effectiveness analysis CUA Cost utility analysis Drs Doctors Dx Dialysis Govt Government
HDF* Hemodiafiltration or acetate free biofiltration HHD Home HD IPD Intermittent Peritoneal Dialysis LCHD Limited care HD MRU Main renal unit (i.e. hospital, HD unless otherwise specified)
NHHD (no:) Nocturnal home HD (no: nights per week) NK Not known Nur Nurse PD Peritoneal dialysis Pt Patient QoL Quality of life RSU Renal satellite unit (HD unless otherwise specified)
SCHD Self-care HD Tec Technicians Tx Transplant

1st author & year	Base year	Study type	Treatments	Setting	Resource use data	Unit cost data & method	Variation (variance etc)	Costs adjusted for pt heterogeneity	Complications	Staff & apportionment	Transport	Comment
McFarlane (2006)	2003	CUA (Markov model)	NHHD MRU	Canada	Literature review	Literature review	Yes (incl. probabilistic sensitivity analysis)	NK (model for both groups)	Yes	Nur NK Drs NK Tec NK	Pt NK Staff NK	
Mohr (2001a, 2001b)	1998	Costs (model)	Centre HD (3) Short daily centre HD (?freq) NHHD (?freq) HHD (short daily, ? freq)	US	Literature review and expert opinion	Govt / National	Extensive sensitivity analyses (incl. cost of Dx)	NK	Yes	Nur NK Drs NK Tec NK	Pt Yes Staff NK	
Mowatt (2003)	2001	CUA (Markov model)	HHD RSU MRU	UK	Literature review	Literature review	Sensitivity analysis	No (mentioned)	Yes	? all averaged Nur Yes Drs Yes Tec Yes	Pt Yes Staff NK	
Piccoli (2004)	? 2001	Costs	HDF* MRU LCHD (3 or 7) HHD (3 or 7) APD	Italy (one area)	? Unit-level (stated bottom up) ?31 pts (not clear)	? Finance	No	NK	NK	Alloc NK Nur Yes Drs Yes Tec NK	Pt NK Staff NK	
Roderick (2005	2000	Costs and QoL	MRU RSU	England & Wales	Notes review (pt-level) & unit-level (interview & questionnaire) MRU 335 pts RSU 394 pts	Literature, Govt, suppliers	Yes (95% CI & sensitivity analysis)	No (discussed)	Yes	Nur Yes (averaged) Drs No Tec No	Pt descriptive Staff descriptive	Overall costs not presented
Soroka (2005)	?	Costs - break even no: of RSU pts	≥ 3 months Dx MRU, RSU	Canada (one area)	Unit-level & pt-level MRU 198 pts RSU 10 pt	Fee schedules Finance	No, but incl. sensitivity analysis	Matched cohort - travel & drug costs (latter ?37+38 pts)	Some (standby)	Nur Yes (? method to measure direct care hours) Drs Yes Tec Yes (billed)	Pt Yes Staff NK	Some Drs fee schedules per week regardless of use
Tediosi (2001)	1996	Costs	All patients MRU RSU PD	Italy (NHS) ? 24 / 9 hospitals	Unit-level (survey) 864 MRU pts 107 RSU pts 436 PD pts	Finance Top-down	No	No (mentioned)	No	? all averaged Nur Yes Drs Yes Tec ? Yes	Pt No Staff No	

Key: CAPD Continuous Ambulatory Peritoneal Dialysis CEA Cost-effectiveness analysis CUA Cost utility analysis Drs Doctors Dx Dialysis Govt Government
HDF* Hemodiafiltration or acetate free biofiltration HHD Home HD IPD Intermittent Peritoneal Dialysis LCHD Limited care HD MRU Main renal unit (i.e. hospital, HD unless otherwise specified)
NHHD (no:) Nocturnal home HD (no: nights per week) NK Not known Nur Nurse PD Peritoneal dialysis Pt Patient QoL Quality of life RSU Renal satellite unit (HD unless otherwise specified)
SCHD Self-care HD Tec Technicians Tx Transplant

Appendix 6 Ethical, research governance and data protection issues

This appendix describes the ethical, research governance and data protection issues. The study received approval from the Southampton and South West Hampshire Research Ethics Committee (A) (05/Q1702/83). It received R & D approval from the Southampton University Hospitals NHS Trust and Portsmouth Hospitals NHS Trust.

Nurse consent

Barcode scanning required the consent of individual nurses. In contrast, consent for work sampling was at the unit-level (senior nurse manager). The observer could not avoid seeing particular nurses (i.e. consented or not) and data collection was not on an individually identifiable basis. Posters were displayed to make staff (and patients) aware that they were being observed as part of a research project.

Patient consent

An information sheet about the study was offered to patients and their visitors, and the researcher was available to discuss issues if required. Patient consent was not necessary for two reasons. First, the study did not affect the patient's care and so to ask for consent may have been unnecessarily stressful to patients. Second, since the study's purpose was to collect nursing time data it was important to capture this information from the patient's first contact (e.g. admission) - a time at which they likely to be more heavily dependent on nursing input.

Data protection issues

On the SUHT ward, each day or before each shift, the researcher obtained a list of patients and entered these *via* a standalone computer onto a removable password protected 'memory stick'. On the renal units, updating of the patient list was only necessary when there were changes.

To assist the nurses tracking patients, data collection forms used patient names rather than the patient ID or hospital number. Barcode lists comprising each patient's direct and indirect care barcodes were mail merged below his / her name. The lists were printed on a laser printer connected directly to the computer (i.e. not networked printer). To cope with new admissions during the shift, sheets of unassigned barcodes were available. Subsequently the researcher assigned these patients unique identifiers and amalgamated them into the main lists.

A 'memory stick' was used to store the subject-identifiable data for mail merging. After the study, the researcher overwrote the memory stick with 'junk' data. Data entry was against numerical identifiers not the patient or nurse name. Electronic data were stored on a password-protected computer and data were backed-up at least daily during the data collection/downloading phase.

Paperwork was stored securely. After data entry, information that could identify subjects was removed from paperwork and shredded. Long-term data storage is at Southampton, in accordance with both the Southampton and Portsmouth Trusts' Research Governance policies at the end of the study.

Appendix 7 Barcode scanning feasibility issues (pre-piloting and costs)

This section outlines the preliminary lessons learned about using barcode scanning to measure nursing time inputs. The whole process was very time consuming. For example, it should have been simple to set-up the barcode scanners, particularly after help from the supplier; however, set-up took about 12 working days. Furthermore, the NHS organisations were undergoing major restructuring, which often delayed negotiations about access to the ward and renal units.

Lessons learned from previous barcode scanning projects at Southampton
Section 3.4.2.1 outlined lessons learned from the previous barcode scanning projects conducted by SUHT. Key points for the current study were:

- Barcode scanning can provide very rich data that can be used for multiple purposes (i.e. of benefit to the researcher and unit).
- Frequent downloads could help minimise scanner problems.
- To reduce the burden, both to administration and to data collectors, the number of different barcodes should be kept to a minimum.
- Staff who would receive insufficient training (e.g. agency staff) should be excluded from data collection (although this was necessary anyway due to Ethics committee requirements about staff-consent).
- The duration of data collection should be 'limited' (e.g. weeks not months) as in Neurosurgery the amount of time captured had increased but accuracy had decreased.
- The co-ordinator should visit the data collectors frequently to monitor progress and give feedback and to increase staff involvement.

Barcode scanning equipment
The barcode scanners previously used at SUHT were no longer suitable because insufficient scanners functioned reliably. Some scanner batteries would not hold their charge and other scanners had proved temperamental to download. Replacement scanners were unavailable because the model (Symbol Datawand III) had been superseded. Furthermore, the software to extract and analyse the barcodes (1.1 Resource Analyser 17/6/96) had been developed specially by a private consultancy (Secta) and would not work with current versions of Windows operating systems.

The scanners for the empirical work required the following specifications: easy to use; small and lightweight (to fit easily into a pocket); portable and cheap (maximum £100 per scanner). This proved the first challenge. It was a steep learning curve to understand the technical aspects. Despite extensive searching on the internet and discussions with barcode scanner suppliers, it appeared that few manufacturers produced barcode scanners to the required specifications. Many scanners were capable of sophisticated data capture and consequently the size and cost were greater. Furthermore, most suppliers seemed disinterested in a 'small' and/or Public Sector order. Moreover, barcode scanner models changed frequently and inventories were often out-of-date. Consequently, the first attempt to order scanners failed because there was insufficient stock available (worldwide) to supply the 30 required. This was extremely frustrating as the newer model, though more compact, had a lower storage capacity. In turn, this made it unfeasible to use both start and stop scans to help identify mistakes when nurses forgot to scan the end of activities.

Finally, 50 scanners (Symbol CS 1504 Consumer Memory Scanner) were ordered. This allowed two scanners per nurse per 8-hour shift based on an estimate of the likely number of scans. The order also include spare capacity, which was important as it transpired that the scanners had a short guarantee (90-days) and product support from the manufacturer was unlikely to extend much beyond five years.

The barcoding technology for the new scanners had improved on the Datawand III, with the following advantages for the CS 1504 scanners:

- recording to the nearest second (not minute),
- easier to use ('point and shoot' rather than a wand that had to be manually swiped across the barcode)
- less training required (see point above).

Initial set-up

The whole set-up procedure to prepare the scanners and establish procedures to download and handle the data was complicated and took more time than anticipated. Initial consultancy from the scanner supplier (1/2 day) was helpful and worthwhile expenditure. For example, communication between the scanner and computer was via a special USB/9-pin serial connection cable and required searching for additional drivers on the scanner manufacturer's website. It transpired that the necessary software to set the scanner time/date, save barcode data files and export data was not included in the scanner price. Then the recommended software (MiniPro 1.0, 2004) would not export dates properly because they were stored in US date format (e.g. 01:00 AM/PM rather than using the 24-hour clock). The supplier could have written specific software to resolve this, but it would have incurred additional expenditure (£650). This was not an attractive option because the bespoke software used previously at SUHT had required ongoing support from a consultancy and became obsolete when the operating system (Windows) changed. Instead, after many hours of trial and error, a solution was found by exporting the barcodes as text (into Excel). In addition, the format of the data made it time consuming to establish validation routines and download the scanner's device identifier as a data check (e.g. the scanners did not differentiate between 12 mid-night and 12 mid-day; all data were stored as 12 AM).

Production of printed barcodes (referred to as labels) required further software (Loftware Label Manager 7, 2003) from the scanner supplier, 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 waiting a reply from the consultant and eventually it took technical advice direct from the software manufacturer to resolve the problem by supplying a different software licence key.

Risk assessment - Radio frequency interference

The CS1504 Consumer Memory Scanner owner's guide states that the equipment complies with the limits of a Class B digital device (Part 15 FCC rules) designed to provide reasonable protection against harmful interference in a residential setting and the ICES-003 Class B and European Union EMC Directives. Furthermore, the University Laser Safety Officer confirmed that although the scanners were laser devices, they did not require any safety measures.

Barcode scanners - equipment and other costs

The table below details the costs of the barcode scanners, other associated equipment and software, and researcher's accommodation costs during data collection. As shown, 11% of expenses for scanning equipment were unexpected.

Barcode scanners and associated costs

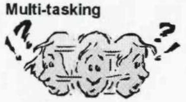
Total cost (£)	Comment	Supplier
Expected - scanning equipment		
4,300	Symbol CS1504 USB memory scanner kit (50 x £86)	Zetes Ltd
32	36" beaded link nickel chains - for scanners (for 50)	Niceday
750	Consultancy (1 day): On-site installation and set-up of scanners, software and training	Zetes Ltd
413	Loftware design and print module (V7.2) labelling software	Zetes Ltd
125	Loftware maintenance - initial 12 month support	Zetes Ltd
5,620	**Sub-total**	
Expected - other expenditure		
633	Hospital accommodation (during data collection) SUHT (3 weeks at £55 per week) = £165, Portsmouth (44 x £10.63 per night) = £467.72	SUHT & QAH
633	**Sub-total**	
6,253	**Overall sub-total for anticipated expenditure**	
Unexpected - scanning equipment		
250	Loftware maintenance - additional 12 month support for two years (required because fault with or loss of USB software key incurred cost of new software)	Loftware
54	MiniPro (barcode extraction) software	MMR Software
337	Batteries for scanners (Energizer EXP76 x 400 - equivalent to 2 battery changes per scanner)	Battery Force
641	**Sub-total (an extra 11% for scanning equipment)**	
6,894	**Overall total**	

Appendix 8 Coding of tasks at Portsmouth

PATIENT-SPECIFIC		NOT PATIENT SPECIFIC
DIRECT CARE Face-to-face with patient	INDIRECT CARE Not face-to-face with patient	GENERAL ACTIVITY
		General cleaning / testing including machines i.e. disinfection / bleaching / citric acid etc Handover
Escorting patients with mobility problems	Making up trolley for patient	Building machines
Putting patient on to dialysis machine		Checking equipment (O_2 / suction etc)
Observations / taking blood / dressings etc	Preparing blood forms	Noting times off dialysis BUT if extra time talking to patient = Direct care
Machine problems (when patient present)		General tea round
Taking patient off dialysis, stripping lines, cleaning machine, chair & trolley, clearing rubbish	Computer work (patient specific)	Computer work for batches of patients (i.e. dialysis sheets)

PATIENT-SPECIFIC		NOT PATIENT SPECIFIC
DIRECT CARE Face-to-face with patient	INDIRECT CARE Not face-to-face with patient	GENERAL ACTIVITY
	Making phone calls (patient specific)	Receiving phone calls But patient specific actions after call = Indirect care Making non-patient specific (e.g. stores)
		Staff meal / personal breaks
		Water sampling

Multi-tasking



The scanners cannot record multi-tasking (each time you scan a barcode you end the previous activity)

General activity	For batches of patients e.g. booking transport for a group of patients
Otherwise,	For patient-specific multi-tasking ... use the code that will attribute your time to the most time consuming / needy patient in the group.

University of Southampton

School of Medicine Community Clinical Sciences Division

Applied Clinical Epidemiology Group

Public Health Sciences & Medical Statistics
Mailpoint 805, Level C, South Academic Block
Southampton General Hospital
Southampton SO16 6YD
Tel: 023 80 796530 Fax: 023 80 796529

Information sheet for nurses
Study to measure patient dependency and nursing time

You are being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Thank you for reading this.

What is the purpose of the study?
Patients vary in their need for nursing care. Some are highly dependent, whilst others need minimal help. We want to measure the nursing time needed for different levels of dependency. Your unit can use this information to help plan the staffing levels needed for the mix of patients.

There are several ways that we could measure nursing time. Each method has to balance how easy it is to collect the data and how good the data are. The study will help us understand about these differences. The first, widely used, method is to ask staff to give an informed guess as to how their workload spreads across patients. We will compare this 'usual practice' with two methods to measure actual nursing time.

The second method is work sampling. This involves a researcher (Tricia Nicholson) walking through the unit at randomly scheduled times. She will note how often nurses are in face-to-face contact with patients to assess the share of time spent on direct patient care.

The two methods described above will have almost no impact on your day-to-day work. Each shift we will need a small amount of information from the nurse in charge of the unit. For example, we will need to know whether there were any new patients. In addition, a few senior nurses will rate the patients' dependency.

We would like your help directly with the third method, bar code scanning. We describe below what this involves. We plan to collect about 2-weeks worth of data after you have had a couple of shifts to get used to what you need to do.

Do I have to take part?
We are inviting all the unit nurses to do the bar code scanning. However, it is up to you to decide whether to take part. This information sheet is yours to keep. If you do decide to take part, we will ask you to sign a consent form and will give you a copy. You will be free to withdraw at any time and without giving a reason. In this case, we will use any data you have collected up to that time, unless you specifically ask us not to do so.

Your nurse manager knows about the study and supports it. However, whether you choose to take part in the bar code scanning, or change your mind later, is a matter between you and the research team. It will not affect the way your manager appraises you.

What will happen to me if I take part?
Bar code scanning is very simple. Each time you scan an activity's bar code, the scanner logs the time. Tricia Nicholson, the researcher will come to the unit to talk to you about the study and answer any questions. She will show you how to use a bar code scanner (see the pictures over the page).

Appendix 9 Nurse information sheet


Scanning a bar code


Hand held scanner

Each shift we will give you a bar code scanner and a small booklet that contains the bar codes. Every time you start a new activity for a patient, you should look up their name in the booklet. You will find two bar codes beneath their name. If you are with the patient, you should scan the 'Direct care' bar code. If you are away from the patient, you should scan the 'Indirect care' bar code. When you start an activity that is not for a specific patient, you should scan the separate 'General activity' bar code. Examples of these activities are admin tasks and meal breaks. Tricia will talk to you about what tasks you should scan for each activity.

When you collect the data, Tricia will be on call if you need help. If you make a 'mistake' such as forgetting to scan the bar code or using the wrong code, do not worry. You can note comments in the bar codes booklet. This will help us to learn how easy it is for you to collect the data and we will use this along with the work sampling records.

At the end of the data collection, Tricia will give you a short anonymous questionnaire about your views on the data collection methods.

Will the information I give be confidential?
We will not pass on information you collect to anyone outside the research team. All paper and computer files will be kept securely. Data held on computer will not be personal and will use study numbers not names.

Please note that Tricia Nicholson is a nurse and as such, must keep to the code of professional conduct. We do not expect there to be problems but should a situation occur that requires action, she will tell a suitable member of staff.

What will happen to the results of the research study?
We will feedback the results to your unit. In addition, we plan to publish the results in peer-reviewed journals to inform health care workers. We will not identify individuals in any report or publication.

Who is funding and organising the research?
Funding for the lead researcher alone (Tricia Nicholson) is from the Department of Health's National Coordinating Centre for Research Capacity Development. The research team includes others at the Universities of Southampton and Oxford.

Contact for further Information
If you would like any more information about this study, please contact:

Tricia Nicholson, Senior Research Fellow
Public Health Sciences & Medical Statistics
Mailpoint 805, Level C, South Academic Block
Southampton General Hospital
Southampton, SO16 6YD

Tel 023 80 794775
Mobile 0781 638 7740
Fax 023 80 796529
Email apn@soton.ac.uk

Thank you for reading this information sheet. We hope that you are interested in this study.

Appendix 10 Nurse consent sheet

University of Southampton

School of Medicine Community Clinical Sciences Division

Applied Clinical Epidemiology Group

Public Health Sciences & Medical Statistics
Mailpoint 805, Level C
South Academic Block
Southampton General Hospital
Southampton
SO16 6YD
Tel: 023 80 796530
Fax: 023 80 796529

Nurse identification number for this study ☐☐

Local contact
Tricia Nicholson (Mrs)
Tel: 023 80 794775 (direct) Mobile 0781 638 7740
Fax: 023 80 796529
Email: apn@soton.ac.uk

CONSENT FORM

Title of Project: **Measurement of patient dependency and nursing time**

Name of Researcher: **Tricia Nicholson**

Please initial box

I confirm that I have read and understood the information sheet dated 16/8/05 (Version 2) for the above study and have had the opportunity to ask questions. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without legal rights being affected. ☐

I understand that the information I provide will be transferred to and stored on a password protected computer. I understand that the results of the study will not identify me by name. ☐

I agree to take part in the above study. ☐

....	....	....
Name of nurse	Date	Signature
Name of person taking consent (if different from researcher)	Date	Signature
Name of researcher taking consent	Date	Signature

1 copy for nurse, 1 copy for researcher

Southampton & SW Hampshire Research Ethics Committee (A) 05/Q1702/83 QAH Version 2 16/8/05

Appendix 11 Nurses' acceptability questionnaire

A questionnaire to assess acceptability to nurses of barcode scanning and work sampling was developed. This included questions adapted from a non-validated questionnaire used in a previous (unpublished) study in primary care. It was necessary to ensure that the questions were appropriate and covered issues of importance to the nurses. The plan was to pilot a draft on five nurses on the SUHT ward using an administered questionnaire, i.e. the researcher reads out the questions and then asks for feedback on the content, wording and coverage. The revised questionnaire would have been sent out after the barcode scanning project.

It was not possible to proceed as planned. During the second week of data collection, the nurses were told that the ward relocation was to be brought forward and it was therefore necessary to shorten the research timescales. For this reason, questionnaire piloting took place during the last three days of the barcode data collection. This also meant it was not possible to use the nurses nominated by the Senior Sister. Either the nurses were on annual leave or were unavailable due to their workload. Therefore, the researcher chose five nurses across different grades from those available on duty on 21-22/11/05.

In addition, due the staffing levels and minimal overlap between shifts it was necessary to abandon the administered questionnaire. Instead, a modified self-administered questionnaire (Version 2) was developed that incorporated the full questionnaire (Version 1) with the feedback questions from the planned interview schedule. The nurses kindly agreed to a brief discussion whilst on their meal breaks.

Through an iterative process, the questionnaire was revised, re-administered and feedback obtained (written and face-to-face) from nurses. Two nurses completed Version 2 and three completed Version 3. At SUHT, nurses subsequently used Version 4 and all responses from the pilot versions (2 and 3) were transcribed onto Version 4 questionnaires before data entry. However, a fifth version shown on the following pages was necessary at Totton in order to make it applicable to the haemodialysis setting.

Measurement of patient dependency & nursing time

University of Southampton

Feedback on measurement of patient dependency & nursing time

Thank you for spending a few minutes to give some feedback.
You do not need to add your name – the questionnaire is anonymous.
Please answer all the questions. You may add extra comments if you wish.

Office use

1. Please enter your grade (e.g. A, ... G) ☐ 1.

Using the bar code scanner

2. Did you find the actual bar code scanner easy to use? *(Please circle)* **No** [1] **Yes** [2] ☐ 2.

3. How many of your <u>shifts</u> did it take you to become confident using the bar code scanner? *(Please circle)* **1 2 3 4 5 6+ Still not confident** [7] ☐ 3.

4. Did you find the bar code scanners reliable (i.e. when scanning the bar codes)? *Please tick <u>one</u>* ☐ 4.

1) Yes, all of the time	
2) Yes, most of the time	
3) No, often difficult	
4) No, difficult most of the time	

Please comment if you wish ☐ 5.

____

____

The bar code lists (of patients and activities)

5. Please indicate how much you used the bar code <u>lists</u> (in the red files) in the following places? *Please <u>tick one column for each row</u>*

	A lot [1]	Sometimes [2]	Never [3]	Office use
Ward – By door				6.
Ward – Nurses' station				7.
Ward – By centrifuge				8.
Clean utility				9.
Kitchen				10.
Office				11.

Measurement of patient dependency & nursing time

University of Southampton

Office use

6. Have you any ideas about how to improve the bar code lists (e.g. layout)? Please comment if you wish ...

12.

____

____

Effect of bar coding on your work

7. Did the bar coding intrude on your relationship with patients?

Please tick one

13.

1) No	
2) Yes, minimal but acceptable	
3) Yes, moderate but acceptable	
4) Yes, a lot but acceptable	
5) Yes, unacceptable	

Please comment if you wish

14.

____

____

8. When did you scan the bar codes ...

Please tick one column for each row

	A lot	Sometimes	'Never'	
1) At the start of the activity				15.
2) Sometime during the activity				16.
3) At the end of the activity				17.

Please comment if you wish

18.

____

____

9. Did you have enough training in the use of the bar code scanners and bar code lists?

Please tick one

19.

1) Yes	
2) Yes, but would have liked more	
3) No, needed more	
4) Had no training	

Measurement of patient dependency & nursing time

University of Southampton

Office use

Please comment if you wish

20.

21.

10. Do you think the amount of information given to you about the project was …

Please tick one

1) Too much ☐
2) Sufficient ☐
3) Adequate, but would have like more ☐
4) Not enough ☐

Please comment if you wish – suggest information you would have liked to have been given

22.

23.

11. Would you be happy to take part in another bar code scanner project?

Please tick one

1) No – Never ☐
2) Yes – Up to one week ☐
3) Yes – Up to two weeks ☐
4) Yes – Up to one month ☐
5) Yes – More than one month ☐

Please comment if you wish

24.

Measurement of patient dependency & nursing time

University of Southampton

Office use

Your view on the information collected

12. Do you think it is useful to know how much time nurses spend with different types of patient?

Please tick one or more

		Office use
1) No	☐	25.
2) Yes – To me	☐	26.
3) Yes – To my team or colleagues	☐	27.
4) Yes – To hospital managers	☐	28.
5) Yes – To others	☐	29.

Please comment if you wish ______________________ 30.

Any further comments you would like to make … ______________________ 31.

Your views on data collection by an observer

13. Did you feel that having an observer interfered with your work? 32.

Please tick one

1) No	☐
2) Yes, minimal but acceptable	☐
3) Yes, moderate but acceptable	☐
4) Yes, a lot but acceptable	☐
5) Yes, unacceptable	☐

Please comment if you wish ______________________ 33.

Measurement of patient dependency & nursing time

University of Southampton

Office use

14. Do you feel that the observer being there made you act differently from normal? — *Please tick one* — 34.

   1) No ☐
   2) Yes, minimal change ☐
   3) Yes, moderate change ☐
   4) Yes, large change ☐
   5) Other ☐

   Please comment if you wish — 35.

   ____

   ____

15. An observer collecting data is an alternative to bar code scanning. Would you be happy to take part in another project where you are observed? — *Please tick one* — 36.

   1) No – Never ☐
   2) Yes – Up to one week ☐
   3) Yes – Up to two weeks ☐
   4) Yes – Up to one month ☐
   5) Yes – More than one month ☐

   Please comment if you wish — 37.

   ____

   ____

Any further comments you would like to make — 38.

____

____

____

____

Thank you for your time

Please return this questionnaire to
Tricia Nicholson or c/o box in Staff Room

Appendix 12 Work sampling paperwork (Totton)

This page shows the barcodes used for work sampling at Totton. Numbers 1-11 correspond to the HD bay positions (with different codes for morning, afternoon and evening sessions to facilitate keeping track of patients at the changeovers between HD sessions).

Morning — Work sampling

1: Nurse, 'Agency', Nurse NO PT, Curtains
2: Nurse, 'Agency', Nurse NO PT, Curtains
4: Nurse, 'Agency', Nurse NO PT, Curtains
5: Nurse, 'Agency', Nurse NO PT, Curtains
6: Nurse, 'Agency', Nurse NO PT, Curtains
7: Nurse, 'Agency', Nurse NO PT, Curtains
8: Nurse, 'Agency', Nurse NO PT, Curtains
9: Nurse, 'Agency', Nurse NO PT, Curtains
10: Nurse, 'Agency', Nurse NO PT, Curtains
11: Nurse, 'Agency', Nurse NO PT, Curtains

Nurse + unknown patient — Nurse (no patient)

'Agency' + unknown patient — 'Agency' (no patient)

Explanation of work sampling barcodes

All nurses were included in work sampling observations. However, the plan was to use the data to validate the study nurses' recordings. Therefore, to facilitate crosschecking, extra codes were needed so that observations of non-study nurses could be excluded.

Label	Interpretation
By bay position	
Nurse	Nurse with patient (direct care)
'Agency'	Non-study nurse with patient (direct care)
Curtains	Not know if nurse present as view of patient obscured (behind curtains / door). This was a particular problem at SUHT as the ward operated with mixed-sex bays. For privacy, patients or nurses often drew the curtain between the beds.
Nurse NO PT	Study nurse at HD bay, but no patient (i.e. indirect care or general activity). Relevant to crosschecking barcode scanning dataset and so ignored for non-study nurses.
General codes	
Nurse + unknown patient	Nurse with patient of unknown identity because interaction outside HD bay (i.e. direct care but patient ID unknown).
'Agency' + unknown patient	As above, but for non-study nurse.
Nurse (no patient)	Nurse not with a patient (i.e. indirect care or general activity).
'Agency' (no patient)	As above, but for non-study nurse.

Work sampling one-hour random observations sheet (20 per hour)

Data collection sheets like the one below were generated in Excel (2003) to give the random start times of observations. The sheet was used to record information that would help assess the validity of work sampling and to facilitate use of the data to validate the nurses' recordings. For example, the 'Not in use' column was used when a HD bay was unoccupied and other information included the number of nurses and patients, and comments about completion of the observations as scheduled.

Date		Hour	

Nurses (total)	Study	Non-Study

Scanner no: [ ]

Time	2
Gap	3

Start zero: 0

Random	Sample	Start time	Gap	Done	Comment	Tot Nurse (no pt)	Nurses	Patients	Not in use
1	1	1	1						
0	2	3	2						
1	3	6	3						
1	4	9	3						
1	5	12	3						
0	6	14	2						
0	7	16	2						
1	8	19	3						
0	9	21	2						
0	10	23	2						
1	11	26	3						
1	12	29	3						
0	13	31	2						
2	14	35	4						
2	15	39	4						
2	16	43	4						
0	17	45	2						
2	18	49	4						
1	19	52	3						
2	20	56	4						
		59	3						

Min	1
Max	4

Missed scans [ ]

Changed scanner [ ]

Appendix 13 Patient dependency-scoring tool for outpatient HD (final version)

....	....	....	....
Last name	**First name**	**Date**	**Completed by (PRINT NAME)**

- Circle each box (1 to 3) that applies (0 = 'normal') i.e. if necessary multiple items within a category
- Circle actual problems, not potential problems. Include the whole session, not just when the patient is on dialysis. If necessary, note other aspect(s) relevant to patient dependency not covered here.
- **To avoid missing items, please note things as you go (as someone else may take the patient off HD).**

	Score	**0**	**1**	**2**	**3**
	Non-attendance		Patient does not attend because in-patient		Score 5 if patient 'refuses' to attend for HD
HD assessment pre-dialysis	**Ease of access (Fistula)**	2 needles good flow	Single needle	Conversion to single needle	1st needling
			Require manipulation	Re-site	Difficulty needling
	Ease of access (Neckline)	Patent good flow	Lines reversed	Poor flow or high pressures	New access required
				Swabs and cultures	Anti-thrombolytic agents required
	Pre-HD BP	Systolic 110-149 & Diastolic <90	Systolic 81-109	Systolic 71-80	Systolic <70
			Systolic 150-199	Systolic >200 or Diastolic >= 90	
	Deviation from target weight (pre-HD)			Under by 1kg	Under by 2kg
			Over 2-3kg	Over 3-4kg	Over 4kg+
	Temperature (pre & post HD)			<35.5	37+
Risk assessment	**Mobility**	Fully mobile	1 person assistance	2 persons assistance	Lifting equipment needed
			Walks with stick	Wheel-chair	Bed required or air bed
	Fluid Status				Dehydrated
			Oedema	Breathless due to fluid	Pulmonary oedema
	Elimination needs	Independent with toilet needs		Requires 1 nurse assistance or vomiting	Incontinent
					2 nurse recirculation
	Communication	Fully conversant		Speech or hearing impaired	Confused or agitated or distressed
				Blind or partially sighted (i.e. needs assistance)	Depressed or upset (i.e. >15 mins 'counselling')
				Language difficulty	New patient
During HD session	**Diabetes**	No	Stable	Frequent blood sugar monitoring	Unstable requiring intervention
	Nutritional requirements		Oral supplements (e.g. Fortisip)	Assistance with feeding required	IV Nutrition required
			Refreshments (i.e. 'tea and toast round')		
	Therapies / interventions		EPO	Vaccinations	HDF or Heparin free HD or Isolated UF
			IV alfacalciferol	IV Iron UKM	IV Antibiotics
			Analgesia (e.g. paracetamol)	Monthly or other blood tests	Transplant bloods (~10+)
					Blood transfusion
	BP during or after dialysis	Systolic 110-149 & Diastolic <90	Systolic 81-109	Systolic 71-80	Systolic <70
			Systolic 150-199	Systolic >200 or Diastolic >= 90	
	CVS stability	Stable	1 episode requiring nursing intervention (e.g. low BP, cramps)	Multiple episodes requiring nursing intervention	Arrhythmias
	Other		Contact MDT (e.g. Dietician, Social Worker, Specialist Nurse, etc)		Medical treatment required (i.e. doctor's advice or visit)
	Dressings		'Quick' dressing(s) i.e. taking up to 5 mins	'Moderate' dressing(s) i.e. taking 6-10 mins	'Long' dressing(s) i.e. taking > 11 mins
	Machine problems	No Problems		Frequent alarms	Machine changed
				Report fault	Clotted circuit
	Taking off	Stop bleeding within 10 mins	Slightly longer to stop bleeding (i.e. 11-20 mins)	Quite a bit longer to stop bleeding (i.e. 21-30 mins)	Much longer to stop bleeding (i.e. >30 mins)
	No: items circled	**Column subtotal**	1 x =	2 x =	3 x =
				Total: enter here *& enter on Proton*	

Appendix 14 Karnofsky Performance Scale

----------------------------	----------------------------	-------------	----------------------------
Last name	First name	Date	Completed by (PRINT NAME)

Karnofsky Performance Scale

Instructions - Please circle the most appropriate score for this patient

Score %	Functional status
100	The patient has no complaints and is without evidence of disease
90	The patient has minor signs/symptoms, but is able to carry out his or her normal activities
80	The patient demonstrates some signs/symptoms and requires some effort to carry out normal activities
70	The patient is able to care for self, but is unable to do his or her normal activities or active work
60	The patient is able to care for self, but requires occasional assistance
50	The patient requires medical care and much assistance with self care
40	The patient is disabled and requires special care and assistance
30	The patient is severely disabled and hospitalisation is indicated; Death is not imminent
20	The patient is very ill with hospitalisation and active life-support treatment necessary
10	The patient is moribund with fatal process proceeding rapidly
0	Dead

Please leave this form with the HD sheet

Please do not write below this line

Karnofsky Performance Scale: | | | |

Co-morbidity indices

Wright / Khan index			
Age		**Other factors**	**Risk group**
<70	And	No co-morbid illness	1 Low
70-80	Or	Age < 70 with one of: Angina Previous MI Cardiac failure CVA COAD Pulmonary Fibrosis Liver diseases (cirrhosis or chronic hepatitis)	2 Medium
	Or	Age < 70 with Diabetes Mellitus	
>80	Or	Any age with 2+ organ dysfunctions and ESRF	3 High
	Or	Any age with Diabetes Mellitus and cardiac / pulmonary disease	
	Or	Any age with visceral malignancy	

Charlson Co-morbidity Index		Weight (score for each condition)
MI Congestive cardiac failure Peripheral vascular disease Cerebrovascular disease Chronic pulmonary disease Connective tissue disease Peptic ulcer disease Mild liver disease Diabetes		1
Hemiplegia Moderate or severe renal disease Diabetes with end organ damage Any tumour (no metastasis within past 5 years) Leukaemia (inc acute and chronic polycythemia vera) Lymphoma (inc Hodgkins, Waldenstroms, Myeloma, Lymphosarcoma)		2
Moderate or severe liver disease		3
Metastatic solid tumour AIDS		6
Plus age weighting	≤ 49 years	+0
	50-59	+1
	60-69	+2
	70-79	+3
	>80	+4

Appendix 16 Results of nurses' acceptability questionnaire (SUHT and Totton)

At both SUHT and Totton, 15 nurses at each site completed the questionnaire. This excluded three 'study' nurses (see Figure 7.1); one at SUHT was away on holiday, one at Totton was off sick, and one at Totton withdrew before administration of the questionnaire.

Using the barcode scanner

Did you find the actual barcode scanner easy to use?	No	Yes	Missing
SUHT	0	15 (100%)	
Totton	0	13 (87%)	2 (13%)

Did you find the barcode scanners reliable (i.e. when scanning the barcodes)?	Yes, all of the time	Yes, most of the time	No, often difficult	No, difficult most of the time
SUHT	5 (33%)	10 (66%)	0	0
Totton	9 (60%)	6 (40%)	0	0

How many of your shifts did it take you to become confident using the barcode scanner?	1	2	3	4	5	6+
SUHT	6 (40%)	6 (40%)	2 (13%)	1 (7%)	0	0
Totton	4 (27%)	9 (60%)	1 (7%)	1 (7%)	0	0

Barcode lists (of patients and activities)

Please indicate how much you used the following types of barcode lists?	A lot	Sometimes	Never	Missing
SUHT				
Personal sheet of listed patients	9 (60%)	6 (40%)	0	
Stickers by patients	12 (80%)	3 (20%)	0	
Lists in bays	4 (27%)	9 (60%)	1 (7%)	1 (7%)
Drug trolley	5 (33%)	3 (20%)	7 (47%)	
Notes trolley (i.e. medical notes)	0	5 (33%)	10 (67%)	
Totton				
Unit by door	8 (53%)	7 (47%)		
Nurses' station	7 (47%)	7 (47%)	1 (7%)	
By centrifuge	1 (7%)	4 (27%)	9 (60%)	1 (7%)
Clean utility room	1 (7%)	4 (27%)	10 (67%)	
Kitchen	1 (7%)	9 (60%)	5 (33%)	

Have you any ideas about how to improve the barcode lists (e.g. layout)? Please comment ...	Site	Comments
	SUHT	Suggestions for individual barcode lists to be: • on smaller paper or one sheet (two nurses) • 'stronger' i.e. on card or in a plastic pocket/laminated (three nurses) Five nurses stated they were happy with the layouts.
	Totton	"Bit more space between them" Request for barcodes in alphabetical order - however these were already available. Three nurses said 'no'.

I.e. Barcode lists located in most places were used to some degree (the least used were the lists on the notes trolleys in SUHT study and by centrifuge and in clean utility room in Totton study).

Effect of barcoding on your work

Did the barcoding intrude on your relationship with patients?	No	Yes, minimal but acceptable	Yes, moderate but acceptable	Yes, a lot but acceptable	Yes, unacceptable
SUHT	6 (40%)	7 (47%)	2 (13%)	0	0
Totton	8 (53%)	6 (40%)	0	1 (7%)	0

Comment	Site	
	SUHT	Patients sometimes reminded nurses to scan
	Totton	"Became a joke - we couldn't speak to them unless they'd been scanned" "Majority of patients happy with what we've been doing"

When did you scan the barcodes	Usually	Sometimes	Never	Missing*
1) At the start of the activity				
SUHT	12 (80%)	1 (7%)	0	2 (13%)
Totton	13 (87%)	2 (13%)	0	0
2) Sometime during the activity				
SUHT	2 (13%)	11 (73%)	0	2 (13%)
Totton	2 (13%)	11 (73%)	0	2 (13%)
3) At the end of the activity				
SUHT	4 (27%)	5 (33%)	2 (13%)	4 (27%)
Totton	6 (40%)	7 (47%)	2 (13%)	0
4) Other - please specify below				
SUHT	0	1 (7%)	1 (7%)	0
Comments SUHT	"Never scanned for the activity after I'd completed it - if I forgot to scan I would fill out a mistakes form" "Sometimes the patient reminded me!" Noted "Rarely" for 'at end of activity'			
Totton	Two nurses stated that they sometimes or often forgot to scan the barcodes			

* Missing some responses because question response format different in earlier pilot versions

Did you have enough training in the use of the barcode scanners and barcode lists?	Yes	Yes, but would have liked more	No, Needed more	Had no training
SUHT	14 (93%)	1 (7%)	0	0
Totton	15 (100%)	0	0	0

Do you think the amount of information given to you about the project was …	Too Much	Sufficient	Adequate, but would have liked more	Not enough
SUHT	0	14 (93%)	1 (7%)*	0
Totton	0	15 (100%)	0	0

*No suggestion about what additional information to include

Would you be happy to take part in another barcode scanner project?	No - Never	Yes - Up to one week	Yes - Up to two weeks	Yes - Up to one month	Yes - More than one month
SUHT	1 (7%)	11 (73%)	2 (13%)	1 (7%)	0
Totton	0	6 (40%)	6 (40%)	3 (20%)	
Comments (Totton)	"As this would allow you to feel that adequate information was being gathered, although I'm sure it was" "It does take a bit of time to do and its something you always have to remember to do but for a short period of time it is" "Easy enough to do once you get in the habit" "2 weeks is long enough at one go. Would do another 2 weeks after a break" "Short term is fine - hard to cope with over longer period"				

Your view on the information collected

Do you think it is useful to know how much time nurses spend with different types of patient?	SUHT	Totton
1) No	0	0
2) Yes - To me	7 (47%)	7 (47%)
3) Yes - To my team or colleagues	11 (73%)	10 (67%)
4) Yes - To hospital managers	11 (73%)	13 (87%)
5) Yes - To others	5 (33%)	6 (40%)

Comments (Totton)
"As long as the data collected is not derogatory in minimising staff levels"
"We know which patients need more care because they take longer to stop bleeding, and as a team we allow for that but I think its useful for managers/Sister to use as justification for staffing levels etc "

Your views on data collection by an observer

Did you feel that having an observer interfered with your work?	No	Yes, minimal but acceptable	Yes, moderate but acceptable	Yes, a lot but acceptable	Yes, unacceptable
SUHT	9 (60%)	5 (33%)	1 (7%)	0	0
Totton	12 (80%)	1 (7%)	0	1 (7%)	0
Comments (Totton)	"Knowing the observer was there made me think twice about scanning"				

Do you feel that the observer being there made you act differently from normal?		No	Yes, minimal change	Yes, moderate change	Yes, large change	Other	Missing
SUHT		12 (80%)	2 (13%)	0	0	0	1 (7%)
Totton		11 (73%)	3 (20%)	0	0	0	1 (7%)
Comments	SUHT	Reminded 'me' to use scanner more (two nurses) "It's quite strange initially having observer especially at night".					
	Totton	"Knowing the observer was there made me think twice about scanning"					

An observer collecting data is an alternative to barcode scanning. Would you be happy to take part in another project where you are observed?	No - Never	Yes - Up to one week	Yes - Up to two weeks	Yes - Up to one month	Yes - More than one month	Missing
SUHT	2 (13%) RGNs	8 (53%)	1 (7%)	2 (13%)	1 (7%)	1 (7%) HCA
Totton	1 (7%) HCA	5 (33%)	6 (40%)	0	0	0
Comment (Totton)	"Would want a break before another 2 week block"					

Additional comments from pilot interviews:

- Issues about remembering when multitasking - multiple patients asking for things.
- Tiredness factor as shift progresses (during 12 hour shifts scanning becomes poorer).
- Sometimes forget whether scanned barcode so scan again.
- Discussed lack of use of comments sheet. Agreed with TN's suggestion of tick boxes e.g. missed meal break, shift overrun, emergencies (when in shift).

Other comments (SUHT)

- "Working long days made the results inaccurate".
- "Sometimes forgot for periods what had scanned/not (e.g. hour or so)
- "Think there needs to be plenty of time so you get into the routine".
- "I did get to enjoy the barcoding, it was okay but I thought it became harder when any colleagues asked you to do something else then I forgot to barcode most of the time trying to do emergency tasks".
- "I think for the time we had and the ward was very busy, a lot of scans were missed".
- "Scanning easier at beginning of shift but when workload escalates due to staff shortages/poorly patient scanning becomes erratic as priority lies with patients & workload + we're not programmed to automatically scan so concerned dependencies will reflect this."

Other comments (Totton)

- "This has been an interesting 2 weeks and have enjoyed having the observer around"
- "Feel that it could have a negative effect - on shifts where dependency is low management may try to cut staffing levels further!"
- "I really enjoyed using the bar coder I will miss it!"
- "I hope this data collections will help each staff how much time they spending for patients and for general activity, and how they can organise things properly".

Appendix 17 Feasibility of work sampling (SUHT and Totton)

Piloting revealed a number of concerns about the suitability of work sampling for patient-level data collection. In the initial implementation at SUHT, it was difficult to track the nurses. At shift changes, many different study and non-study staff were present. Storerooms and the staff room were located off the ward and so the researcher assumed that no direct care occurred off the ward. There were mixed sex bays and often curtains were drawn around beds for privacy, which obscured the view of nurses/patients. This led to inaccuracies in coding of direct care, although it was sometimes possible to hear whether a nurse was present. Ward records were not always up-to-date. Hence, the duty rota did not always show if a nurse failed to attend or was moved to another ward; and help received for a few hours was not recorded. This hampered calculation of overall nursing time. Similarly, the boards displaying patient names and locations were not always up-to-date. This made it difficult to ensure that observations were attributed to the correct patient. In response, the researcher tried to liaise with a senior nurse to check the staffing and patients before starting each work sampling observation block. Despite concerns, piloting of work sampling continued because compared with the SUHT ward, the Totton HD unit was small, screens were rarely used and the patient group was more stable.

Work sampling at Totton was unsuccessful. The unit had wide pillars and a cramped layout that obstructed the observer's view and made it easy to make mistakes (missing nurses or attributing nurses to the wrong HD bay and hence patient). To avoid obstructing the nurses, the observer used 'fixed' viewing positions rather walking through the unit, but this missed nurse-patient interactions in the waiting room. To continue work sampling beyond the pilot would have required extra resources. It was difficult for the observer to maintain concentration for prolonged periods (with only 10-minute breaks) and whilst trying to support the nurses' barcode scanning. The researcher was aware that 8% of observations were 'incorrect' as shown below.

Work sampling - known observation 'mistakes'

Reason	Occasions	Importance
Missed	51 (3%)	Reduced number of observations and affects ability to detect differences
Nurse/patient interruption	34 (2%)	
Other reason	17 (1%)	
Early/Late	97 (5%)	Affects randomness of observations - acceptable if not missed systematically
Total	148 (8% of all work sampling observations)	

After data collection, the researcher realised that it would be difficult to convert the work sampling data into direct care times per HD session. To maximise observation time, the schedule had covered half the unit's opening hours each day ('mornings' one week and 'evenings' the second week), but this cut-across the afternoon HD sessions. It was difficult to decide how to work out the 'patient-equivalent' sessions because the proportion of each patient's session observed was unknown. (Patients did not attend at exactly the same time each visit and it was difficult to keep track of patients' arrival and departure times, which otherwise could only be guessed from computer records of time on/off HD.) Therefore, adding together the time data for the two part-sessions (across the two weeks) was considered unreliable. If the afternoon sessions had been excluded, it was unclear how to estimate the comparable nursing hours available to patients. In retrospect, data collection should have been continuous, but this was not possible for one researcher.

In conclusion, it was not possible to reliably link work sampling data to specific patients and the technique was dropped from the main data collection.

Appendix 18 Dependency scoring issues (Totton)

The table below shows a number of issues that arose during the Totton study about the completion of the dependency-scoring tool. It was possible to address some issues for the Portsmouth data collection. Actions taken were: increased nurse training, inclusion of instructions on the dependency form (as previously there were none), and changing the layout to try to improve clarity and hence arithmetic. Other broader issues remained unresolved despite discussion with senior nurses from all the HD units.

Issue	Why important
Issues addressed for the Portsmouth study	
Inconsistencies in scoring. Some nurses scored every level that applied whilst others scored the worst level for each item. This problem particularly applied to access, mobility and therapies.	The maximum possible score was 42 using the worst case on each level, but 109 using multiple responses (although these scores were not necessarily clinically feasible). Scoring inconsistencies resulted in a difference for 50 out of 306 ratings (16%). Most only differed by one point (15%), but one was a 5-point difference.
Additional points added for items not included in rating.	Inconsistency in scoring by arbitrary inclusion of additional items that may not be relevant.
Only scoring when actually on HD rather than whole session when patient present.	Did not reflect changes if patient became ill (e.g. hypotensive) after coming off HD.
Mistakes in calculations of overall dependency scores.	Discrepancies between study raw data and total for routine unit data entered on computer. Data missing on computer on 4 occasions (1%); 6% additions under calculated (mostly by 1-2 points, but up to 6 points) and 3% over calculated.
Unresolved broader issues	
Scaling levels assigned to dependency rating items.	May not reflect differences in relative workload or time differences for these factors.
Discriminatory power of some dependency items (e.g. nutrition scored 1 on 97% occasions whilst other items had minimal use, e.g. elimination scored once).	Impact on ability to differentiate patient workload.
Analysis issues due to number of items and multiple levels on each item.	Large number of items and levels in relation to number of patients precluded many statistical analyses to investigate dependency tool. Ideally, most would require collapsing into dummy variables (Yes/No).

Appendix 19 'Spot checks' of barcode scanning (Portsmouth)

The data collection sheet below was used for barcode scanning validation observations at Portsmouth.

Date / /06

Measurement of patient dependency & nursing time

University of Southampton

Bay	Time	Nurses	Activity	Sid	✓✗
Waiting			DC IC GA		
N. Station			DC IC GA		
Other **NO PT**			DC IC GA		
Other **PT**			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
			DC IC GA		
N. Station			DC IC GA		
Other **NO PT**			DC IC GA		
Sluice			DC IC GA		
Clean utility			DC IC GA		
Kitchen			DC IC GA		
Fluid store			DC IC GA		
TOTAL					
Total exp					

Notes:
DC = Direct care, IC = Indirect care, GA = General activity, Sid = Staff ID

Appendix 20 Patient dependency-scoring tool - verification of data

The table below examines the items in the dependency-scoring tool to see how each might be assessed for inter-rater reliability. It shows when the items are assessed, how easy it would be verify them and whether verification is likely to be objective.

Inter-rater reliability - ease of verification of dependency data

When assessed	Easily verified	Nature of assessment	Aspects
Start of HD	Yes	Objective	• Deviation of weight from target • Temperature
Start of HD	No	Subjective	• Vascular access (2 sets of items)
Ongoing*	Yes	Objective	• Blood pressure (2 sets of items)
Ongoing	(Yes)	Relatively objective	Relatively stable patient characteristics: • Mobility items • Three communication items (e.g. sight or hearing impaired)
Ongoing	(Yes)	Relatively objective	• Diabetes - blood glucose • Nutritional requirements • Therapies/interventions (e.g. drugs administered, bloods taken) • Other (referral or medical treatment required) • Machine problems
Ongoing	No	Subjective	• Fluid status • Elimination needs • Three communication items • Cardiovascular stability (i.e. episode requiring nursing intervention) • Dressings
End of HD	No	Relatively subjective	• Time to stop bleeding

Key / notes:
* Ongoing means assessed throughout the HD session
Non-attendance not included above

Appendix 21 Estimation of nursing costs - data sources for resource use and unit costs, and assumptions used

Table 1 shows the source of cost data required to estimate the top-down or bottom-up unit costs per hour of nursing time and nursing costs per HD session.

Table 1 Sources of cost data

Aspect	Source (costs in 2006)
Salaries	National scales from NHS Employers (2006): Used midpoint and highest salaries.
Salary oncosts	
National Insurance (NI)	HM Revenue and Customs (2005): **For annual pay:** Salary <£5,035 NI = -(5,035 - 4,368) x 3.5% = -£23.35 (rebate). Salary £5,035 to £33,540 NI = (Salary - 5,035) x 9.3% - £23.35. Salary >£33,540 NI = (33,540 - 5,035) x 9.3%+ (Salary - 33,540) x 12.8% - £23.35. **For monthly pay:** Salary <£420 NI = -(420 - 364) x 3.5% = -£1.96 (rebate). Salary £ 420 to £2795 NI = (Salary - 420) x 3.5% - £1.96. Salary >£2795 NI = (2795 - 420) x 9.3% + (Salary - 2795) x 12.8% - £1.96. Note: Due to the NI bands, the order in which NI is applied matters (i.e. to components or total costs. After application of NI the sum of components ≠ total cost).
Superannuation	Department of Health (2007d): Paid at 14% applicable to basic pay and enhancements for unsocial hours, but not overtime (if paid).
Payments for unsocial hours	NHS Whitley Council (2004): Extra payments: Basic rate x 0.3 for all Saturday and twilight shifts, and basic pay x 0.6 for all Sunday shifts.
Hours paid per nurse per year	Hours paid per year: Weeks per year x hours worked per week per WTE = (365 / 7) x 37.5 = 1955 hours.

Table 2 shows the steps to calculate the top-down nursing costs per HD session. In addition, for comparison with the bottom-up estimates two other outputs calculated were the average cost per hour paid and average cost per hour worked.

Table 2 Steps in calculation of average top-down costs in study

Step	Calculation required (costs in 2006)
Preliminary steps	
Basic rate cost per hour	At each grade: Basic rate cost per hour = Annual salary (midpoint or highest)[1] divided by hours paid per year (1955 hours from Table 1)[2]
Basic cost components calculated using i) midpoint, ii) highest salary	Estimated pay expenditure using cost components for each individual nurse from duty rotas (regardless of whether study nurse or not): A. Basic pay = Total hours (productive and unproductive time across the 28 days) x basic rate cost per hour (see above) B. Pay for unsocial hours = [Saturday or twilight hours x basic rate cost per hour x 30%] + [Sunday hours x basic rate cost per hour x 60%] C. Total nurse's pay = A+ B D. Employer's superannuation contribution = 14% x C E. Employer's NI contribution = Rates (monthly rates from Table 1) applied to C F. Pay expenditure (incl. unsocial hours and oncosts) = C + D + E
Total estimated nursing pay expenditure	= Sum of nurses pay expenditure (incl. unsocial hours and oncosts)[3] = £73,122 (midpoint salaries) or £83,347 (highest salaries)
Number of HD sessions	1641 HD sessions delivered from study (see section 9.1).
Nursing cost per HD session	= Total estimated nursing pay expenditure divided by HD sessions delivered in study = £73,122 / 1641 = £44.56 per hour (midpoint salaries) or = £83,347 / 1641 = £50.79 per hour (highest salaries)
Other outputs	
Total nursing hours paid during study	From duty rotas (all nurses regardless of whether in study): 5111 hours paid (total productive and unproductive hours) 3900 hours worked
Average cost per hour paid	= Total estimated nursing pay expenditure divided by hours paid in study = £73,122 / 5111 = £14.31 per hour (midpoint salaries) or = £83,347 / 5111 = £16.31 per hour (highest salaries)
Average cost per hour worked	= Total estimated nursing pay expenditure divided by hours worked in study = £73,122 / 3900 = £18.75 per hour (midpoint salaries) or = £83,347 / 3900 = £21.37 per hour (highest salaries)

Notes:

[1] A 'true' top-down cost uses actual not estimated pay expenditure for each nurse's individual salary point (not midpoint or highest point as here). Actual pay expenditure includes all payments for unsocial hours and salary oncosts (as here), plus payment for overtime or agency nurses and any relevant enhancements for specialities (e.g. psychiatric or geriatric nursing) and locality payments (e.g. London or other location allowance to attract staff.

[2] Most actual pay expenditure is calculated on a monthly basis i.e. (Annual salary divided by 12) x proportion of WTE. However, a basic rate cost per hour (as here) is used to calculate enhancements for unsocial hours (or hourly paid staff).

[3] A few shifts were worked as overtime but not identified on the duty rota and so were treated as usual working hours in estimating pay expenditure.

Table 3 shows the resource use data from study required for bottom-up costing.

Table 3 Resource use data from study required for bottom-up costing

Resource use	Source or calculation required (costs in 2006)
Number of HD sessions	1641 HD sessions delivered from study (see section 9.1).
Percentage of patient-specific nursing time	*From study overall 54% (from Table 9.4), RGN 55%, HCA 50%*
Hours paid per nurse per year	1955 (from Table 1)
Average unproductive time per nurse	During study = 161 days / 34.1 WTE nurses = 4.7 unproductive days in 28 days Extrapolated to one year = 4.7 x 365/28 = 61 unproductive days per year or = 61 days x 7.5 hours per day = 461 unproductive hours per year
Hours worked	Hours worked per year: Expected = 1560 hours (from Curtis 2007). Actual = Hours paid per year - Unproductive hours per year = 1955.4 - 460.5 = 1495 hours (rounded)
Average whole time equivalents (WTE)	WTE (over 4-weeks of study) = Hours paid at each grade / Total hours paid / WTE hours per week x Weeks in study = Hours paid at each grade / 5111 / 37.5 x 4. Average WTE overall = Sum of WTE at each grade = 34.1. WTE for Band 5 RGNs = 21.2 Percentage of hours for Band 5 RGNs = 62%
Resource use: Nursing hours	Patient-specific time from study: 1. Overall mean 73 mins per patient per HD session (from Table 9.17) or 1.22 hours. 2. Extra patient-specific time per HD session mean 8 mins (95% CI 4 to 11 mins) for patient ineligible for RSU care (from Table 9.15, model with eligibility for RSU care) or 0.133 hours (95% CI 0.067 to 0.183 hours). *Percentage of hours paid at unsocial hours rates* (worked out separately for HCAs and RGNs): i) Hours worked on Saturday or twilight shifts divided by total hours (i.e. productive and unproductive) ii) Hours worked on Sunday shifts divided by total hours (overall percentages shown in Table 9.23). For RGNs: 21% for Saturdays and twilight duty rates (at 30% extra), 12% for Sundays (at 60% extra). A few shifts were worked as overtime but not identified on the duty rota and so were treated as usual working hours.

Table 4 shows the steps to calculate the bottom-up nursing costs per HD session and the extra cost per patient ineligible for RSU care after the preliminary steps needed to calculate the average cost per hour paid, per hour worked and per patient-specific hour.

Table 4 Steps in calculation of average bottom-up costs in study - initially illustrated for Band 5 RGN

Step	Calculation required (costs in 2006)
Preliminary steps	
Basic annual salary components using midpoint salary (and also using highest salary)	**Here shown for Band 5 RGN midpoint salary** Basic annual salary = £21,646 Annual salary at Saturday or twilight rate = £21,646 x 30% = £6,494 Annual salary at Sunday rate = £21,646 x 60% = £12,988
Annual salary cost including unsocial hours in proportion to those worked in study	= £21,646 + (£6,494 x 21%) + (£12,988 x 12%) = £24,612
Annual salary cost including oncosts	Superannuation and NI rates applied to £24,612 = £29,854
Unit costs per hour (incl. unsocial hours and oncosts)	per hour paid = £29,854 / 1955 = £15.27 per hour worked (expected) = £29,854 / 1560 = £19.14 per hour worked (actual) = £29,854 / 1495 = £19.97 per patient-specific hour (expected) = £19.14 / 55% = £34.55 per patient-specific hour (actual) = £19.97 / 55% = £36.06
Contribution of Band 5 RGN nurses to average unit costs per hour	= % of total hours paid for Band 5 RGN nurses during study x unit cost per hour: = 62% x £19.14 = £11.87 per hour worked (expected) = 62% x £19.97 = £12.38 per hour worked (actual) = 62% x £34.55 = £21.42 patient-specific hour (expected) = 62% x £36.06 = £22.36 patient-specific hour (actual)
Average nursing unit cost per hour (midpoint salaries)	Sum of contribution (as above) for each nursing grade: £14.31 per hour paid £17.94 per hour worked (expected) £18.72 per hour worked (actual) £33.12 per patient-specific hour (expected working hours) £34.56 per patient-specific hour (actual working hours)
Cost per HD session	= Cost per patient-specific hour x patient-specific time per session (in hours) = £33.12 x 1.22 = £40.27 (expected working hours) or = £34.56 x 1.22 = £42.03 (actual working hours)
Extra cost per HD session per patient ineligible for RSU care	= Extra nursing time per patient x average unit cost per hour of patient-specific time = 0.133 x £33.12 = £4.42 (expected working hours) or 0.133 x £34.56 = £4.61 (actual working hours)
Extra cost per year per patient ineligible for RSU care	= Extra cost per HD session per patient ineligible for RSU care x 156 sessions per year = £689 (expected working hours) = £719 (actual working hours)

Table 5 summarises the cost outputs calculated using both the top-down and bottom-up approaches.

Table 5 Summary of unit costs per hour and per HD session from top-down and bottom-up estimation

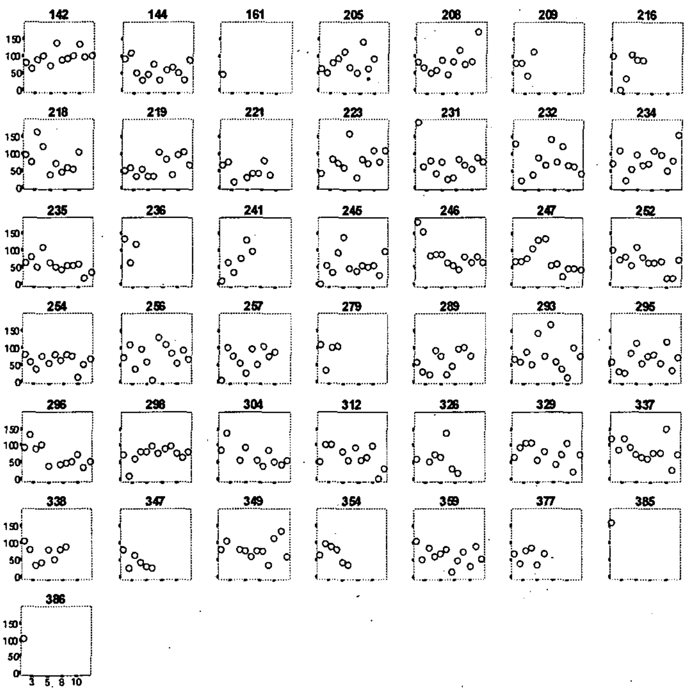
Average unit cost	Method	Midpoint salary	Highest salary	Dif from top-down
Per hour paid	Top-down	£14.31	£16.31	-
	Bottom-up	£14.31	£16.32	< 0.1%
Per hour worked	Top-down	£18.75	£21.37	-
	Bottom-up, expected hours	£17.94	£20.45	4%
	Bottom-up, actual hours	£18.72	£21.34	< 0.2%
Per hour of patient-specific time	Top-down	n/a	n/a	-
	Bottom-up, expected hours	£33.12	£37.72	-
	Bottom-up, actual hours	£34.56	£39.36	-
Cost per HD session	Top-down	£44.56	£50.79	-
	Bottom-up, expected hours	£40.27	£45.87	10%
	Bottom-up, actual hours	£42.03	£47.87	6%
	Bottom-up, actual hours, revised for missing patient-specific time*	£44.39	£50.56	<0.5%

Notes: Cost year 2006
n/a not applicable as would require weighting by nurse-level (bottom-up) data.
* Revised time estimate from Table 9.23. .

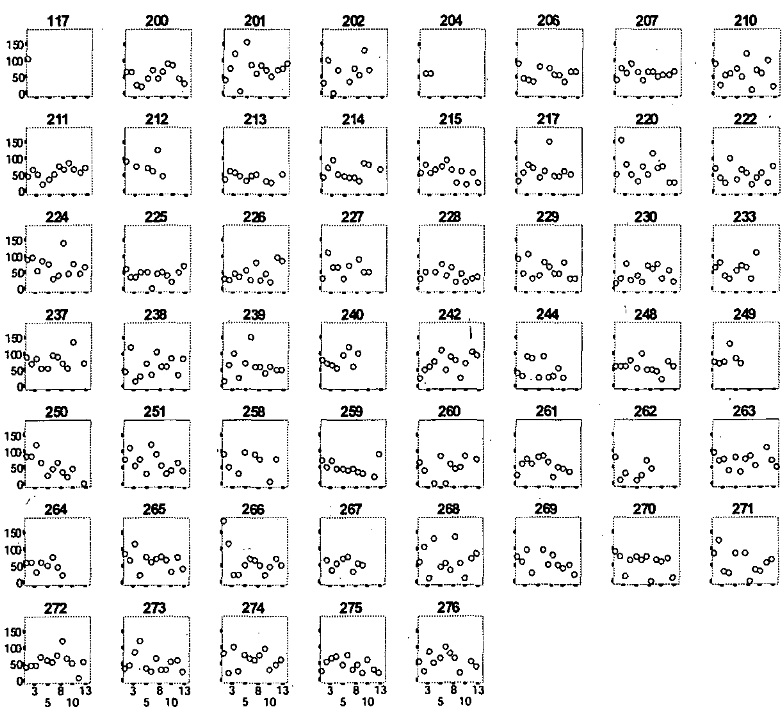
Appendix 22 Scatter plots of total patient-specific time by visit (Portsmouth)

Each 'box' represents a patient's data over the duration of the study (patient's ID above the box). The y-axis is the overall patient-specific time per session and the x-axis represents successive HD visits. The plots show much variation within and between patients. There was no obvious pattern of patients having consistently high or low patient-specific times per session.

Patients ineligible for RSU care (n = 43, one other patient had no time data))



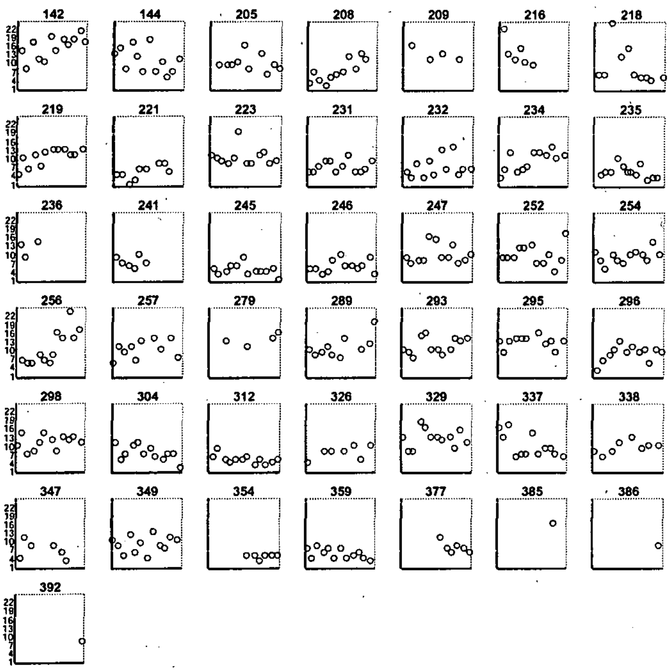
Patients eligible for RSU care (n = 108, first 54 shown here)



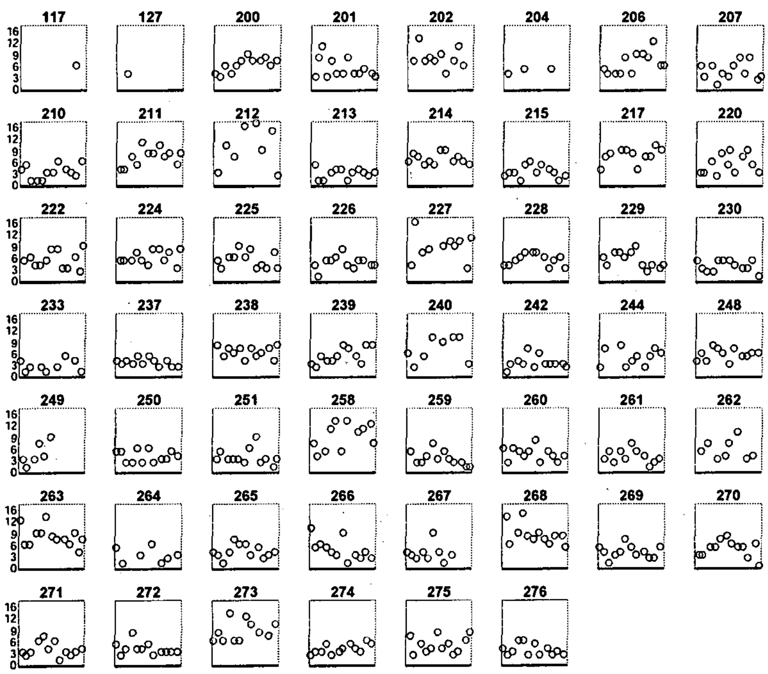
Appendix 23 Scatter plots of dependency scores by visit (Portsmouth)

Each 'box' represents a patient's data over the duration of the study (patient's ID above the box). The y-axis is the dependency score and the x-axis represents successive HD visits. The plots show much variation within and between patients and no patients with consistently high dependency scores.

Patients not eligible for RSU care (n= 43, *one other patient had no dependency* data, max dependency score 25)



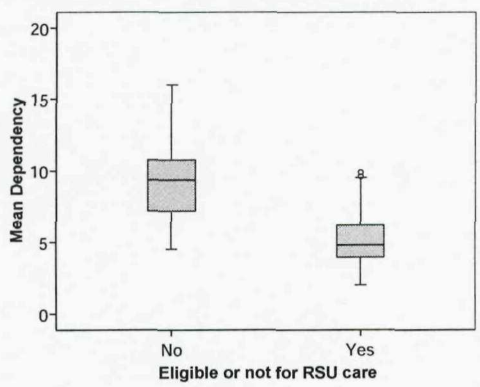
Patients eligible for RSU care (n = 108, first 54 shown here, max score 17)

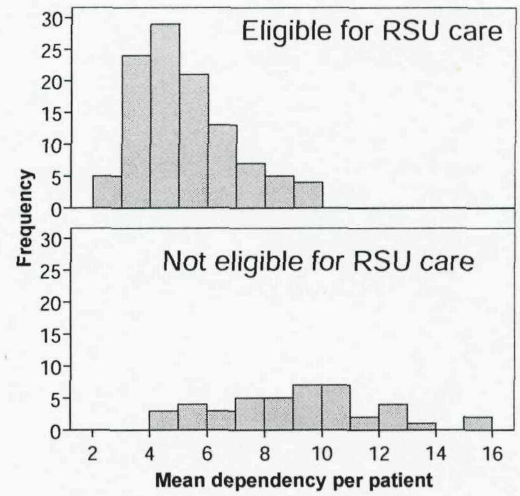


Appendix 24 Plots of patient dependency scores (Portsmouth)

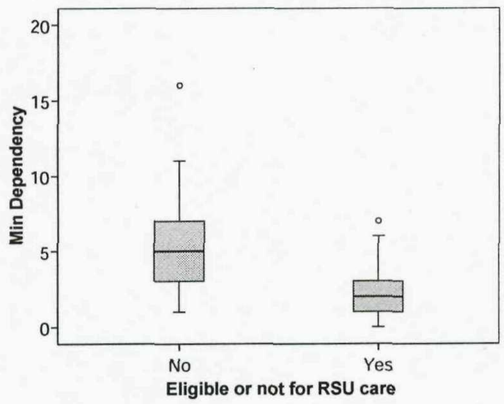
The graphs below show patients' summary dependency scores (i.e. using each patient's mean, minimum and maximum score). The boxes represent the median and inter-quartile range (IQR); circles represent outlier cases that are 1.5 to 3 times the IQR (box lengths) from the upper or lower edge of the box.

Patients' mean dependency score

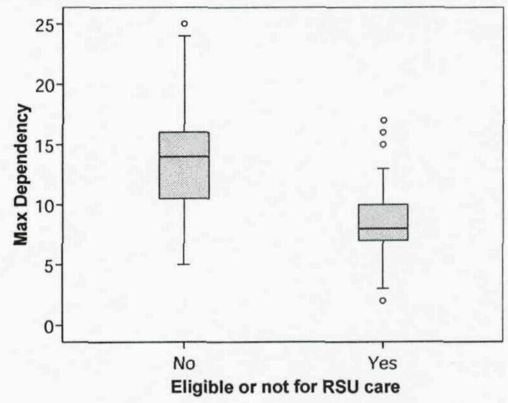




Patients' minimum dependency score



Patients' maximum dependency score



Appendix 25 Audit of dependency data quality (Portsmouth)

The table below compares the data recorded on 163 dependency forms (over 13 days from 1-14/11/06) with the HD chart or dependency score on the computer (Proton). The last column assesses the impact on the dependency scores.

Results of audit of dependency data quality

Item	Quality of data recorded on dependency forms (compared with HD notes chart)	Impact on dependency
Dependency score on computer c.f. dependency form	Scores entered on computer mostly agreed (88%) with those on or calculated from dependency forms. The remainder were higher or lower by up to 3 points.	None overall*
Target weight & pre-HD weight used to calculate weight deviation from target	Target weight and / or pre-HD weight were missing from 5 (3%) HD charts (and dependency forms). Inaccuracies in 36 (22%) dependency forms as follows, weight deviation: • Recorded on dependency form, but not present on 19 (12%) HD charts • More or less on dependency form than recorded on 11 (7%) HD charts • Not recorded on dependency form, but present on 5 (3%) HD charts • Recorded on dependency form, but raw data absent 1 HD chart.	Mixed, but overall over reported
Temperature pre-HD	Temperature >37^{0}C not recorded, but present on 3 (2%) HD charts.	Under reported
Blood glucose measurements recorded	• Not recorded for 9 (6%) patients known to be diabetic (different coloured nursing notes folder), but on 5 charts (3%) noted as stable. • Mismatch between noted stability of diabetes and frequency of monitoring: • For 20 forms, the initial blood glucose was raised (using threshold > 8 mMol/L), but only 9 charts showed glucose re-checked. • Recorded as frequent monitoring, but only recorded once in one chart.	Under reported
EPO administered	3 dependency forms noted EPO administered, but *no record of this on HD chart.*	Over reported
Iron administered	1 dependency form noted IV iron administered, but no record of this on HD chart.	Over reported
Blood pressure pre-HD	• Incorrectly reported - under valued on 10 (6%) dependency forms • Not reported when present on 16 (10%) charts • Reported but raw data missing from 1 HD chart.	Under reported
Blood pressure recorded during and / or after HD	Due to multiple readings, blood pressure could be high or low during and / or after HD** • Incorrectly reported - under valued on 9 (5%) charts • Not reported when present on 75 (44%) charts • Reported but raw data missing on 3 (2%) charts.	Under reported
Other notes on HD chart in relation to other items on dependency form	On 14 occasions, items were noted on the HD chart that were not scored on the dependency form and represented an underscoring by 33 points overall.	Under reported
HD chart signed	*32 (20%) of HD charts were not signed.*	*Not applicable*

* Because analyses used scores calculated from raw data

** Denominator 170 to account for 7 patients with high and low readings on audit

Glossary

This glossary applies to terms used in the thesis, although in some cases there are alternative definitions.

Accurate	Correct and valid (i.e. measures what it purports to measure)
Average cost	Total cost divided by quantity of output
Cost	The value of resources used in production
Direct care	Care that is face-to-face
Fixed costs	Costs for inputs that do not vary by outputs
General activity	Care that does not relate to a specific patient
Health Care Assistant	'Un-qualified' (unregistered) nurse incorporating health care assistants, health care support workers, or auxiliaries and dialysis assistants (e.g. SATOs)
Indirect care	Care for a specific patient that is not face-to-face
Marginal cost	The change in total cost for an extra unit of output
Opportunity cost	The value of benefits from the resources used for their best alternative, regardless of whether bought
Overheads	Shared resource use such as human resources and estates, which do not directly link to the output of interest
Patient heterogeneity	Variation between patients, for example by physical, mental, and social characteristics, clinical diagnosis, procedures or illness severity. In the context of the thesis, patient heterogeneity is used particularly in relation to patients' variations in resource use
Patient-specific	Direct and indirect care
Precise	Exact (e.g. having narrow 95% confidence intervals)
Price	The value consumers are willing to pay for a product, which incorporates both production costs and profit
Productive time	Productive time is for shifts worked and comprises general activity and patient-specific time
Reliable	Dependable and repeatable
Resources	The inputs (labour, capital and materials) used to produce goods or services
Unproductive time	Time for annual leave, study leave, sickness, etc., covered by nurses' annual salaries
Valid	Measures what it purports to measure
Variable costs	Costs for inputs that vary by quantity of output

References

Adam T, Evans DB, Koopmanschap MA (2003a) Cost-effectiveness analysis: Can we reduce variability in costing methods? International Journal of Technology Assessment in Health Care 19(2): 407-420.

Adam T, Evans D, Murray C (2003b) Econometric estimation of country-specific hospital costs. Cost Effectiveness and Resource Allocation 1(1): 3.

Agar JW, Knight RJ, Simmonds RE, Boddington JM, Waldron CM, Somerville CA (2005) Nocturnal haemodialysis: An Australian cost comparison with conventional satellite haemodialysis. Nephrology 10(6): 557-570.

Aird B (1996) Activity-based cost management in health care - another fad? International Journal of Health Care Quality Assurance 9(4): 16-19.

Altman DG (1991) Practical statistics for medical research. Chapman and Hall, London.

Altman DG & Bland JM (1995) Statistics notes: Absence of evidence is not evidence of absence. British Medical Journal 311(7003): 485.

Ansell D, Feest T, Byrne C (2002) UK Renal Registry Report 2002. 5th Annual Report. UK Renal Registry, Bristol.

Ansell D, Feest T, Byrne C, Ahmad A (2003) UK Renal Registry Report 2003. 6th Annual Report. UK Renal Registry, Bristol.

Ansell D, Feest T, Ahmad A, Rao R (2004) UK Renal Registry Report 2004. 7th Annual Report. UK Renal Registry, Bristol.

Ansell D, Feest T, Rao R, Williams A, Winearls C (2005) UK Renal Registry Report 2005. 8th Annual Report. UK Renal Registry, Bristol.

Ansell D, Feest TG, Tomson C, Williams AJ, Warwick G (2006) UK Renal Registry Report 2006. 9th Annual Report. UK Renal Registry, Bristol.

Ansell D, Feehally J, Feest TG, Tomson C, Williams AJ, Warwick G (2007) UK Renal Registry Report 2007. 10th Annual Report. UK Renal Registry, Bristol.

Audit Commission (2008) The right result? Payment by Results 2003-07. Audit Commission, London.

Averill RF (1991) Development. In Fetter RB, Brand DA, Gmache D eds. DRGs Their design and development. Health Administration Press, Ann Arbor, Michigan. Ch 2: 28-41.

Balogh R (1992) Audits of nursing care in Britain: A review and a critique of approaches to validating them. International Journal of Nursing Studies 29(2): 119-133.

Bardsley M & Coles J (1989) Case mix and management issues. In Bardsley M, Coles J, Jenkins L eds. DRGs and health care. The management of case mix. 2nd edition. King's Edward's Hospital Fund for London, London. Ch 7: 106-112.

Bardsley M, Coles J, Jenkins L eds. (1989) DRGs and health care. The management of case mix. 2nd edition. King's Edward's Hospital Fund for London, London.

Barnes RM (1958) Motion and time study. 4th edition. Wiley & Sons, New York.

Beck EJ, Beecham J, Mandalia S, Griffith R, Walters MDS, Boulton M, Miller DL (1999) What is the cost of getting the price wrong? Journal of Public Health Medicine 21(3): 311-317.

Beddhu S, Bruns FJ, Saul M (2000) A simple comorbidity scale predicts clinical outcomes and costs in dialysis patients. American Journal of Medicine 108(8): 609-613.

Beech R, Gulliford M, Roderick P (2004) Renal disease. In Stevens A, Raftery J, Mant J, Simpson S eds. Health Care Needs Assessment. Volume 1. 2nd Edition. Radcliffe Publishing Ltd, Oxford. Ch 2: 75-140.

Beecham J, Knapp M, Fenyo A (1993) Costs, needs and outcomes in community mental health care. In Netten A & Beecham J eds. Costing community care. Theory and practice. Ch 11: 162-176.

Bernardini J, Callen S, Fried L, Piraino B (2004) Inter-rater reliability and annual rescoring of the Charlson comorbidity index. Advances in Peritoneal Dialysis 20: 125-127.

Bjorvatn A (2005) Evaluation of decentralized hemodialysis in Norway: A cost-benefit analysis. Dialysis & Transplantation 34(10): 684-691+730.

Black G (2005) Introduction to accounting and finance. Financial Times/Prentice Hall, Harlow.

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

BMA (British Medical Association and Royal Pharmaceutical Society of Great Britain) (2006) British National Formulary. No. 52. BMJ Publishing and RPS Publishing, London.

Boland A, Haycox A, Bagust A, Fitzsimmons L (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).

Boyle PJ, Kudlac H, Williams AJ (1996) Geographical variation in the referral of patients with chronic end stage renal failure for renal replacement therapy. Quarterly Journal of Medicine. 89(2): 151-157.

Brand DA, Smits HL, Cooney LM, Schneider KC, Murtaugh CM, Fetter RB (1991) The PDG system for long-term care. In Fetter RB, Brand DA, Gmache D eds. DRGs Their design and development. Health Administration Press, Ann Arbor, Michigan. Ch 13: 307-332.

Bratt JH, Foreit J, Chen PL, West C, Janowitz B, De Vargas T (1999) A comparison of four approaches for measuring clinician time use. Health Policy and Planning 14(4): 374-381.

Brazier J & Deverill M (1999) A checklist for judging preference-based measures of health related quality of life: Learning from psychometrics. Health Economics 8(1): 41-51.

Brazier J, Deverill M, Green C, Harper R, Booth A (1999) A review of the use of health status measures in economic evaluation. Health Technology Assessment 3(9).

Brazier J, Roberts J, Tsuchiya A, Busschbach J (2004) A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics 13(9): 873-884.

Briggs AH & Gray AM (1999) Handling uncertainty when performing economic evaluation of health care interventions. Health Technology Assessment 3(2).

Briggs A, Clark T, Wolstenholme J, Clarke P (2003) Missing.... presumed at random: Cost-analysis of incomplete data. Health Economics 12(5): 377-392.

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

Brown S & Green A eds. (2006) Introductory guide NHS finance in the UK 8th edition. Healthcare Financial Management Association, Bristol.

Burke TA, McKee JR, Wilson HC, Donahue RM, Batenhorst AS, Pathak DS (2000) A comparison of time-and-motion and self-reporting methods of work measurement. Journal of Nursing Administration 30(3): 118-125.

Buxton MJ & West RR (1975) Cost-benefit analysis of long-term haemodialysis for chronic renal failure. British Medical Journal 2(5967): 376-379.

Caine N, Sharples LD, Hollingworth W, French J, Keogan M, Exley A, Hodgkins D, Bilton D (2002) A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic. Health Technology Assessment 6(27).

Carpenter I, Perry M, Challis D, Hope K (2003) Identification of registered nursing care of residents in English nursing homes using the Minimum Data Set Resident Assessment Instrument (MDS/RAI) and Resource Utilisation Groups version III (RUG-III). Age & Ageing. 32(3): 279-285. [erratum in Age Ageing. 2003 32(6): 686].

Carr-Hill R & Jenkins-Clarke S (2003) Improving the effectiveness of the nursing workforce. Short report of analysis of NISCM dataset. Centre for Health Economics, University of York, York.

CCOHTA (1996) A guidance document for the costing process. Version 1.0. Canadian Coordinating Office for Health Technology Assessment. Ottawa, Canada.

Chandna SM, Schulz J, Lawrence C, Greenwood RN, Farrington K (1999) Is there a rationale for rationing chronic dialysis? A hospital based cohort study of factors affecting survival and morbidity. British Medical Journal 318: 217-223.

Charlson ME, Pompei P, Ales KL (1987) A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases 40(5): 373-383.

Cheshire and Merseyside Renal Transport Action Learning Set (2006). Recommendations for the provision of a patient centred renal transport service. From http://www.kidney.org.uk/campaigns/Dialysis/PatientTransport/Renal-Transport-Cheshire-and-Merseyside.pdf (accessed 24/11/06).

Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P (2003) The role of modelling in prioritising and planning clinical trials. Health Technology Assessment 7(23).

Churchill DN, Lemon BC, Torrance GW (1984) A cost-effectiveness analysis of continuous ambulatory peritoneal dialysis and hospital hemodialysis. Medical Decision Making 4(4): 489-500.

Coles J (1989) Attributing costs and resource use to case types. In Bardsley M, Coles J, Jenkins L eds. DRGs and health care. The management of case mix. 2nd edition. King's Edward's Hospital Fund for London, London. Ch 6: 85-98.

Connell N, Lees D, Lees P, Powell P, Stafford P, Sutcliffe C (1996) Costing and contracting in the NHS: A decision support approach. In Bourn M & Sutcliffe C eds. Management accounting in healthcare. Chartered Institute of Management Accountants, London. Chapter 3: 30-40.

Connell C, Lees P, Powell PL, Sutcliffe CMS, Lees D, Searle H, Stafford P, Haines M (1997) Developing decision support systems: A study of healthcare management. The Chartered Institute of Management Accountants, London.

Conner-Spady B & Suarez-Almazor ME (2003) Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Medical Care 41(7): 791-801.

Cooper R & Kaplan RS (1988) Measure costs right: Make the right decisions. Harvard Business Review (Sept-Oct): 96-103.

Cromwell J, Maier J, Gage B, Drozd E, Osber D, Richter E, Greenwald L, Goldman H (2004) Characteristics of high staff intensive Medicare psychiatric inpatients. Health Care Financing Review. 26(1): 103-117.

Curtis L (2007) Unit costs of health and social care 2007. 15th edition. Personal & Social Services Research Unit, University of Kent, Canterbury.

Curtis L & Netten A (2004) Editorial. New developments and data sources for unit costs. In Curtis L & Netten A eds. Unit costs of health and social care 2004. 12th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 1-9.

Curtis L & Netten A (2005) Unit costs of health and social care 2005. 13th edition. Personal & Social Services Research Unit, University of Kent, Canterbury.

Curtis L & Netten A (2006) Unit costs of health and social care 2006. 14th edition. Personal & Social Services Research Unit, University of Kent, Canterbury.

Dawson D & Street A (1998) Reference Costs and the pursuit of efficiency in the 'new' NHS. Discussion paper 161. Centre for Health Economics, University of York, York.

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.

De Wit GA, Ramsteijn PG, de Charro FT (1998) Economic evaluation of end stage renal disease treatment. Health Policy 44(3): 215-232.

Department of Health (1996) Renal Purchasing Guidelines. NHS Executive, London.

Department of Health (2000a) The NHS Plan: A plan for investment, a plan for reform. The Stationery Office. London.

Department of Health (2000b) Towards a strategy for nursing research & development. Department of Health, London.

Department of Health (2002a) Delivering the NHS Plan: Next steps on investment, next steps on reform. The Stationery Office. London.

Department of Health (2002b) Reforming NHS Financial Flows. Introducing payment by results. Department of Health, London.

Department of Health (2006a) Indicative tariff. Renal Dialysis Indicative Tariff 2006-07. From http://www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/NHSFinancialReforms/DH_4127039 (accessed 14/4/08).

Department of Health (2006b) Market Forces Factors 2006/07. .

Department of Health (2006c) National Schedule of Reference Costs - Appendix NSRC4 - NHS Trust and PCT combined reference cost schedules. From http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_062884 (accessed 23/11/07).

Department of Health (2006d) Our health, our care, our say: Anew direction for community services. The Stationery Office, Norwich.

Department of Health (2006f) Payment by Results Guidance 2007/08. Version 1 (December). Department of Health, Leeds.

Department of Health (2007a) Coding for Success: Simple technology for safer patient care. Department of Health, London.

Department of Health (2007b) Definition: patient-level information and costing systems. Department of Health, London.

Department of Health (2007c) HCHS Pay & Prices Series 2005-06 - Pay & Prices Series 2005-06. HCHS Pay & price inflation. From http://www.info.doh.gov.uk/doh/finman.nsf/af3d43e36a4c8f8500256722005b77f8/312613abdec153f5802572bc00321222?OpenDocument (accessed 18/3/08).

Department of Health (2007d) NHS Pensions Scheme. Department of Health, London.

Department of Health (2007e) Payment by results guidance 2008-09 annexes - amended 17 December 2007. (Annex B). Department of Health, London.

Department of Health (2007f) Reference Costs Collection Guidance 2006-07. Department of Health (PbR - Finance and costing team), London.

Department of Health (2008a) National Schedule of Reference Costs 2006-07 - Appendix NSRC4 - NHS trust and PCT combined reference cost schedules. From http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_082571 (accessed 25/3/08).

Department of Health (Payment by Results Finance and Costing Team) (2008b) NHS costing manual 2007/08. Department of Health.

Department of Health Renal Team (2004) The National Service Framework for Renal Services. Part One: Dialysis and Transplantation. Department of Health, London.

Donaldson C, Gerard K, Jan S, Mitton C, Wiseman V (2005) Economics of health care financing. The visible hand. 2nd edition. Palgrave Macmillan, Basingstoke.

Doshi JA, Glick HA, Polsky D (2006) Analyses of cost data in economic evaluations conducted alongside randomized controlled trials. Value in Health 9(5): 334-340.

Drummond MF & Jefferson TO (1996) Guidelines for authors and peer reviewers of economic submissions to the BMJ. British Medicial Journal 313(7052): 275-283.

Drummond M & McGuire A eds. (2001) Economic evaluation in health care. Merging theory with practice. Oxford University Press, Oxford.

Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL (2005) Methods for the economic evaluation of health care programmes. 3rd edition. Oxford University Press, Oxford.

Dupont WD (2002) Statistical modeling for biomedical researchers. A simple introduction to the analysis of complex data. Cambridge University Press, Cambridge.

Eastham J (2006) Workload measurement and costs in a U.K. Accident and Emergency department: a health economics analysis. PhD thesis Keele University.

Evans C & Crawford B (1999) Patient self-reports in pharmacoeconomic studies. Their use and impact on study validity. PharmacoEconomics 15(3): 241-256.

Excel – Microsoft Office Excel (2003) Version 11. Microsoft Corporation. Washington.

Feest TG, Rajamahesh J, Byrne C, Ahmad A, Ansell D, Burden R, Roderick PJ (2005) Trends in adult renal replacement therapy in the UK: 1982-2002. Quarterly Journal of Mediciine 98(1): 21-28.

Fetter RB (1989) Introduction. In Bardsley M, Coles J, Jenkins L eds. DRGs and health care. The management of case mix. 2nd edition. King's Edward's Hospital Fund for London, London. 5-10.

Fetter RB (1991) Background. In Fetter RB, Brand DA, Gmache D eds. DRGs Their design and development. Health Administration Press, Ann Arbor, Michigan. Ch 1: 3-27.

Finkler SA, Knickman JR, Hendrickson G, Lipkin M, Jr., Thompson WG (1993) A comparison of work-sampling and time-and-motion techniques for studies in health services research. Health Services Research 28(5): 577-597.

Fitzpatrick R, Davey C, Buxton MJ, Jones DR (1998) Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment 2(14).

Frazis H, Stewart J (2004) What can time-use data tell us about hours of work? Monthly Labor Review 127(12): 3-9.

Freeman JL, Fetter RB, Hayoung P, Schneider KC, Lichtenstien JL, Bauman WA, Duncan CC, Hughes JS, Freeman DH, Palmer GR (1991) Refinement. In Fetter RB, Brand DA, Gmache D eds. DRGs Their design and development. Health Administration Press, Ann Arbor, Michigan. Ch 3: 57-78.

Freund L, Burrows-Hudson S, Preisig P (1998) Development of a patient classification system for chronic hemodialysis patients. American Journal of Kidney Diseases 31(5): 818-829.

Garber AM & Phelps CE (1997) Economic Foundations of Cost-Effectiveness Analysis. Journal of Health Economics 16(1): 1-31.

Gerard K, Nicholson T, Mullee M, Mehta R, Roderick P (2004) EQ-5D versus SF-6D in an older, chronically ill patient group. Applied Health Economics & Health Policy. 3(2): 91-102.

Gibbs I, McCaughan D, Griffiths M (1991) Skill mix in nursing: a selective review of the literature. Journal of Advanced Nursing 16(2): 242-249.

Giuffrida A, Gravelle H, Sutton M (2000) Efficiency and administrative costs in primary care. Journal of Health Economics 19(3): 983-1006.

Glick HA, Polton DP, Schulman KA (2001) Trial-based economic evaluations: An overview of design and analysis. In Drummond M & McGuire A eds. Economic evaluation in health care. Merging theory with practice. Oxford University Press, Oxford. Ch 6: 112-140.

Glick H, Doshi JA, Sonnad SS, Polsky D (2007) Economic evaluation in clinical trials. Oxford University Press, Oxford.

Goddard A & Ooi K (1998) Activity-based costing and central overhead cost allocation in universities: A case study. Public Money & Management 18(3): 31-38.

Gold MR, Siegel JE, Russell LB, Weinstein MC eds. (1996) Cost-effectiveness in health and medicine. Oxford University Press, New York.

Gonzalez-Perez JG, Vale L, Stearns SC, Wordsworth S (2005) Hemodialysis for end-stage renal disease: A cost-effectiveness analysis of treatment options. International Journal of Technology Assessment in Health Care. 21(1): 32-39.

Graham B & McGregor K (1997) What does a GP consultation cost? British Journal of General Practice 47(416): 170-172.

Graves N, Walker D, Raine R, Hutchings A, Roberts JA (2002) Cost data for individual patients included in clinical studies: No amount of statistical analysis can compensate for inadequate costing methods. Health Economics 11(8): 735-739.

Grieve R, Nixon R, Thompson SG, Normand C (2005) Using multilevel models for assessing the variability of multinational resource use and cost data. Health Economics 14(2): 185-196.

Gruen R & Howarth A (2005) The pricing process in different settings. In Financial Management in Health Services. Open University Press, Maidenhead. Ch 6: 66-78.

Gunda S, Thomas M, Smith S (2005) National survey of palliative care in end-stage renal disease in the UK. Nephrology Dialysis & Transplantation. 20: 392-395.

Gunn GT, Douglas PH (1942) The Production Function for American Manufacturing for 1914. The Journal of Political Economy 50(4): 595-602.

Halliday RG & Darba J (2003) Cost data assessment in multinational economic evaluations: Some theory and review of published studies. Applied Health Economics and Health Policy 2(3): 149-155.

Hansen BL (1960) Work sampling for modern management. Prentice-Hall: Englewood Cliffs New Jersey.

Harrison J (2004) Addressing increasing patient acuity and nursing workload. Nursing Management. 11(4): 20-25.

Heerey A, McGowan B, Ryan M, Barry M (2002) Microcosting versus DRGs in the provision of cost estimates for use in pharmacoeconomic evaluation. Expert Review of Pharmacoeconomics and Outcomes Research 2(1): 29-33.

Hemmelgarn BR, Manns BJ, Quan H, Ghali WA (2003) Adapting the Charlson Comorbidity Index for use in patients with ESRD. American Journal of Kidney Diseases 42(1): 125-132.

Himmelfarb J, Chertow GM (2005) Medicare ESRD Prospective Payment System: Weighing the Evidence. Journal of the American Society of Nephrology 16(5): 1164-1165.

Hirth R (2007) The organization and financing of kidney dialysis and transplant care in the United States of America. International Journal of Health Care Finance and Economics 7(4): 301-318.

HM Revenue and Customs (2005) Contracted-out contributions for employers with Contracted-out Salary Related Schemes. CA39 National Insurance Contributions Tables (2006 to 2007). From http://www.hmrc.gov.uk/nitables/2006/ca39.pdf (accessed 7/4/08).

Holmes D, Teresi J, Lindeman DA, Glandon GL (1997a) 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, Goldman D (1997b) 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.

Horrocks S, Anderson E, Salisbury C (2002) Systematic review of whether nurse practitioners working in primary care can provide equivalent care to doctors. British Medical Journal 324(7341): 819-823.

Horton NJ & Lipsitz SR (1999) Review of software to fit generalized estimating equation regression models. The American Statistician 53(2): 160-169.

http://www.dh.gov.uk/en/Policyandguidance/Humanresourcesandtraining/Modernisingworkforceplanninghome/index.htm (accessed 12/12/07).

http://www.oecd.org/dataoecd/61/56/39653523.xls Purchasing Power Parities (PPPs for GDP) for OECD countries since 1980 (accessed 24/3/08).

http://www.performance.doh.gov.uk/HPSSS/TBL_E3.HTM Health and Personal Social Services Statistics. Section E – Expenditure. E3 Analysis of health authority and NHS trust expenditure (year end 31 March) (accessed 23/11/07).

Hughes D (1991) Costing Consultations in General practice: Towards a Standardized Method. Family Practice 8(4): 388-393.

Hurst K (2002) Selecting and applying methods for estimating the size and mix of nursing teams. Nuffield Institute for Health, Leeds.

Hutchinson TA, Boyd NF, Feinstein AR, Gonda A, Hollomby D, Rowat B (1979) Scientific problems in clinical scales, as demonstrated in the Karnofsky Index of performance status. Journal of Chronic Diseases 32: 661-666.

Hutton J (2001) Unit costs for multi-country economic evaluations. In Netten A, Rees T, Harrison G eds. Unit costs of health and social care 2001. 9th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 31-34.

Ifudu O, Mayers J, Matthew J, Tan CC, Cambridge A, Friedman EA (1994) Dismal rehabilitation in geriatric inner-city hemodialysis patients. JAMA: The Journal of the American Medical Association 271(1): 29-33.

Ifudu O, Paul HR, Homel P, Friedman EA (1998) Predictive value of functional status for mortality in patients on maintenance hemodialysis. American Journal of Nephrology 18(2): 109-116.

Itami H & Kaplan RS (1980) An activity analysis approach to unit costing with multiple interactive products. Management Science 26(8): 826-839

Jacobs P, Ohinmaa A, Brady B (2005) Providing systematic guidance in pharmacoeconomic guidelines for analysing costs. Pharmacoeconomics 23(2): 143-153.

Jan S, Kumaranayake L, Roberts J, Hanson K, Archibald K (2005) Economic analysis for management and policy. Open University Press, Maidenhead.

Jassal SV, Brissenden JE, Raisbeck A, Roscoe JM (1998) Comparative cost-analysis of two different chronic care facilities for end-stage renal disease patients. Geriatric Nephrology & Urology 8(2): 69-76.

Johannesson M, Jonsson B, Karlsson G (1996) Outcome measurement in economic evaluation. Health Economics 5(4): 279-296.

Johnston K, Buxton MJ, Jones DR, Fitzpatrick R (1999) Assessing the cost of health care technologies in clinical trials. Health Technology Assessment 3(6).

Jones KR (1992) Variations in the hemodialysis treatment process. Clinical Nursing Research 1(1): 50-66.

Kaplan RS (1990) Measure costs right: Making the right decision. (Analysis of activity based accounting). The CPA Journal Online Feb.

Karnofsky D & Burcherval J (1949) The Clinical Evaluation of Chemotherapeutic Agents in Cancer. In Macleod C ed. Evaluation of chemotherapeutic agents. Columbia University Press, New York. 191-205.

Keane WF & Collins AJ (1994) Influence of co-morbidity on mortality and morbidity in patients treated with hemodialysis. American Journal of Kidney Diseases 24(6):1010-1018.

Kennedy A, Leigh-Brown A, Torgerson D, Campbell J, Grant A (2002) Resource use data by patient report or hospital records: Do they agree? BMC Health Services Research 2(1): 2.

Kernick DP & Netten A (2002) A methodological framework to derive the cost of the GP consultation. Family Practice 19(5): 500-503.

Khan IH, Catto GR, Edward N, Fleming LW, Henderson IS, MacLeod AM (1993) Influence of coexisting disease on survival on renal-replacement therapy. Lancet 341(8842): 415-418.

Kidney Epidemiology and Cost Center (2004) Methodology for Developing a Basic Case-Mix Adjustment for the Medicare ESRD Prospective Payment System. Kidney Epidemiology and Cost Center, The University of Michigan, Ann Arbor.

Kinley H, Czoski-Murray C, George S, McCabe C, Primrose J, Reilly C, Wood R, Nicolson P, Healy C, Read S, Norman J, Janke E, Alhameed H, Fernandez N, Thomas E (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).

Kirkwood BR & Sterne JAC (2003) Essential medical statistics. 2nd edition. Blackwell Science, Oxford.

Klarman HE, Francis JO, Rosenthal GD (1968) Cost effectiveness analysis applied to the treatment of chronic renal disease. Medical Care 6(1): 48-54.

Klein R (2007) Outcomes of variation in hospital nurse staffing in English hospitals: A lesson for policy makers. International Journal of Nursing Studies 44(2): 167-168.

Knapp MRJ & Beecham J (1993) Reduced list costings: Examination of an informed short cut in mental health research. Health Economics. 2: 313-322.

Kroeker A, Clark WF, Heidenheim AP, Kuenzig L, Leitch R, Meyette M, Muirhead N, Ryan H, Welch R, White S, Lindsay RM (2003) An operating cost comparison between conventional and home quotidian hemodialysis. American Journal of Kidney Diseases 42(1 Suppl): 49-55.

Lader D, Short S, Gershuny J (2006) The Time Use Survey, 2005. How we spend our time. Office for National Statistics, London.

Lamb G, Kip KE, Mapes D, Burrows-Hudson S, Cesta T, Zazworsky D (2004) Reliability and validity of a risk assessment tool for patients with kidney failure. Nephrology Nursing Journal: Journal of the American Nephrology Nurses' Association 31(3): 267-270.

Larson-Lohr V (2003) Productivity and acuity tool for wound care and hyperbaric medicine centers. Undersea & Hyperbaric Medicine. 30(4): 271-275.

Laurant M, Reeves D, Hermens R, Braspenning J, Grol R, Sibbald B (2004) Substitution of doctors by nurses in primary care. Cochrane Database of Systematic Reviews (4).

Laurila J, Suramo I, Brommels M, Tolppanen EM, Koivukangas P, Lanning P, Standertskjold-Nordenstam G (2000) Activity-based costing in radiology. Application in a pediatric radiological unit. Acta Radiologica 41(2): 189-195.

Lave JR & Lave LB (1970) Hospital Cost Functions. The American Economic Review 60(3): 379-395. American Economic Association.

Lave JR, Lave LB, Silverman LP (1972) Hospital cost estimation controlling for case-mix. Applied Economics 4(3): 165-180.

Lee H, Manns B, Taub K, Ghali WA, Dean S, Johnson D, Donaldson C (2002) Cost analysis of ongoing care of patients with end-stage renal disease: The impact of dialysis modality and dialysis access. American Journal of Kidney Diseases 40(3): 611-622.

Lee RH, Bott MJ, Forbes S, Redford L, Swagerty DL, Taunton RL (2003) Process-based costing. Journal of Nursing Care Quality 18(4): 259-266.

Liang K-Y & Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73(1): 13-22.

Lim TO, Lim YN, Wong HS, Ahmad G, Singam TS, Morad Z, Suleiman AB, Rozina G, Ong LM, Hooi LS, Shaariah W, Tan CC, Loo CS (1999) Cost effectiveness evaluation of the Ministry of Health Malaysia dialysis programme. Medical Journal of Malaysia 54(4): 442-452.

Lindsay RM (2004) The London, Ontario, Daily/Nocturnal Hemodialysis Study. Seminars in Dialysis 17(2): 85-91.

Lindsay RM, Leitch R, Heidenheim AP, Kortas C (2003) The London daily/nocturnal hemodialysis study - Study design, morbidity, and mortality results. American Journal of Kidney Diseases 42(1 Suppl 1): S5-S12.

Little R (2005) Economic evaluation of intermediate care. In Roe B & Beech R eds. Intermediate and continuing care. Policy and practice. Blackwell Publishing, Oxford. Ch 9: 119-135.

Loftware Label Manager 7 (2003) Version 7.2.2.11. Loftware Inc, York ME.

Longworth L & Bryan S (2003) An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health Economics. 12(12): 1061-1067.

Luce BR, Manning WG, Siegel JE, Lipscomb J (1996) Estimating cost in cost-effectiveness analysis. In Gold MR, Siegel JE, Russell LB, Weinstein MC eds. Cost-effectiveness in health and medicine. Oxford University Press, New York. Ch 6: 176-213.

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

Mallick N (1994) Report of the Health Care Strategy Unit. Review of renal services. Part I: Findings and recommendations and Part II: Evidence for the review. Department of Health, Health Care Strategy Unit, London.

Mallick NP (1997) The costs of renal services in Britain. Nephrology Dialysis & Transplantation. 12 Suppl 1: 25-28.

Manca A, Rice N, Sculpher MJ, Briggs AH (2005) Assessing generalisability by location in trial-based cost-effectiveness analysis: The use of multilevel models. Health Economics 14(5): 471-486.

Mankiw NG (2004) Principles of economics. International student edition. 3rd. Thomson South-Western, Mason, Ohio.

Manning WG & Mullahy J (2001) Estimating log models: To transform or not to transform? Journal of Health Economics. 20(4): 461-494.

Manning WG, Basu A, Mullahy J (2005) Generalized modeling approaches to risk adjustment of skewed outcomes data. Journal of Health Economics 24(3): 465-488.

Manns BJ, Mortis GP, Taub KJ, McLaughlin K, Donaldson C, Ghali WA (2001) The Southern Alberta Renal Program database: a prototype for patient management and research initiatives. Clinical & Investigative Medicine - Medecine Clinique et Experimentale 24(4): 164-170.

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

Matthews JN, Altman DG, Campbell MJ, Royston P (1990) Analysis of serial measurements in medical research. British Medical Journal 300 (6719): 230-235.

McClellan WM, Anson C, Birkeli K (1991) Functional status and quality of life predictors of early mortality among patients entering treatment for end stage renal disease. Journal of Clinical Epidemiology 44(1): 83-89.

McFarlane PA, Pierratos A, Redelmeier DA (2002) Cost savings of home nocturnal versus conventional in-center hemodialysis. Kidney International 62(6): 2216-2222.

McFarlane PA, Bayoumi AM, Pierratos A, Redelmeier DA (2003) The quality of life and cost utility of home nocturnal and conventional in-center hemodialysis. Kidney International 64(3): 1004-1011.

McFarlane PA, Bayoumi AM, Pierratos A, Redelmeier DA (2006) The impact of home nocturnal hemodialysis on end-stage renal disease therapies: A decision analysis. Kidney International 69(5): 798-805.

Mellett HF, Marriott N, Harries S (1993) Financial management in the NHS: A *manager's handbook. Chapman & Hall, London.*

Meyer D (1984) Manpower planning. One. An American approach. Nursing Times 80(34): 52-54.

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

MiniPro 1.0 (2004) CS1504 Barcode Extraction software. MMR Software, Bethesda.

Miskulin D (2005) Characterizing comorbidity in dialysis patients: principles of measurement and applications in risk adjustment and patient care. Peritoneal Dialysis International 25(4): 320-332.

Mistry H, Buxton M, Longworth L, Chatwin J, Peveler R (2005) Comparison of general practitioner records and patient self-report questionnaires for estimation of costs. The European Journal of Health Economics (Health Economics in Prevention and Care) 6(3): 261-266.

Mogyorosy Z & Smith P (2005) The main methodological issues in costing health care services. A literature review. CHE Research report 7. Centre for Health Economics, University of York, York.

Mohr PE (2001) The economics of daily dialysis. Advances in Renal Replacement Therapy 8(4): 273-279.

Mohr PE, Neumann PJ, Franco SJ, Marainen J, Lockridge R, Ting G (2001) The case for daily dialysis: Its impact on costs and quality of life. American Journal of Kidney Diseases 37(4): 777-789.

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

Morris R, MacNeela P, Scott A, Treacy P, Hyde A (2007) Reconsidering the conceptualization of nursing workload: literature review. Journal of Advanced Nursing 57(5): 463-471.

Mowatt G, Vale L, Perez J, Wyness L, Fraser C, MacLeod A, Daly C, Stearns SC (2003) Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of home versus hospital or satellite unit haemodialysis for people with end-stage renal failure. Health Technology Assessment 7(2).

Mullen PM (2000) When is delphi not delphi? University of Birmingham, Health Services Management Centre, Birmingham. Discussion Paper 37.

Mullen PM (2003) Delphi: Myths and reality. Journal of Health, Organisation and Management.17(1): 37-52.

Netten A (1993) Costs, prices and charges. In Netten A & Beecham J eds. Costing community care. Theory and practice. Ashgate, Aldershot. Ch 7: 87-102.

Netten A (1996) Editorial: Changing estimates of unit costs over time. In Netten A & Dennett J eds. Unit costs of health and social care 1996. 4th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 7-13.

Netten A (1999) Editorial: Editorial. Unit costs and performance indicators. In Netten A, Dennett J, Knight J eds. Unit costs of health and social care 1999. 7th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 7-10.

Netten A (2001) Editorial: Presentation and types of unit cost information. In Netten A, Rees T, Harrison G eds. Unit costs of health and social care 2001. 9th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 7-12.

Netten A (2002) Editorial: Updating unit costs of health and social care. In Netten A & Curtis L eds. Unit costs of health and social care 2002. 10th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 7-15.

Netten A (2003) Editorial: New developments and changes in guidance on the discount rate. In Netten A & Curtis L eds. Unit costs of health and social care 2003. 11th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 7-13.

Netten A & Beecham J eds. (1993) Costing community care. Theory and practice. Ashgate, Aldershot.

Netten A & Beecham J (1999) Estimating unit costs. In Netten A, Dennett J, Knight J eds. Unit costs of health and social care 1999. 7th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 11-18.

Netten A & Curtis L (2005) Adapting and adjusting unit cost information. In Curtis L & Netten A eds. Unit costs of health and social care 2005. 13th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 1-8.

Netten A, Knight J, Dennett J, Cooley R, Slight A (1998) A 'Ready Reckoner' for Staff Costs in the NHS 1998, Volume II, Methodology. Discussion paper 1476. Personal & Social Services Research Unit, University of Kent. Canterbury.

Netten A, Dennett J, Knight J (1999) Unit costs of health and social care 1999. 7th edition. Personal & Social Services Research Unit, University of Kent, Canterbury.

NHS Employers (2006) 25 May 2006 Pay Circular (AforC) 1/2006. Pay and conditions for NHS Staff covered by the Agenda for Change agreement.

NHS Information Authority (2003) Version 3.5 Healthcare Resource Group Documentation Set. NHS Information Authority, Winchester.

NHS Whitley Council (2004) NHS Whitley Council provisions (extracts). Working or providing emergency cover outside normal hours. References. Agenda for Change Final Agreement. Chapter 2: Interim Regime. NHS Terms and Conditions of Service Handbook. From: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets /@dh/@en/documents/digitalasset/dh_4096592.pdf (accessed 20/8/2005).

NICE (National Institute for Clinical Excellence) (2002) Guidance on home compared with hospital haemodialysis for patients with end-stage renal failure. Technology Appraisal Guidance No. 48. National Institute for Clinical Excellence, London.

NICE (National Institute for Clinical Excellence) (2004) Guide to the methods of technology appraisal. National Institute for Clinical Excellence, London.

NICE (National Institute for Health and Clinical Excellence) (2006) About NICE guidance: what does it mean for me? Information for patients, carers and the public -

an interim guide. From http://www.nice.org.uk/nicemedia/pdf/AboutGuidance.pdf (accessed 25/3/08).

Nicholson T, Roderick P (2007) International Study of Health Care Organization and Financing of renal services in England and Wales. International Journal of Health Care Finance and Economics. 7(4):282-299.

Nisenbaum HL, Birnbaum BA, Myers MM, Grossman RI, Gefter WB, Langlotz CP (2000) The costs of CT procedures in an academic radiology department determined by an activity-based costing (ABC) method. Journal of Computer Assisted Tomography 24(5): 813-823.

O'Brien BJ, Spath M, Blackhouse G, Severens JL, Dorian P, Brazier J (2003) A view from the bridge: Agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Economics. 12 (11): 975-981.

Oddone E, Guarisco S, Simel D (1993) Comparison of housestaff's estimates of their workday activities with results of a random work-sampling study. Academic Medicine 68(11): 859-861.

Oddone E, Weinberger M, Hurder A, Henderson W, Simel D (1995) Measuring activities in clinical trials using random work sampling: Implications for cost-effectiveness analysis and measurement of the intervention. Journal of Clinical Epidemiology 48(8): 1011-1018.

Oliver MN, Goodwin MA, Gotler RS, Gregory PM, Stange KC (2001) Time use in clinical encounters: Are African-American patients treated differently? Journal of the National Medical Association 93(10): 380-385.

Oostenbrink JB & Al MJ (2005) The analysis of incomplete cost data due to dropout. Health Economics 14(8): 763-776.

Oostenbrink JB, Koopmanschap MA, Rutten FFH. (2002) Standardisation of costs: The Dutch manual for costing in economic evaluations. PharmacoEconomics 20(7): 443-454.

Patel A, Foster J, Martin F (2003) The costs of intermediate care schemes. In Netten A & Curtis L eds.. Unit costs of health and social care 2003. 11th edition. Personal & Social Services Research Unit, University of Kent, Canterbury. 15-18

Patten J (1984) Kidney patients. House of Commons Official Report (Hansard) p710 cols 309-310.

Peeters P, Rublee D, Just PM, Joseph A (2000) Analysis and interpretation of cost data in dialysis: Review of Western European literature. Health Policy 54(3): 209-227.

Pelletier D & Duffield C (2003) Work sampling: Valuable methodology to define nursing practice patterns. Nursing & Health Sciences. 5(1): 31-38.

Petrou S, Murray L, Cooper P, Davidson LL (2002) The accuracy of self-reported health care resource utilization in health economic studies. International Journal of Technology Assessment in Health Care. 18(3): 705-710.

Phelps CE & Mushlin AI (1991) On the (near) equivalence of cost-effectiveness and cost-benefit analyses. International Journal of Technology Assessment in Health Care 7(1): 12-21.

Piccoli GB, Bermond F, Mezza E, Soragna G, Burdese M, Jeantet A, Segoloni GP, Piccoli G (2004) Home hemodialysis a la carte: A tailormade program (1998-2003). Journal of Nephrology 17(1): 76-86.

Pindyck RS & Rubinfeld DL (2005) Microeconomics. 6th edition. Pearson Prentice Hall, Upper Saddle River, New Jersey.

Plough AL, Salem SR, Shwartz M (1984) Case mix in end-stage renal disease. Differences between patients in hospital-based and free-standing treatment facilities. New England Journal of Medicine 310(22): 1432-1436.

Rafferty AM, Clarke SP, Coles J, Ball J, James P, McKee M, Aiken LH (2007) Outcomes of variation in hospital nurse staffing in English hospitals: Cross-sectional analysis of survey data and discharge records. International Journal of Nursing Studies 44(2): 175-182.

Raikou M & McGuire A (2004) Estimating medical care costs under conditions of censoring. Journal of Health Economics 23(3): 443-470.

Raikou M, Briggs A, Gray A, McGuire A (2000) Centre-specific or average unit costs in multi-centre studies? Some theory and simulation. Health Economics 9(3): 191-198.

Reinhardt U (1972) A Production Function for Physician Services. The Review of Economics and Statistics 54(1): 55-66.

Richardson G, Maynard A, Cullum N, Kindig D (1998) Skill mix changes: substitution or service development? Health Policy 45(2): 119-132.

Riewpaiboon A, Malaroje S, Kongsawatt S (2007) Effect of costing methods on unit cost of hospital medical services. Tropical Medicine & International Health 12(4): 554-563.

Roberts RO, Bergstralh EJ, Schmidt L, Jacobsen SJ (1996) Comparison of self-reported and medical record health care utilization measures. Journal of Clinical Epidemiology. 49(9): 989-995.

Roderick P, Clements S, Stone N, Martin D, Diamond I (1999) What determines *geographical variation in rates of acceptance onto renal replacement therapy in* England? Journal od Heath Services Research & Policy 4(3): 139-146.

Roderick P, Davies R, Jones C, Feest T, Smith S, Farrington KI (2004) Simulation model of renal replacement therapy: Predicting future demand in England. Nephrology Dialysis & Transplantation 19: 692-701.

Roderick P, Nicholson T, Armitage A, Mehta R, Mullee M, Gerard K, Drey N, Feest T, Greenwood R, Lamping D, Townsend J (2005) An evaluation of the costs, effectiveness and quality of renal replacement therapy provision in renal satellite units in England and Wales. Health Technology Assessment 9(24).

Roe B (2005) Service development and evaluation in intermediate care. In Roe B & Beech R eds. Intermediate and continuing care. Policy and practice. Blackwell Publishing, Oxford. Ch 5: 61-77.

Royal College of Nursing (2003) The future nurse: a discussion paper. Royal College of Nursing, London. Sanderson H (1989) DRGs in the NHS. In Bardsley M, Coles J, Jenkins L eds. DRGs and health care. The management of case mix. 2nd edition. King's Edward's Hospital Fund for London, London. Ch 5: 74-84.

Schulman K, Burke J, Drummond M, Davies L, Carlsson P, Gruger J, Harris A, Lucioni C, Gisbert R, Llana T, Tom E, Bloom B, Willke R, Glick H (1998) Resource costing for multinational neurologic clinical trials: Methods and results. Health Economics 7(7): 629-638.

Sculpher M, Claxton K, Akehurst R (2005) It's just evaluation for decision-making: Recent developments in, and challenges for, cost-effectiveness research. In Smith PC, Ginnelly L, Sculpher M eds. Health policy and economics: Opportunities and challenges. Open University Press, Maidenhead. Ch 1: 8-41.

Seshamani M & Gray A (2004) Ageing and health-care expenditure: The red herring argument revisited. Health Economics 13(4): 303-314.

Shepherd J, Jones J, Hartwell D, Davidson P, Price A, Waugh N (2007) Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of mild chronic hepatitis C: A systematic review and economic evaluation. Health Technology Assessment 11(11).

Shepperd S & Iliffe S (2005) Hospital at home versus in-patient hospital care. Cochrane Database of Systematic Reviews (Issue 3).

Shwartz M, Young DW, Siegrist R (1995) The ratio of costs to charges: how good a basis for estimating costs? Inquiry 32(4): 476-481.

Sittig DF (1993) Work-sampling: A statistical approach to evaluation of the effect of computers on work patterns in healthcare. Methods of Information in Medicine. 32(2): 167-174.

Smet M (2002) Cost characteristics of hospitals. Social Science & Medicine 55(6): 895-906.

Smet M (2004) Multi-product costs and standby capacity derived from queuing theory: The case of Belgian hospitals. Applied Economics 36(13): 1475-1487.

Soroka SD, Kiberd BA, Jacobs P (2005) The marginal cost of satellite versus in-center hemodialysis. Hemodialysis International 9(2): 196-201.

Sovie MD (1988) Variable costs of nursing care in hospitals. Annual Review of Nursing Research 6: 131-150.

Spilsbury K & Meyer J (2001) Defining the nursing contribution to patient outcome: lessons from a review of the literature examining nursing outcomes, skill mix and changing roles. Journal of Clinical Nursing 10(1): 3-14.

SPSS 14 (Statistics Package for the Social Sciences) (2005) Version 14.02 SPSS Inc. Chicago.

Standaert B, Goldstone J, Lu ZJ, Erder MH, Yin JL (2002) Economic analysis of filgrastim use for patients with acute myeloid leukaemia in the UK: A comparison of collection methods of resource use data. PharmacoEconomics 20(10): 665-674.

Stanton J (2005) Renal dialysis: Counting the cost versus counting the need. Clio Medica 75: 217-241.

Stat/Transfer 8 (2003). Circle Systems Inc., Seattle.

Stata 9 (2005) Inter-cooled Stata for Windows. Version 9.2. Stata Corporation, College Station, Texas, USA.

Stewart JA & Short FA (1999) Time accuracy of a bar code system for recording resuscitation events: Laboratory trials. Resuscitation 42(3): 235-240.

Stone PW, Chapman RH, Sandberg EA, Liljas B, Neumann PJ (2000) Measuring costs in cost-utility analyses: Variations in the literature. International Journal of Technology Assessment in Health Care 16(1): 111-124.

Swindle R, Lukas CV, Meyer DA, Barnett PG, Hendricks AM (1999) Cost analysis in the Department of Veterans Affairs: consensus and future directions. Medical Care 37(4 Suppl): AS3-AS8.

Taira DA, Seto TB, Siegrist R, Cosgrove R, Berezin R, Cohen DJ (2003) Comparison of analytic approaches for the economic evaluation of new technologies alongside multicenter clinical trials. American Heart Journal 145(3): 452-458.

Taylor A, George S, Thomas E, Smith H, Lees P, Sheron L (1998) C/PHC/01/7.95 Workload measurement in primary care using barcode technology. Final report to South and West Regional R&D Directorate. Health Care Research Unit, University of Southampton (unpublished).

Tediosi F, Bertolini G, Parazzini F, Mecca G, Garattini L (2001) Cost analysis of dialysis modalities in Italy. Health Services Management Research 14(1): 9-17.

The NHS Confederation (2006) Why we need fewer hospital beds. The NHS Confederation, London.

The Working Time Regulations 1998. Statutory instrument 1998 No. 1833. The Stationery Office, London.

Thompson SG, Nixon RM, Grieve R (2006) Addressing the issues that arise in analysing multicentre cost data, with application to a multinational study. Journal of Health Economics 25(6): 1015-1028.

Thurston NK & Libby AM (2002) A Production Function for Physician Services Revisited. The Review of Economics and Statistics 84(1): 184-191.

Townsend J, Wolke D, Hayes J, Dave S, Rogers C, Bloomfield L, Quist-Therson E, Tomlin M, Messer D (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).

Tsuchiya A & Williams A (2001) Welfare economics and economic evaluation. In Drummond M & McGuire A eds. Economic evaluation in health care. Merging theory with practice. Oxford University Press, Oxford Ch 2: 22-45.

Tsuchiya A, Brazier J, Roberts J (2006) Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. Journal of Health Economics 25(20): 334-346.

Valderrábano F, Berthoux FC, Jones EH, Mehls O (1996) Report on management of renal failure in Europe, XXV, 1994 end stage renal disease and dialysis report. The EDTA-ERA Registry. European Dialysis and Transplant Association-European Renal Association. Neph Dial Transplant 11(Suppl 1): 2-21.

Waller J (1999) Consultations in General Practice - What do they cost? Centre for Innovation in Primary Care, Sheffield.

Walsh B, Steiner A, Warr J, Sheron L, Pickering R, Southampton NLU Evaluation Team (2003) Nurse-led inpatient care: Opening the 'black box'. International Journal of Nursing Studies 40: 307-319.

Watts T, Jones M, Wainwright P, Williams A (2001) Methodologies analysing individual practice in health care: a systematic review. Journal of Advanced Nursing 35(2): 238-256.

Wheeler JR, Messana JM, Turenne MN, Hirth RA, Pozniak AS, Pan Q, Chuang CC, Slish K, Tedeschi P, Roys EC, Wolfe RA (2006) Understanding the basic case-mix adjustment for the composite rate. American Journal of Kidney Diseases 47(4): 666-671.

Whynes DK & Walker AR (1995) On Approximations in Treatment Costing. Health Economics. 4(1): 13-39.

Winkelmayer WC, Weinstein MC, Mittleman MA, Glynn RJ, Pliskin JS (2002) Health economic evaluations: The special case of end-stage renal disease treatment. Medical Decision Making 22(5): 417-430.

Wordsworth S & Ludbrook A (2005) Comparing costing results in across country economic evaluations: The use of technology specific purchasing power parities. Health Economics 14(1): 93-99.

Wordsworth S, Ludbrook A, Caskey F, MacLeod A (2005) Collecting unit cost data in multicentre studies. Creating comparable methods. The European Journal of Health Economics 6(1): 38-44.

Wright LF (1991) Survival in patients with endstage renal disease. American Journal of Kidney Diseases 17: 25-28.

Yuen P (2005) Office of Health Economics Compendium of Health Statistics 2005-2006, 17th edition. Oxford: Radcliffe Publishing.

Zorn CJW (2001) Generalized estimating equation models for correlated data: A review with applications. American Journal of Political Science 45(2): 470-490.

Zupancic JA & Richardson DK (2002) Characterization of neonatal personnel time inputs and prediction from clinical variables - a time and motion study. Journal of Perinatology. 22(8): 658-663.