Health Technology Assessment 2003; Vol. 7: No. 26

Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment

JG Williams
WY Cheung
DR Cohen
HA Hutchings
MF Longo
IT Russell

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment

JG Williams[1*]
WY Cheung[1]
DR Cohen[2]
HA Hutchings[1]
MF Longo[2]
IT Russell[3]

[1] Centre for Postgraduate Studies, The Clinical School, University of Wales Swansea, UK
[2] School of Care Sciences, University of Glamorgan, Pontypridd, UK
[3] Department of Health Sciences and Clinical Evaluation, Alcuin College, University of York, UK

* Corresponding author

Declared competing interests of authors: none

Published September 2003

This report should be referenced as follows:

Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT. Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment. *Health Technol Assess* 2003;**7**(26).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE.

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

This has meant that the HTA panels can now focus more explicitly on health technologies ('health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care) rather than settings of care. Therefore the panel structure was replaced in 2000 by three new panels: Pharmaceuticals; Therapeutic Procedures (including devices and operations); and Diagnostic Technologies and Screening.

The HTA Programme will continue to commission both primary and secondary research. The HTA Commissioning Board, supported by the National Coordinating Centre for Health Technology Assessment (NCCHTA), will consider and advise the Programme Director on the best research projects to pursue in order to address the research priorities identified by the three HTA panels.

The research reported in this monograph was funded as project number 94/06/27.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

G

Abstract

Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment

JG Williams,[1]* WY Cheung,[1] DR Cohen,[2] HA Hutchings,[1] MF Longo[2] and IT Russell[3]

[1] Centre for Postgraduate Studies, The Clinical School, University of Wales Swansea, UK

[2] School of Care Sciences, University of Glamorgan, Pontypridd, UK

[3] Department of Health Sciences and Clinical Evaluation, Alcuin College, University of York, UK

* Corresponding author

Objectives: To estimate the feasibility, utility and resource implications of electronically captured routine data for health technology assessment by randomised controlled trials (RCTs), and to recommend how routinely collected data could be made more effective for this purpose.

Data sources: Four health technology assessments that involved patients under care at five district general hospitals in the UK using four conditions from distinct classical specialties: inflammatory bowel disease, obstructive sleep apnoea, female urinary incontinence, and total knee replacement. Patient-identifiable, electronically stored routine data were sought from the administration and clinical database to provide the routine data.

Review methods: Four RCTs were replicated using routine data in place of the data already collected for the specific purpose of the assessments. This was done by modelling the research process from conception to final writing up and substituting routine for designed data activities at appropriate points. This allowed a direct comparison to be made of the costs and outcomes of the two approaches to health technology assessment. The trial designs were a two-centre randomised trial of outpatient follow-up; a single-centre randomised trial of two investigation techniques; a three-centre randomised trial of two surgical operations; and a single-centre randomised trial of perioperative anaesthetic intervention.

Results: Generally two-thirds of the research questions posed by health technology assessment through RCTs could be answered using routinely collected data. Where these questions required analysis of NHS resource use, data could usually be identified. Clinical effectiveness could also be judged, using proxy measures for quality of life, provided clinical symptoms and signs were collected in sufficient detail. Patient and professional preferences could not be identified from routine data but could be collected routinely by adapting existing instruments. Routine data were found potentially to be cheaper to extract and analyse than designed data, and they also facilitate recruitment as well as have the potential to identify patient outcomes captured in remote systems that may be missed in designed data collection. The study confirmed previous evidence that the validity of routinely collected data is suspect, particularly in systems that are not under clinical and professional control. Potential difficulties were also found in identifying, accessing and extracting data, as well as in the lack of uniformity in data structures, coding systems and definitions.

Conclusions: Routine data have the potential to support health technology assessment by RCTs. The cost of data collection and analysis is likely to fall, although further work is required to improve the validity of routine data, particularly in central returns. Better knowledge of the capability of local systems and access to the data held on them is also essential. Routinely captured clinical data have real potential to measure patient outcomes, particularly if the detail and precision of the data could be improved.

Contents

List of abbreviations and terms vii

Executive summary ix

1 Introduction 1
Background 1
Aim 1
Objectives 1
Research questions 1
Scope of routine data 1

2 Methodology 3
Overview 3
Method 3
Cost comparison 4
Qualitative issues 4

3 Exemplar A: shared care of inflammatory bowel disease 5
Background 5
Objective 5
Methods 5
Designed data 5
Routine data 5
Comparison of designed and routine datasets 12

4 Exemplar B: community diagnosis of obstructive sleep apnoea 15
Background 15
Objective 15
Methods 15
Designed data 15
Routine data 16
Comparison of designed and routine datasets 19

5 Exemplar C: a randomised study to assess and compare the clinical effectiveness of two surgical techniques for the treatment of stress urinary incontinence 21
Background 21
Objective 21
Methods 21
Designed data 21
Routine data 23
Comparison of clinical outcome using designed and routine datasets 24

6 Exemplar D: autologous blood transfusion in total knee replacement surgery 31
Background 31
Objective 31
Methods 31
Designed data 31
Routine data 33
Comparison of designed and routine datasets 38

7 Proxy outcome measures from routine data 39
Introduction 39
Methods 39
Results 48
Discussion 49
Conclusions 51

8 Health economics: resource implications of using routine data in health technology assessment 53
Aims 53
Principles 53
Method for aim 1: estimating the costs of designed data 53
Method for aim 2: estimating the costs of routine data 54
Method for aim 3: assessment of costs against effectiveness 55
Method for aim 4: estimating the marginal value of designed data 55
Results for aims 1–3: designed data versus routine data studies 56
Results for aim 4: marginal value of better studies 60
Discussion 62

9 Health technology assessment through routine data 65
Availability of routine data 66
Confidentiality of routine data 66
Validity of routine data 67
Identification of relevant systems 68
Access to identified systems 68
Case identification and matching 69
Utility of routine data 69
Potential of routine data 69
Enhancing routine data for RCTs 69

Design and conduct of health technology assessment by RCTs through routine data . 70
Conclusions and recommendations for further research 70

Acknowledgements 73

References 75

Appendix 1 Shared care of inflammatory bowel disease 79

Appendix 2 Diagnosis of obstructive sleep apnoea 87

Appendix 3 Surgery for stress incontinence 93

Appendix 4 Autologous blood transfusion 101

Appendix 5 SF-36 107

Appendix 6 The UK Inflammatory Bowel Disease (IBD) Questionnaire 111

Appendix 7 Contingent valuation exercise part 1 and 2 115

Health Technology Assessment reports published to date 119

Health Technology Assessment Programme 127

List of abbreviations and terms

AA	Automobile Association
AHI	apnoea/hypopnoea index
Analyst 1	Dr WY Cheung
Analyst 2	Dr H Hutchings
CFIS	Contracting for Information Services (UHW)
CI	confidence interval
CIS	clinical information system
CXR	chest X-ray
DC	Professor D Cohen
DT	Dr D Thomas
DVT	deep vein thrombosis
ECG	electrocardiogram
GeneCIS®	Generic Clinical Information System
GIT	gastrointestinal tract
HES	hospital episode statistics
HISS	Hospital Information Support System
HRQOL	health-related quality of life
IBD	inflammatory bowel disease
ICD-10	International Statistical Classification of Disease and Related Health Problems Version 10
IIQ	Incontinence Impact Questionnaire
ITR	Professor IT Russell
ITU	intensive therapy unit
JGW	Professor JG Williams
LOS	length of stay
MGL	Mr MG Lucas
nCPAP	nasal continuous positive airways pressure
nvCJD	New variant Creutzfeldt–Jacob disease
O/E	on examination
OPCS-4	Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures – fourth revision
OSA	obstructive sleep apnoea
PAS	Patient Administrative System (Morriston)
PE	Dr P Ebden
PEDW	Patient Episode Database (Wales)
PIMS	Patient Information Management System (Neath)
PMS	Patient Management System (Singleton)
PRCS	Postoperative red cell salvage
QALY	quality-adjusted life years
QoL	quality of life
QS1	administrative data on outpatients
RCT	randomised controlled trial
SCOPE	System Care Orientated Patient Environment (Llanelli)

List of abbreviations and terms *continued*

SF-36	Short Form-36 item health-related quality-of-life questionnaire
Theatre Man	operating theatre information system
TURP	transurethral resection of prostate
TKR	total knee replacement
UDI	Urogenital Distress Inventory
UHW	University Hospital of Wales, Cardiff
UKIBDQ	United Kingdom Inflammatory Bowel Disease Questionnaire
VAS	visual analogue scale
WTP	willingness to pay

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

Data are widely collected routinely in healthcare and increasingly held in electronic form. These data are used for a wide variety of purposes, such as health technology assessment without randomisation, although the value of this has been disputed. The randomised controlled trial (RCT) is the design of choice for health technology assessment, but data are usually collected for the sole purpose of evaluation. The value of using routinely collected data for prospective health technology assessment by RCTs has not previously been explored.

Objectives

The objectives were to estimate the feasibility, utility and resource implications of electronically captured routine data for health technology assessment by RCTs, and to recommend how routinely collected data could be made more effective for this purpose.

Methods

The project assessed the feasibility of extending the practice of health technology assessment through the use of routine data by replicating the analysis of four RCTs. The original trials were taken as designed, and the trial population as randomised. The research process was then modelled from data definition to final writing up, substituting routine for designed data activities throughout. In other words, the project simulated a novel form of health technology assessment by RCTs, using existing electronic data. The four exemplars addressed different interventions (shared care for inflammatory bowel disease, home assessment of obstructive sleep apnoea, urethral sling surgery for female urinary incontinence, and autologous blood transfusion during total knee replacement). For each of these four RCTs, two analyses were undertaken, one using designed data and the other routine data. The analyses were carried out independently before discussion and reconciliation of the findings. This led to conclusions about the feasibility, validity, utility and cost of using routine data for health technology assessment.

Results

The study has shown that some of the research questions posed by health technology assessment through RCTs can indeed be answered using routinely collected data. Where these questions require analysis of NHS resource use, data can usually be identified. Clinical effectiveness can also be judged, using proxy measures for quality of life (QoL), provided clinical symptoms and signs are collected in sufficient detail. Patient and professional preferences cannot be identified from routine data but could be collected routinely by adapting existing instruments.

Routine data are potentially cheaper to extract and analyse than designed data. In addition, they facilitate recruitment. They also have the potential to identify patient outcomes captured in remote systems that may be missed in designed data collection.

Notwithstanding these potential benefits, the study confirmed previous evidence that the validity of routinely collected data is suspect, particularly in systems that are not under clinical and professional control. There are also potential difficulties in identifying, accessing and extracting data, and in the lack of uniformity in data structures, coding systems and definitions. While data validity remains suspect there is likely to be resistance among researchers to the use of routine data for health technology assessment by RCTs.

Conclusions

Routine data have the potential to support health technology assessment by RCTs. The cost of data collection and analysis is likely to fall, although further work is required to improve the validity of routine data, particularly in central returns. Better knowledge of the capability of local systems and access to the data held on them is also essential.

Routinely captured clinical data have real potential to measure patient outcomes, if the data were collected in detail and with precision.

Research recommendations

There is a need for further research to:

- test prospectively the feasibility of health technology assessment by RCTs through routine data
- classify the research data needed for health technology assessment, and to map these data to potential routine sources
- assess the feasibility, cost and effects of greater clinical ownership and responsibility for hospital episode statistics
- explore the feasibility and cost of local information laboratories aimed at maximising access to, and the utility of, routine data
- understand and change clinicians' and researchers' attitudes to routine data, particularly as validity and availability improves
- define standards to ensure the uniformity and validity of data collected by different local and national systems
- explore the use of surrogate clinical data for measuring patient-focused outcomes
- explore the feasibility and cost of routine completion of health-related QoL questionnaires in clinical practice
- explore the feasibility and cost of routine capture of patient preference data.

Chapter 1
Introduction

Background

This study focused on the feasibility, utility and cost of using electronically stored routine data to undertake or support the assessment through randomised controlled trials (RCTs) of health technologies in clinical settings. McKee identified routine data as those data whose primary reason for collection is other than audit (or, by implication, research);[1] such data include hospital episode statistics (HES) held in central returns,[2] data from hospital and community information systems;[3] cancer registries; systems established to manage specific programmes, such as breast or cervical screening; radiology, pathology, pharmacy and accident and emergency systems; administrative and patient-focused demographic and clinical data stored on local systems such as Patient Administrative System (PAS), Hospital Information Support System (HISS) and departmental clinical systems.[4]

In the UK, substantial work has been done using routine data linked to clinical data to detect geographical variations in perinatal, infant and childhood mortality;[5,6] prevalence of low birth weight;[7] mean height of school children;[8] rates of accidental injury;[9] respiratory illness;[10] and cancer incidence and mortality.[11] It has been argued that routinely available activity data in primary and secondary care could be used to monitor and promote equity of service utilisation and informed contracting.[12] There are also examples of using routine data to evaluate quality of care, for example in health surveillance of preschool children,[13] obstetric practice;[3] and various acute specialities including general medicine, general surgery, gynaecology, trauma, orthopaedics and geriatrics.[14]

Large databases of routinely collected data have been used in non-randomised retrospective studies to compare transurethral resection of prostate (TURP) and open prostatectomy,[15] and to study the appropriateness of cardiovascular procedures[16] and cholecystectomy.[17] There has been interest in the use of large databases for technology and quality assessment without randomisation,[18,19] but the value of this approach has been disputed.[20]

Aim

The primary aim of the study was to explore whether data routinely collected in electronic form could reliably be used as a basis for health technology assessment by RCTs, and whether and how such data could be enhanced for this purpose.

Four examplar RCTs were selected that addressed different interventions:

- shared care for inflammatory bowel disease
- home assessment of obstructive sleep apnoea
- urethral sling surgery for female urinary incontinence
- autologous blood transfusion during total knee replacement (TKR).

Objectives

- To estimate the feasibility, utility and resource implications of using electronically captured routine data for health technology assessment by RCTs.
- To recommend how routinely collected data could be made more effective for this purpose.

Research questions

The following questions will be addressed in this study:

- How far can the objectives of each of the four exemplar health technology assessments be met by analysis of routinely collected data?
- Where all the objectives of the health technology assessment cannot be met by routine data, what conclusions can be drawn?
- Could the shortfall have been met by enhancing or augmenting the routine data?
- What are the cost implications of using routinely collected rather than purpose-designed data for health technology assessment?

Scope of routine data

All electronically stored patient-specific data, collected as part of the process of delivery of

TABLE 1 *Overview of location and scope of routinely collected data*

Database	Where located	Data collected
PEDW	National Assembly for Wales	Administrative, demographic data, diagnoses and procedures on inpatient and day cases
QS1	National Assembly for Wales	Administrative data on outpatients
PAS (e.g. SCOPE, PMS, PIMS)	Trust	Administrative, demographic and some clinical data
CIS (e.g. GeneCIS)	Clinical department	Administrative, demographic and detailed clinical data on all patient contacts in hospital
Pathology	Pathology	Demographic and pathological data
Radiology	Radiology	Demographic data and imaging reports
Laboratory	Laboratory	Demographic data and test results
Theatre	Operating theatre	Demographic, administrative and procedure data
Casemix (e.g. CFIS)	Trust	Administrative, demographic, financial and some clinical data
GP	Primary care	Demographic, administrative and clinical data

SCOPE: System Care Orientated Patient Environment; PMS: Patient Management System; PIMS: Patient Information Management System; CFIS: Contracting for Information Services.

healthcare, are included as falling within the definition of routine data. Some of these data are collected by provider organisations, specifically for central returns or contracting, while other data are collected by healthcare professionals or clinical teams to inform the delivery of care to individual patients (*Table 1*). In Wales, routine data are forwarded from all hospitals to the Health Care Management Information Services at the Welsh Health Common Services Authority in Cardiff (now Health Solution Wales). By this process individual data are collected on more than 500,000 patients using hospital beds each year, and held on the all-Wales inpatient and day patient database [Patient Episode Data for Wales (PEDW)]. More limited data are routinely collected on outpatient attendance (QS1). Within hospitals detailed data are collected on patient administration (PAS) or case-mix management systems, and used for internal management and informing and monitoring contracts with commissioners (and, previously, fund-holding general practices). Many clinical teams collect data specifically for the purposes of individual patient management using clinical information systems. GeneCIS® is a generic clinical information and management system[21] that supports gastroenterology at Neath and urology at Morriston. The information gained has been used to monitor the quality and quantity of services and to facilitate new clinical developments.[22] Data are also held on databases that support service departments such as radiology, pathology laboratories and theatres.

At the time of this study there were no routinely collected central returns on activity or morbidity in primary care, which have been addressed elsewhere in specific studies using sentinel practices, such as the national studies of morbidity in general practice.[23] Fortunately, there are indications that diagnostic data held on general practitioner (GP) information systems are generally valid.[24] A project at the National Assembly for Wales is piloting the routine extraction of data from such GP systems and initial results suggest that this may be a very useful source of valid information.

Chapter 2
Methodology

Overview

The project assessed the feasibility of extending the practice of health technology assessment through the use of routine data by replicating four RCTs using routine data in place of the data already collected for the specific purpose of the assessments. This was done by modelling the research process from conception to final writing up and substituting routine for designed data activities at appropriate points. This allowed a direct comparison to be made of the costs and outcomes of the two approaches to health technology assessment.

The four health technology assessments involved patients under care at five district general hospitals (Morriston, Llanelli, Singleton, Neath and the University Hospital of Wales, Cardiff) involving four distinct clinical disciplines: gastroenterology, respiratory medicine, urology and anaesthesia. The trial designs were a two-centre randomised trial of outpatient follow-up;[25,26] a single-centre randomised trial of two investigation techniques;[27] a three-centre randomised trial of two surgical operations;[28] and a single-centre randomised trial of perioperative anaesthetic intervention.[29,30] The exemplars are described in more detail in following chapters, and in Appendices 1–4.

For each of these four health technology assessments, two analyses were undertaken, one using designed data and the other using routine data. The analyses were undertaken by two statisticians working independently, as shown in *Table 2*.

For the purpose of this study the analysts working on the designed dataset revisited the analysis they had undertaken for the original trial. Thereafter, careful comparison of the four analyses using routine data with the corresponding analyses using designed data led to general conclusions about the feasibility, validity, utility and cost of using routine data for health technology assessment.

Method

The first stage of the investigation was to review each original study protocol, and to identify appropriate sources of routinely collected electronically stored data for analysis. Sources of data were identified by the secondary analysts, in discussion with members of the project team. Items that could be used as proxies for outcome or process variables or as possible covariates were identified. Within each exemplar, analysis proceeded according to the general principles appropriate to the particular design and the specific procedures set out in the relevant protocol. Each statistician discussed the principles of these analyses with the project leader [JG Williams (JGW)], the relevant clinician [JGW, MG Lucas (MGL), D Thomas (DT) or P Ebden (PE)] and the consultant statistician [IT Russell (ITR)] at the beginning of each analysis and as necessary thereafter. However, no details were exchanged for any given health technology assessment until each of the analyses for that health technology assessment had been deemed complete by the relevant statistician. The studies continued as planned, using designed datasets, but were analysed independently using the routine datasets. Responsibility for the analyses by the two statisticians is shown in *Table 2*.

This approach reduced the possibility of analytical bias caused by preconceptions and insight gained from undertaking the same study by different methods. A common analytical strategy developed from the protocols of the exemplars and refined by the consultant statistician (ITR) was applied to both designed and routine data, to reduce the effect of differences in approach between the two analysts.

As soon as both analyses were complete for a given exemplar, they were rigorously compared by the two statisticians, under the direction of ITR, with support from the relevant clinician and the project manager. One key feature of this task was the comparison of each outcome and process variable in the designed dataset with the nearest proxy in the routine dataset. By pooling these two datasets it was often possible to identify whether a simple enhancement or augmentation of the routine data in the direction of the designed data could increase the validity and utility of the routine data for the purpose of HTA. Thus, the analysis of the designed dataset was regarded as the 'gold

TABLE 2 *Summary of exemplars*

Study	Technology	Condition	Design	Sample size	Analysis of designed HTA	Development and analysis of routine HTA	Appendix
A	Shared care	Inflammatory bowel disease	Two-centre randomised trial	180	Analyst 1	Analyst 2	1
B	Home assessment	Obstructive sleep apnoea	Single-centre randomised trial	102	Analyst 2	Analyst 1	2
C	Urethral sling	Female urinary incontinence	Three-centre randomised trial	165	Analyst 1	Analyst 2	3
D	Autologous blood transfusion	Total knee replacement	Single-centred randomised trial	231	Analyst 2	Analyst 1	4

HTA: health technology assessment; Analyst 1: Dr WY Cheung; Analyst 2: Dr H Hutchings.

standard'. Repeating this process for each of the four exemplars generated general conclusions to add to the specific conclusions of each. The results of the analysis for each exemplar are given in Chapters 3–6.

Cost comparison

This study aimed to estimate the difference in research costs between the original (designed data) studies and their routine data counterparts. It was thus solely concerned with those cost elements that vary with the type of data used. All other costs of conducting the studies have been excluded. For the routine data studies, the time taken spent on the various activities by those involved in the studies was monitored directly.

However, when the present study was conceived, the four exemplar (designed data) studies were at various stages of completion. Staff employed on the projects were undertaking a wide variety of activities, only some of which would be affected if a routine data approach had been taken instead. At that time there was no reason for them to monitor how much of their time was being spent on different activities. To disentangle the costs of the relevant activities, staff who had worked on each study were asked to estimate the proportion of total project time they had devoted to a list of activities which varied with the type of data used. The cost of each activity was estimated by dividing the total funding which had been secured for each relevant researcher using these proportions. These proportions were used to estimate the hours apportioned to various activities, using the total funded time as the base. Such a retrospective 'top-down' estimate is inevitably crude. Moreover, the total cost apportioned to the various activities did not include the time of some staff who contributed to the projects but who were funded from other sources. Accordingly, to reflect the true cost of the designed data studies (i.e. the value of the resources used), the time of these non-funded staff was added where appropriate.

In the case of one designed data study, however, a detailed record of the time spent on different activities had been kept. This allowed a more accurate 'bottom-up' approach to be used in this case. Details of the methods used are given in Chapter 8.

Qualitative issues

Issues encountered in the identification of data sources, extraction of data and questions of validity were documented throughout the study and collated. These are discussed in Chapter 9.

On the basis of their experience, the authors explored the process whereby RCTs can be designed for, and conducted using, routine data. This is also discussed in Chapter 9.

Chapter 3

Exemplar A: shared care of inflammatory bowel disease[25,26]

Background

The routine outpatient follow-up of patients with chronic relapsing disease leads to a rising workload and consultations that are often unnecessary or inappropriate. An open-access system of follow-up offers potential advantages. This was evaluated for patients with inflammatory bowel disease (IBD).

Objective

To evaluate whether open-access follow-up of patients with IBD is better than routine, booked appointments.

Methods

Design
Pragmatic two-centre RCT.

Setting
Two district general hospitals in Swansea and Neath, south-west Wales, UK.

Subjects
One hundred and eighty adult patients (78 Crohn's disease, 77 ulcerative or indeterminate colitis, 25 ulcerative or idiopathic proctitis) recruited from outpatient clinics between October 1995 and November 1996.

Intervention
Open-access follow-up according to patient need.

Control
Routine outpatient appointments at intervals determined by the physician at consultation in the clinic.

Main outcome measures
Generic [Short Form-36 item health-related quality of-life questionnaire (SF-36)] and disease-specific [United Kingdom Inflammatory Bowel Disease Questionnaire (UKIBDQ)] quality of life (QoL) measured at 6-monthly intervals for 2 years, number of primary and secondary care contacts, total resource use and views of patients and GPs.

Designed data

Sources of designed data
Data on health-related quality of life (HRQoL) were collected using validated questionnaires completed by patients at entry and every 6 months for 2 years. Contacts and resource use were extracted from patient records, both electronic and paper. This was undertaken by the research team in secondary care, and by postal request to GPs for primary care. Patient and practitioner views and patient-borne costs were obtained by postal questionnaire at the end of the study. Some GPs were also interviewed.

Analysis of designed data
Analysis was by intention to follow-up. To counteract the effect of possible differences in baseline HRQoL scores, changes in individual QoL scores from baseline were analysed using *t* tests. Numeric data were analysed using paired *t* tests. Because of skewness, resource data were bootstrapped (1000 replications).[31] Preference data were analysed by chi-squared tests.

Results of analysis of designed data
There were no significant differences in generic or disease-specific QoL. Open-access patients had fewer day-case visits ($p < 0.05$) but there were no statistically significant differences in any other resource use variables, patient-borne costs or total societal costs. GPs ($p < 0.001$) and their patients ($p < 0.001$) both preferred open access.

Conclusions from designed data
Open-access follow-up delivers the same quality of care as routine outpatient care, and is preferred by patients and GPs. Both methods of follow-up had broadly similar overall costs. Better methods of ensuring urgent access to outpatient clinics are needed.

Routine data

Sources of routine data
Patient-identifiable, electronically stored routine data were sought from the PEDW, hospital

TABLE 3 *Nature and sources of routine data*

Designed data	Routine data	Source of routine data
Study start date	As designed study	Original study recruitment
Study end date	As designed study	Original study recruitment
Randomisation code	As designed study	Original study recruitment
Postcode	As designed study	Original study recruitment
Diagnostic group	As designed study	Original study recruitment
Centre	As designed study	Original study recruitment
GP code	As designed study	Original study recruitment
GP consultations at the surgery	As designed study	GP data
GP home visits	As designed study	GP data
Drug usage (frequency, dose and specific drug)	As designed study	GP data
Outpatient visits: IBD related	As designed study	GP data, GeneCIS
Day-case visits: IBD related	As designed study	GP data, GeneCIS
Inpatient stay: IBD related	As designed study	GP data, PEDW
Number and type of test or investigation	As designed study	GP data, PEDW, GeneCIS, radiology, pathology, Theatre Man, PAS
Cost of drugs	As designed study	GP data
Cost of consultations at the surgery	As designed study	GP data
Cost of GP home visits	As designed study	GP data
Cost of outpatient visits: IBD related	As designed study	GP data
Cost of inpatient visits: IBD related	As designed study	GP data
Cost of tests and investigations	As designed study	GP data
Total cost of primary care	As designed study	GP data
Total cost of secondary care	As designed study	GP data
Baseline UKIBDQ score	Surrogate UKIBDQ score	GeneCIS
24 month UKIBDQ score	Surrogate UKIBDQ score	GeneCIS
Baseline SF-36 score	Surrogate SF-36 score	GeneCIS
24 month SF-36 score	Surrogate SF-36 score	GeneCIS
Mileage cost to GP's surgery	Imputed mileage cost from home to GP's surgery	AA Route Finder
Mileage cost to hospital	Imputed mileage cost from home to hospital	AA Route Finder
Mileage cost for outpatient appointments	Imputed mileage cost from home to outpatient appointments	AA Route Finder
Mileage cost for-day case visits	Imputed mileage cost from home to day-case visits	AA Route Finder
Mileage cost for inpatient visits	Imputed mileage cost from home to inpatient visits	AA Route Finder

AA: Automobile Association.

administrative systems, GP computer systems, radiology systems, pathology systems and departmental clinical systems.

Each data source was queried to extract all the data for each of the study patients from the date of recruitment to the study end date (at 24 months). *Table 3* lists the variables analysed in the designed study, the variable or surrogate used from routine data sources, and the sources used. Electronically stored data were extracted according to specified matching criteria, by information managers or technicians with experience of the individual systems. GeneCIS contained a written description of the clinical code, so little interpretation was required. The International Statistical Classification of Disease and Related Health Problems Version 10 (ICD-10)[32] and the Office of Population Censuses and Surveys Classification of Surgical Operations and Procedures – fourth revision (OPCS-4)[33] coding manuals were used to interpret the clinical diagnostic and procedure data held on PEDW and PAS.

Analysis of routine data

Comparisons were made between the open-access and routine groups at study recruitment and at 24 months. Clinical effectiveness was determined by comparing the presence of individual and combinations of relevant symptoms, signs or diagnoses, and costs determined by comparing primary and secondary resource use and patient-borne costs.

A comparison of all the available data was made at baseline and 24 months to determine whether there were any differences in clinical effectiveness, resource use and patient and GP preferences between the open-access and routine groups. Categorical data were analysed statistically using the chi-squared test and numerical data by the two-sample *t* test or the Mann–Whitney *U* test. Resource data were bootstrapped if the distribution was skewed.

To compare clinical effectiveness in terms of changes in health status, the UKIBDQ and SF-36 questionnaires were used as templates to determine which symptoms, signs or diagnoses were relevant to IBD. The presence of each comparable health characteristic was compared between the two groups at baseline and 24 months. This is described in more detail in Chapter 7.

Resource use was determined by directly comparing the number of investigations, blood tests, procedures, biopsies, drug use, personnel seen, and hospital and GP visits. The total NHS cost for primary and secondary care was also compared between the two groups. The patient-borne costs were determined by the total number of GP and hospital visits, using AA standard mileage costs based on the distance travelled.

Each of the systems available was examined to determine whether comments were held that indicated the type of care preferred by patient and GP. For example, GeneCIS offers a facility to insert free text, but this had not been used to record patient preferences in a structured way.

Availability and validity of routine data

The system that contained the most useful data for assessing changes in health status was the departmental clinical system GeneCIS. This system was in place at one of the study sites (Neath) and contained data on just over half of the original group of patients. At each hospital visit, relevant symptoms, signs and diagnoses were noted in the system and a comparison of relevant symptoms could be compared at baseline and 24 months. PEDW data contained diagnostic and procedure codes, but these only related to 53 of the original sample and the dates were often not comparable with the original baseline and 24-month dates. The hospital information systems also contained diagnostic and procedure codes but these were available only for some of the original sample from Morriston. Of the 81 patients, only 23 had codes listed on the system, making the numbers too small to make any meaningful comparisons. All the systems available contained information that could be used to compare NHS resource use and patient-borne costs.

Table 4 shows the percentage of study patients identified with at least one episode of care by each data source.

Results of analysis of effectiveness using routine data

No significant differences existed between the two groups at baseline in relevant symptoms, signs or diagnoses, with the exception of urgency. There was a significantly higher number of control patients with this symptom [$p < 0.05$, 95% confidence interval (CI) of the difference, –0.15 to –0.006]. The compiled composite UKIBDQ score showed no difference between the two groups. There were also found to be no differences in the numbers of patients with comparable SF-36 symptoms between the two groups.

At 24 months the only symptom that reached significance was the presence of abdominal pain,

TABLE 4 *Validity of routine data*

Data source	No. (%) of study patients identified (*n* = 180 overall: 99 Neath; 81 Morriston)	Comment
GP data systems (various at both sites)	155 (86)	Held on computer by general practices
PEDW	52 (29)	IBD study involved outpatient appointments, which are not routinely documented on PEDW
Neath GeneCIS	99 (100)	Clinical system only available at Neath
Neath PIMS	0 (0)	Change of computer system in progress. Department could not provide data in time for deadline
Neath Pathology	99 (100)	Had to print out results; could not save to disk
Neath Theatre Man	98 (99)	Administrative and procedure data only
Neath Radiology	97 (98)	Radiology system only available at Neath
Morriston PAS	81 (100)	Only 28% of patients had diagnostic and procedure codes documented
Morriston Pathology	81 (100)	Had to print out results; could not save to disk
Morriston Theatre Man	0 (0)	System was down for a prolonged period when data extraction was requested

with more open-access patients exhibiting this symptom ($p < 0.01$; 95% CI of the difference, 0.05 to 0.45). There were no differences in any of the other comparable IBD symptoms or in the composite IBD score. There was a significant difference in the general health of the patients at 24 months. A greater number of open-access patients had a symptom code of condition improved compared with those in the routine group ($p < 0.05$; 95% CI of the difference, 0.33 to 0.03). There were no other differences in surrogate SF-36 symptoms between the two groups at 24 months.

There were no differences in the number of consultations with the GP or the practice nurse and no differences between the two groups in the reasons for GP visits or the number requiring home visits. Drug use was similar for the two groups. Although the total NHS cost per patient had a tendency to be higher in the open-access group, this did not reach significance. There was no significant difference between the two groups in the total NHS cost for primary care.

There was no difference in the number of hospital outpatient or inpatient visits between the two groups for IBD-related or non-IBD-related conditions. Patients in the control group were admitted as day cases more frequently than those in the open-access group ($p < 0.05$; CI of the difference, –0.57 to –0.14). There were no differences in the type of personnel seen in the hospital, with the exception that more control patients were seen by a consultant ($p < 0.05$; CI of the difference, –0.02 to –0.23). There were no differences in the number of tests, procedures, investigations or biopsies performed in each group. There was no significant difference in the total NHS cost for secondary care between the two groups.

The patient travel costs for primary and secondary care (inpatient and outpatient) visits showed no significant differences between the two groups. The control group, however, incurred a significantly greater cost for day-case hospital visits ($p < 0.01$; CI of the difference, –208.1 to –31.9).

Discussion of routine data analysis

There were no major differences in the reported symptoms, signs or diagnoses at baseline and 24 months, with the exception that more open-access patients had an improvement in their condition at 24 months. This result should be interpreted cautiously, however, since the number of cases was small and the information was only available on just over half of the original group of patients. Complete UKIBDQ (total and subscale) scores and SF-36 (subscale) scores could not be compiled since comparable items were only found for a proportion of the questions from the original questionnaires. The composite surrogate SF-36 and UKIBDQ scores showed no difference between the two groups.

TABLE 5(a) *Differences in health outcomes between baseline and 24 months: comparison between designed and routine data*

Designed data[a] (*n*)	**Mean difference between groups (95% CI)**	**Routine data: surrogate variable from GeneCIS (*n* = 97)[b]**	**Mean difference between groups (95% CI)**
UKIBDQ Bowel Function 1 (160)	0.27 (–7.12 to 7.67)	Diarrhoea, incontinent of faeces, urgency	0.11 (–0.13 to 0.36)
UKIBDQ Bowel Function 2 (160)	–3.46 (–10.88 to 3.96)	Abdominal tenderness, site of GIT pain, epigastric pain, abdominal cramps, griping pain, colicky abdominal pain, upper abdominal pain, lower abdominal pain, generalised pain, wind, abdomen feels distended, abdomen feels swollen, abdominal discomfort, abdominal swelling, nausea, vomiting	0.12 (–0.21 to 0.44)
UKIBDQ Emotional Function (160)	–1.28 (–5.91 to 3.36)	O/E depressed, O/E nervous, O/E anxious, person with feared complaint, anxiousness, anxiety state unspecified, irritable, O/E nervous	0.02 (–0.20 to 0.24)
UKIBDQ Social Function (160)	0.40 (–5.71 to 6.50)	No surrogate available	
UKIBDQ Systemic Function (159)	2.16 (–5.56 to 9.90)	Tired all the time, fatigue, tiredness symptom, O/E looks ill, non-organic sleep disorder, insomnia, appetite symptom	0.03 (–0.16 to 0.22)
SF-36 Vitality (163)	–3.72 (–10.72 to 3.27)	Tired all the time, fatigue, tiredness symptom	–0.01 (–0.17 to 0.14)
SF-36 General Health (160)	–3.48 (–8.92 to 1.97)	Feels well, feels ill	–0.02 (–0.61 to 0.57)
SF-36 Mental Health (163)	–3.74 (–9.88 to 2.39)	O/E depressed, depressed, depressive disorder	0.08 (–0.08 to 0.25)
SF-36 Bodily Pain (163)	–2.50 (–10.35 to 5.03)	Central abdominal pain, colicky pain, epigastric pain, generalised abdominal pain, lower abdominal pain, right flank pain, right iliac fossa pain, suprapubic pain, upper abdominal pain, cervalgia, O/E tenderness/pain, O/E in pain, site of GIT pain, complains of a pain, griping pain, pain in limb	0.13 (–0.10 to 0.36)
SF-36 Physical Function (162)	–3.66 (–10.55 to 3.24)	No surrogate available	
SF-36 Role Physical (157)	–2.67 (–16.8 to 11.4)	No surrogate available	
SF-36 Social Function (163)	–0.39 (–8.89 to 8.10)	O/E anxious, O/E nervous, person with a feared complaint, anxiousness, anxiety state unspecified	0.00 (–0.08 to 0.08)
SF-36 Role Emotional (154)	–5.29 (–22.0 to 11.4)	No surrogate available	

[a] Different sample sizes for different variables in the designed data study because of missing items.
[b] Sample size for GeneCIS is less than in Table 4 as two patients withdrew during the study.
GIT: gastrointestinal tract; O/E: on examination.

TABLE 5(b) *Resource use in secondary care: comparison between designed and routine data with bootstrapping*[a]

	Mean difference between groups (95% CI)			
	Designed data (*n* = 178)	**Source**	**Routine data**	**Source[b] (*n*)**
Resource variable				
Outpatient visits IBD-related problems	–0.73 (–1.50 to .21)	Notes/PAS	–0.62 (–1.68 to 0.40) –0.75 (–2.04 to 0.29) –0.25 (–1.53 to 1.33)	GP (145) Hospital (81) GeneCIS (97)
Inpatient days IBD-related problems	0.40 (–0.51 to 1.74)	Notes/PAS	0.34 (–0.06 to 1.52) 0.50 (–1.53 to 2.87)	GP (140) Hospital (81)
Day-case visits IBD-related problems, mainly investigations	–0.18 (–0.34 to –0.03)	Notes/PAS	–0.32 (–0.57 to –0.14) –0.09 (–0.34 to 0.13) –0.38 (–0.81 to –0.01)	GP (145) Hospital (81) GeneCIS (97)
Total investigations and tests	–1.15 (–3.44 to 0.89)	Pathology	–1.38 (–4.22 to 1.92)	GeneCIS (97)
Full blood count	–0.28 (–0.96 to 0.46)	Pathology	0.48 (–0.24 to 1.07) 0.04 (–1.46 to 1.37) –0.27 (–1.09 to 0.68)	GP (153) Hospital (81) GeneCIS (97)
Erythrocyte sedimentation rate and C-reactive protein	–0.35 (–1.09 to 0.18)	Pathology	–0.06 (–0.39 to 0.23) 0.27 (–0.82 to 1.15) –0.25 (–1.00 to 0.68)	GP (153) Hospital (81) GeneCIS (97)
Biochemical profile	–0.27 (–1.01 to 0.79)	Pathology	0.30 (–0.07 to 0.79) 0.02 (–1.49 to 2.23)	GP (153) Hospital (81)
Colonoscopy	–0.16 (–0.30 to –0.03)	Notes/PAS	–0.09 (–0.25 to 0.03) –0.12 (–0.27 to 0.03)	GP (153) Hospital (81)
Rigid sigmoidoscopy	–0.03 (–0.14 to 0.09)	Notes/PAS	–0.002 (–0.08 to 0.12) –0.02 (–0.15 to 0)	GP (153) Hospital (81)
Flexible sigmoidoscopy	0.02 (–0.17 to 0.20)	Notes/PAS	–0.01 (–0.09 to 0.05) –0.05 (–0.34 to 0.35)	GP (153) Hospital (81)
Biopsy	–0.01 (–0.16 to 0.18)	Pathology	–0.01 (–0.07 to 0.04) 0.02 (–0.29 to 0.42) –0.24 (–0.98 to 0.42)	GP (153) Hospital (81) GeneCIS (97)
Vitamin B_{12}, folate and ferritin	–0.06 (–0.35 to 0.22)	Pathology	0.03 (–0.33 to 0.43) 0.02 (–0.34 to 0.41)	GP (153) Hospital (81)
Costs				
Outpatient costs IBD-related problems (£)	–46.1 (–94.76 to 13.03)	Hospital	–39.32 (–105.88 to 25.06) –47.26 (–128.31 to 18.32)	GP (145) Hospital (81)
Inpatient costs IBD-related problems (£)	59.3 (–75.93 to 261.46)	Hospital	51.14 (–9.56 to 227.89) 74.25 (–229.55 to 430.50)	GP (140) Hospital (81)
Cost of tests and investigations (£)	–57.43 (–135.68 to 28.60)	Pathology	–28.39 (–83.46 to 18.73) –52.72 (–174.24 to 78.57)	GP (153) Hospital (81)
Total NHS cost in secondary care (£)	–44.23 (–239.43 to 213.53)	Pathology + hospital	–21.37 (–134.61 to 172.11) –25.73 (–442.71 to 391.04)	GP (140) Hospital (81)
Patient-borne costs secondary care (£)	–15.26 (–30.38 to 1.51)	Patient reported	–1.63 (–9.36 to 5.01)	AA + GP (139)

[a] Original analysis of the designed data (as reported in the *British Medical Journal*[25] and Final Report[26] of the study) used non-parametric tests for resource use in the absence of software for bootstrapping.
[b] Different sample sizes for routine data from GP systems because of missing items.

TABLE 5(c) *Resource use in primary care: comparison between designed and routine data with bootstrapping*

	Mean difference between groups (95% CI)			
	Designed data ***n* = 155**	**Source**	**Routine data** ***n* = 155**	**Source**
Resource variable				
No. of GP consultations in surgery	0.67 (–1.90 to 2.73)	GP	Data and source as designed data	
No. of GP home visits	–0.03 (–0.40 to 0.36)	GP	Data and source as designed data	
Drugs (no. of tablets/enemas)				
Immunosuppressive	87.02 (–129.07 to 239.88)	GP	Data and source as designed data	
Maintenance	426.4 (–9.25 to 824.51)	GP	Data and source as designed data	
Antidiarrhoeal	–7.63 (–403.28 to 212.36)	GP	Data and source as designed data	
Steroid enema	–1.03 (–12.97 to 8.05)	GP	Data and source as designed data	
Costs				
GP consultations (£)	5.92 (–13.21 to 20.88)	GP	Data and source as designed data	
Home visits (£)	–1.34 (–20.23 to 18.61)	GP	Data and source as designed data	
Total cost of drugs (£)	116.6 (–28.77 to 243.65)	GP	Data and source as designed data	
Total NHS costs in primary care (£)	121.2 (–39.32 to 252.75)	GP	Data and source as designed data	
Patient-borne costs for primary care (£)	11.18 (–1.61 to 23.92)	Patient diary	1.99 (–7.23 to 16.32)	GP + AA

TABLE 5(d) *Total societal costs: comparison between designed and routine data with bootstrapping*

	Mean difference between groups (95% CI)	
Resource variable	**Designed data**	**Routine data**
Total NHS costs (£)	61.61 (–189.11 to 331.45) (*n* = 156)	103.1 (–114.64 to 323.76) (*n* = 139)
Total patient-borne costs (£)	–5.63 (–31.40 to 19.75) (*n* = 156)	0.97 (–13.67 to 21.69) (*n* = 115)
Total societal costs (£)	55.98 (–208.43 to 344.83) (*n* = 156)	104.9 (–107.31 to 355.82) (*n* = 139)

The only data source that provided information on health change over the period of the study was the clinical information system GeneCIS. This system contained information regarding symptoms, signs and diagnoses, and could be useful in comparing health change in future studies. It is routinely used at one of the study sites and it could be adapted slightly to ensure that more detailed information was collected. If such a system was implemented widely it could provide a large amount of useful data for research purposes without the need for purposely collecting data. One shortfall of this system is that there may be an element of over-reporting of symptoms, signs and diagnoses. Fortunately, the system requires that symptoms, signs and diagnoses be removed on subsequent visits if they are no longer present.

All the routine data systems available provided some information on resource use and patient-borne costs. These data highlighted that the type of follow-up had no major impact on NHS resource use in primary and secondary care. Using AA standard mileage costs in conjunction with information on frequency of GP and hospital visits, it was found that patient-borne costs were similar in the two groups. Costs of new interventions or treatments can have an impact on their implementation. The routine data available were useful in determining resource use and could reduce the need to collect data on costs.

None of the systems available had any information concerning the preferences of GPs or patients on the type of care. So, routine data were not able to answer this question and purpose-designed data collection would be needed.

Thus, the available routine data system could answer two out of the three original research

TABLE 6 *Reconciliation by research question*

Research question	Definitive conclusions from analysis of designed data	Conclusions from analysis of routine data	Comments
To evaluate the clinical effectiveness in terms of changes in health status (using generic SF-36 and disease-specific UKIBDQ QoL questionnaires)	No difference in generic or disease-specific quality of life	No difference in displayed individual symptoms, signs or diagnoses or composite surrogate UKIBDQ and SF-36 scores. More open-access patients were given a diagnostic code of 'symptom improved' at 24 months	1. Doctors may over-report symptoms, signs and diagnoses because of the nature of the clinical information system (GeneCIS). Symptoms, signs and diagnoses must be removed on subsequent visits if they are no longer present 2. Data only available on the Neath sample of patients 3. Only a proportion of the symptoms, signs and diagnoses were comparable with the SF-36 and UKIBDQ questions.
Evaluation of NHS resource use and patient-borne costs	Open access had fewer day visits but there were no significant differences in any other resource use variables or in patient-borne costs	Open access had fewer day visits but there were no significant differences in any other resource use variables or in patient-borne costs.	1. Some difficulties in obtaining all the information from the available hospital systems (slow retrieval rate by hospital personnel, computer failure) 2. Some hospital data only available for a proportion of the study patients 3. Routine data systems provided a large amount of data for comparison of resource costs and patient-borne costs
Evaluation of patient and GP preferences (from designed questionnaires)	GPs and patients preferred open-access appointments	Unable to draw any conclusions	No routine data available to compare patient and GP preferences on any of the available electronic systems

questions: those relating to effectiveness and cost but not individual preferences were answerable to a great degree using the available routine data systems. Wider implementation of clinical information systems along with expansion of data collected would expand the function of such systems in answering specific research questions. During the course of data collection there were some difficulties in retrieving data, due either to pressures on personnel or to system failures. Prospective discussions with relevant data personnel would also aid the collection of such material.

Conclusions from routine data

Using the electronic routine data available there were found to be no major differences between the two groups in clinical effectiveness, resource use or patient-borne costs. GP and patient preferences could not be identified.

Comparison of designed and routine datasets

Table 5(a) shows remarkable consistency in the estimated changes in health outcome over 24 months whether they are derived from designed data or from the GeneCIS system. That *Tables 5(b, c)* show greater consistency is even less surprising given that the source of data is the same. However, the sample sizes are smaller for routine data because these data are not all accessible.

Table 6 shows that routine data found no significant differences in QoL of patients at

baseline or 24 months between the two study groups. NHS resource use and patient-borne costs were similar, apart from routine patients having more day visits. However, routine data were unable to draw any conclusions about patients' or GPs' preferred type of appointment. None of the routine data sources available had recorded any qualitative data that allowed such a comparison to be made.

One of the systems (GeneCIS) does have the provision for recording such information, although it does not do so routinely. Although the recording of such information routinely may incur additional costs, it would have the benefit of providing detailed qualitative information that may reduce the need for purpose-designed data.

Broadly speaking, the conclusions drawn were similar to the results from the primary analysis. In summary, the routine data could answer two of the three original research questions to a large degree.

Chapter 4

Exemplar B: community diagnosis of obstructive sleep apnoea[27]

Background

The number of patients referred for evaluation with suspected obstructive sleep apnoea (OSA) is increasing, resulting in a growing demand for diagnostic studies. The majority of centres lack the time, money and experienced staff to perform full polysomnography, so the use of more limited sleep studies is being advocated for the diagnosis of patients.

Objective

To evaluate the diagnostic validity and costs of home monitoring compared with inpatient investigation of OSA.

Methods

Design

Pragmatic, single-blinded, cross-over, RCT. Patients were randomised to receive home monitoring followed by inpatient monitoring or vice versa. Physicians were blinded to the results of the second diagnosis and the decision to treat with nasal continuous positive airways pressure (nCPAP) was made on the basis of the first diagnosis alone.

Setting

A small district general hospital in Llanelli, Wales, UK.

Subjects

One hundred and two patients referred with suspected OSA, recruited between July 1995 and February 1997.

Intervention

Synectics Microdigitrapper S home sleep system.

Control

Inpatient monitoring using both Visi-Lab Sleep System (Version 3) and Compumedics P-Series Remote Sleep System.

Main outcome measures

The apnoea/hypopnoea index (AHI) derived from home and hospital monitoring systems and the cost of performing each arm of the study were compared to ascertain the sensitivity and specificity of the home monitoring system, level of agreement between the home and inpatient system, number of positive diagnoses made, NHS resource use and patient-borne costs.

Designed data

Sources of designed data

A direct comparison between home and inpatient diagnosis, including health economics (in terms of resource use and patient-borne costs) and two distinct inpatient diagnostic systems for sleep apnoea was made. Data relating to patient diagnosis were extracted from the individual systems. Information relating to resource use and patient-borne costs was recorded on a designed paper record form following an interview with the patient.

Analysis of designed data

The strength of agreement between the AHIs derived using the inpatient and home sleep systems was assessed using the Bland and Altman method.[34] The sensitivity and specificity of home monitoring compared with inpatient monitoring was calculated. Costs of both diagnostic methods were assessed. Data on resource use were not bootstrapped as they presented very low skew coefficient.[35]

Results of analysis of designed data

The median AHI was similar for inpatient (1.4; range 0–77.0) and home (1.6; range 0–45.4) diagnosis. The inpatient system diagnosed three more cases (25) than the home system (22) and there was an 83% level of agreement between the two systems with low failure rates. Using the Bland and Altman method, the mean difference was 2.2 (range 25.8 to –21.4; 95% CI, 4.8 to –0.4). The mean cost of home monitoring was £108.90 per patient compared with £334.07 for inpatient diagnosis ($p < 0.05$).

Conclusions from designed data

Monitoring for OSA at home compared well with inpatient diagnosis. Since home monitoring was considerably cheaper it may provide a useful alternative to inpatient investigation in a large proportion of cases.

Routine data

Sources of routine data

Patient-identifiable, electronically stored routine data were sought from the PEDW and the local hospital information system SCOPE (*Table 7*). Patients' names, dates of birth, postcodes and GP codes were used as identifiers. Data were also sought from primary care systems, but access was refused by many GPs in the light of changes in the law on confidentiality.[36]

Analysis of routine data

The original study was a cross-over trial and study patients should have at least two episodes of care. Of the 86 study patients identified on PEDW, only 19 (22%) were reported to have more than one episode, while of the 90 identified on SCOPE, only 14 (16%) were reported to have more than one episode. Therefore, the routine data for this

TABLE 7 *Nature and sources of routine data*

Designed data	Routine data	Source of routine data
Study start date	As designed study	Original study recruitment
Study end date	As designed study	Original study recruitment
Postcode	As designed study	Postcode finder on web
Randomisation code	As designed study	Original study recruitment
Date of recruitment	As designed study	Original study recruitment
Sensitivity of home monitoring Specificity of home monitoring Level of agreement between home and inpatient monitoring Mean bias between home and inpatient monitoring	ICD-10 codes for first episode within 6 months of recruitment date, including: G473 (Sleep apnoea) R06 (Abnormalities of breathing) Z03 (Medical observation and evaluation for suspected diseases and conditions)	PEDW, SCOPE
No. of patients with a positive diagnosis of OSA	Patients with episodes coded in ICD-10 code G473 within 6 months of the recruitment date	PEDW, SCOPE
	OPCS codes for all episodes within 6 months of recruitment date	PEDW, SCOPE
	Patient borrowing nCPAP machine	ECG department
Capital, service and disposable cost of Visi-Lab sleep system	As designed study	Hospital finance department
Capital, service and disposable cost of Compumedics sleep system	As designed study	Hospital finance department
Capital, service and disposable cost of Synectics equipment	As designed study	Hospital finance department
Staff time	As designed study	ECG department
Grade of staff involved	As designed study	Research proposal
Staff hourly rate	As designed study	Published information
Patient mileage	Imputed patient mileage	AA Route Finder
Patient travel cost	Imputed travel cost	AA
Patient lost productivity cost	No appropriate surrogate	
Estimate for repeat monitoring	No appropriate surrogate	
No. of visits	GP data	GP systems

TABLE 8 *Validity of routine data*

Data source	No. (%) of study patients identified (n = 102)	Comment
PEDW data	85 (83)	
SCOPE data	90 (88)	
Hospital ECG department	102 (100)	
GP data	34 (33)	Issues of patient consent raised by GPs

study were analysed as a simple RCT. Only 33% of GP records were available and these were excluded from the analysis. Categorical data were statistically analysed by the chi-squared test and numerical data by a two-sample t test.

The effectiveness of the two monitoring systems in picking up OSA was assessed by the difference between the two study groups in proportion of patients diagnosed with OSA (ICD-10 code: G473) on the first patient episode found within the study period in PEDW and SCOPE.

Coding policies in the hospital appeared to change during this study. Before 1997, patients admitted to rule out OSA might have been coded to Z03.8 'Observation for other suspected diseases and conditions', or to R06.8 for 'Unspecified Apnoea', if they did not turn out to have OSA. After 1997, coders would code the conditions treated. If OSA was diagnosed this would be the primary diagnosis. If it was not confirmed, the main diagnosis would be the main symptoms treated. Therefore, differences between the two study groups were assessed by the proportion of patient episodes with diagnostic codes R06, Z03 and other diseases.

Costs of the home monitoring system in secondary care were estimated by the typical cost of 30 minutes of grade H nurse time to explain the system to patients, capital, service and disposable costs.

To impute travel costs it was assumed that home monitoring patients made two visits to the hospital outpatient department (one to set up the equipment and one to return it) and hospital monitoring patients made one visit for their overnight monitoring. The cost of return trips was estimated from data on distances and cost per mile published by the AA.

Availability and validity of routine data

Electronically stored data were extracted by information managers or technicians expert in the individual systems according to specified matching criteria. Interpretation of the clinical codes was confirmed with the All Wales Clinical Coding Tutor.

The standard treatment for OSA is nCPAP at home. Therefore, findings related to the clinical diagnosis made were triangulated with the number of study patients issued an nCPAP machine according to the manually kept index card of the hospital ECG department, as well as those who had a procedure performed according to SCOPE or PEDW within 6 months of their recruitment dates (*Table 8*).

Results of analysis of routine data

The percentage of patients with an ICD-10 code of G473 'Sleep Apnoea' reported was high, but percentages of patients with other ICD-10 codes reported were low. Very few patients were reported to have had a procedure performed. There were more patients reported to have borrowed an nCPAP machine, but the percentage was still much lower than those who were reported to have had a positive diagnosis of OSA according to routine data.

There was no important difference between the findings derived from PEDW or SCOPE. None of the differences affected the comparison between inpatient and home monitoring.

There was no major difference between the study groups in reported positive diagnosis of OSA, symptoms related to abnormal breathing, observations related to monitoring, procedures performed or number of nCPAP machines issued.

The home monitoring cost significantly more in patient travel but less in hospital stays. However, costs estimated from routine data lacked information about both the need for repeated monitoring when initial monitoring was inconclusive and patients' time off work.

Table 9 summarises the comparison of the two study groups according to routine data from the different information systems.

TABLE 9 *Results of analysis of routine data*

Routine data (*n* by source and group)	Home monitoring	Inpatient monitoring	Mean difference (95% CI)
Proportion with OSA (code G473)			
(SCOPE: home = 45, inpatient = 45)	0.98	0.98	0.00 (–0.06 to 0.06)
(PEDW: home = 45, inpatient = 40)	0.91	0.98	–0.06 (–0.16 to 0.03)
Proportion with any other ICD-10 code			
(SCOPE: home = 45, inpatient = 45)	0.09	0.07	0.02 (–0.01 to 0.13)
PEDW: home = 45, inpatient = 40)	0.20	0.13	0.08 (–0.08 to 0.23)
Proportion with any procedure code			
(SCOPE: home = 45, inpatient = 45)	0.04	0.09	–0.04 (–0.15 to 0.06)
(PEDW: home = 45, inpatient = 40)	0.11	0.07	0.04 (–0.09 to 0.16)
Proportion borrowing nCPAP machine			
(ECG: home = 50, inpatient=52)	0.54	0.40	0.14 (–0.06 to 0.33)
Mean costs per patient (£) [SD]			
Inpatient stay			
(SCOPE: home = 45, inpatient = 45)	0	173.39	–173.39 (–193.40 to –153.38)
(PEDW: home = 45, inpatient = 40)	0	165.38	–165.38 (–183.33 to –147.43)
Capital/service cost			
(NHS Trust: home = 50, inpatient = 52)	18.90	38.85	–19.95 CI not appropriate; no SD
Disposables			
NHS Trust (home = 50, inpatient = 52)	14.5	8.61	5.88 CI not appropriate; no SD
Staff time			
(NHS Trust: home = 50, inpatient = 52)	10.30	6.17	4.13 CI not appropriate; no SD
Mean total cost per patient (£)			
(SCOPE: home = 45, inpatient = 45)	43.70	227.01 [67.55]	–183.32 (–203.33 to –163.31)
(PEDW: home = 45, inpatient = 40)	43.70	238.93 [38.83]	–195.23 (–206.73 to –183.73)
Patient travel cost			
(SCOPE: home = 45, inpatient = 45)	13.08 [9.28]	6.35 [5.31]	6.73 (3.56 to 9.90)
(PEDW: home = 45, inpatient = 40)	13.02 [9.31]	6.30 [4.96]	6.71 (3.53 to 9.89)
Sum of service and patient-borne costs per patient (£) [SD]			
Total societal cost[a]			
(SCOPE: home = 45, inpatient = 45)	56.78 [9.28]	233.37 [67.97]	–176.59 (–196.91 to –156.26)
(PEDW: home = 45, inpatient = 40)	56.83 [9.38]	245.44 [39.43]	–188.62 (–200.83 to –176.40)

[a] Incomplete estimate as cost of lost productivity not available.

Discussion of routine data analysis

The proportion of patients with a positive diagnosis of OSA reported in PEDW and PAS was unexpectedly high (over 90%), compared with the proportion of patients with any procedures performed on them (<10%) or issued an nCPAP machine (47%) within 6 months of the recruitment date. Nevertheless, there was no reason to suspect any difference between the study groups in recording the diagnoses in the routine data systems.

There was no information on the grade of staff and exact staff time spent with patients for the two systems. The current estimate was projected from the research proposal and the subjective judgement of staff. These may not reflect what actually happened in the clinic or ward.

The cost estimates from routine data have to be interpreted with caution. There was no appropriate surrogate from the routine data systems for instances of repeated monitoring. There was no specified field in the electronic databases to record the number of technical failures. There was also no record in the electronic databases about the number of attempts made

TABLE 10(a) *Diagnostic accuracy: comparison between designed and routine data*

Designed data (*n*)	Mean difference between groups (95% CI)	Routine data: surrogate variable (*n* by source)	Mean difference between groups (95% CI)
Difference in AHI (events per hour) between the two monitoring systems (84)	2.2 (4.8 to –0.4)	No available surrogate	
Percentage of patients with a positive diagnosis (102)	–0.04 (–0.23 to 0.15)	Proportion of patients with ICD-10 codes of G473	
		SCOPE (90)	0.00 (–0.06 to 0.06)
		PEDW (85)	–0.06 (–0.16 to 0.03)
		Proportion of patients with any procedure codes	
		SCOPE (90)	–0.04 (–0.15 to 0.06)
		PEDW (85)	0.04 (–0.09 to 0.16)
		Proportion of patient borrowing an nCPAP machine ECG (102)	–0.14 (–0.06 to 0.33)
Summary of within-subject comparison of AHI readings from the two monitoring systems			
Sensitivity of home monitoring against inpatient system (85)	63%	No available surrogate	
Specificity of home monitoring against inpatient system (85)	86%	No available surrogate	
Level of agreement between home and inpatient system (85)	83%	No available surrogate	

within the same episode before an investigative procedure was successful. This could have led to bias in the current study. As staff were less familiar with the home monitoring system, there could have been more instances of technical failures.

The lack of appropriate surrogates for patients' time off work from routine data might have caused an underestimate of the cost of the inpatient monitoring system. However, it was the hospital's standard practice to admit patients for OSA monitoring in the evenings and discharge them in the mornings. This should have minimised patients' time off work through inpatient monitoring.

Conclusions from routine data

The home monitoring system picks up a similar proportion of patients with OSA to the inpatient monitoring system. However, it is not possible to estimate the true proportion of study patients with OSA from the routine data analysis. Home monitoring costs significantly less than inpatient monitoring. However, the estimated costs of the two monitoring systems from routine data are incomplete and likely to be imprecise.

Comparison of designed and routine datasets

Table 10(a) shows broadly similar differences between the study groups in the percentage of patients with a positive diagnosis of OSA whether they are derived from designed or routine data. *Table 10(b)* shows remarkable consistency between the findings of the two analyses. This was not surprising given the similar sources of data.

Table 11 shows that routine data found no significant differences in diagnostic accuracy between the two monitoring systems. There was consistent over-reporting of the proportion of study patients with a positive OSA diagnosis across the electronic routine data sources, compared with the designed data. The secondary cost estimate came from a partial analysis of the total cost. Nevertheless, conclusions about resource use were similar to those of the designed analysis.

In summary, the routine data answered the two original research questions to some degree.

TABLE 10(b) *Resources used in diagnosis: comparison between designed and routine data*

	Mean difference between groups (95% CI)			
Resource variable (*n*)	**Designed data**	**Source**	**Routine data**	**Source**
Patient-borne costs (£)				
Patient travel (102)	5.33 (2.36 to 8.31)	Patient self-report	6.73 (3.56 to 9.90) 6.71 (3.43 to 9.99)	SCOPE (*n* = 90) PEDW (*n* = 85)
Lost productivity (101)	–4.58 (–32.82 to 23.66)	Patient self-report	–	None
Total patient-borne costs (101)	0.79 (–27.89 to 29.47)	Patient self-report	–	Incomplete
Service costs (£)				
Inpatient stay (102)	–169.62 (CI not appropriate; no SD)	Patient notes	–173.39 (–193.40 to –153.38) –165.38 (–183.33 to –147.43)	SCOPE (*n* = 90) PEDW (*n* = 85)
Capital/service cost (102)	–19.95 (CI not appropriate; no SD)	NHS Trust	–19.95 (CI not appropriate; no SD)	As designed data
Disposables (102)	5.88 (CI not appropriate; no SD)	NHS Trust	5.88 (CI not appropriate; no SD)	As designed data
Staff time (100)	–31.42 (–39.08 to –23.76)	Directly monitored	4.13 (CI not appropriate; no SD)	ECG department of Trust (*n* = 102)
Sum of all service costs (£)				
Total secondary care cost, excluding repeats (100)	–215.10 (–222.76 to –207.44)		–183.32 (–202.33 to –163.31)[a] –195.23 (–208.17 to –182.29)[a]	SCOPE (*n* = 90) PEDW (*n* = 82)
Sum of patient-borne and service costs (£)				
Total societal cost, excluding repeats (99)	–210.68 (–240.88 to –180.49)		–183.32 (–203.33 to –163.31)[a] –195.23 (–206.73 to –183.73)[a]	SCOPE (*n* = 90) PEDW (*n* = 81)
Repeat monitoring (102)	6.09 (–0.61 to 12.79)	Directly monitored	–	None
Sum of patient-borne and service costs, including repeat monitoring (£)				
Total societal cost, including repeats (99)	–204.34 (–235.00 to –173.68)		–183.32 (–203.33 to –163.31)[a] –195.23 (–206.73 to –183.73)[a]	SCOPE (*n* = 90) PEDW (*n* = 81)

[a] Incomplete estimate as no routine data available for lost productivity and repeat monitoring.

TABLE 11 *Reconciliation by research questions*

Research question	Definitive conclusion from analysis of designed data	Conclusion from analysis of routine data	Comments
To compare the diagnostic value of home versus inpatient monitoring for OSA	Monitoring for OSA at home compared well with inpatient diagnosis	Home monitoring picked up a similar proportion of patients with OSA as the inpatient system	Consistent overestimation of proportion of patients with a positive OSA diagnosis across the electronic routine data systems
To compare the cost of home versus inpatient monitoring for OSA	Hospital monitoring was more than three times as expensive as home monitoring	Home monitoring costs more than hospital monitoring in patient travelling costs. Home monitoring cost less than hospital monitoring inpatient costs	Secondary analysis estimated only part of the total cost

Chapter 5

Exemplar C: a randomised study to assess and compare the clinical effectiveness of two surgical techniques for the treatment of stress urinary incontinence[28]

Background

Urinary incontinence is a common problem in women, which causes considerable morbidity. Many procedures have been described for the treatment of stress incontinence but none has gained universal acceptance. A meta-analysis of the world literature covering the past 50 years has revealed fewer than ten properly conducted RCTs of surgery for this problem.

Early evidence suggests that a sling fashioned from the abdominal wall fascia to support the bladder neck achieves consistently reliable results with a durable outcome, but with the disadvantage of requiring extensive dissection and often causing pain at the site of fascial harvest. A recent modification has been to harvest a much smaller piece of fascia and suspend it from Nylon threads attached to the abdominal wall. There is potential for both easier surgery and avoidance of pain using this technique, but it has not previously been rigorously evaluated.[37]

Objective

To establish the morbidity and clinical efficacy of two techniques of fascial sling for urinary incontinence.

Methods

Design
Pragmatic three-centre RCT.

Setting
Three hospitals [Singleton, Morriston and the University Hospital of Wales (UHW) at Cardiff].

Subjects
One hundred and sixty-five women with clinically proven stress urinary incontinence undergoing surgery between October 1996 and February 1999, and followed up for 12 months.

Intervention
Short pubovaginal fascial sling, harvested from leg fascia and mounted on Nylon thread.

Control
Conventional long sling, harvested from leg fascia.

Main outcome measures
Data were collected pre- and postoperatively and at 3, 6 and 12 months following surgery. Primary end-points were reduction in leaked urine volume, changes in HRQoL scores [Urogenital Distress Inventory (UDI) and Incontinence Impact Questionnaire (IIQ)], symptoms of urinary urgency, voiding difficulty and patient satisfaction. Secondary end-points were postoperative pain, operative time, length of stay in hospital, and immediate and late complications including readmissions to hospital within 12 months.

Designed data

Sources of designed data
HRQoL scores were collected using validated questionnaires completed by patients at entry, and 3, 6 and 12 months after surgery. Data on operative time were collected from the operating theatre information system, Theatre Man, at Singleton, but were not available from this source at Morriston or UHW, and were specifically recorded for the study. Patients' pain and site of pain were scored at 24 hours, 4 or 5 days and 3 months after the operation. Volume of leakage (by 1 hour pad tests), symptoms of urinary urgency and voiding difficulty were assessed at study entry, 3, 6 and 12 months at the clinic. Data on patient satisfaction, complications, readmissions and length of stay (LOS) in hospital were collected using patient questionnaires and routine electronic data sources at 3, 6 and 12 months after surgery. Blinded reviews of data on complications, readmissions and voiding difficulty were carried out by clinicians to establish relevance.

TABLE 12 *Nature and sources of routine data*

Designed data	Routine data	Source of routine data
Study start date	As designed study	Original study recruitment
Study end date	As designed study	Original study recruitment
Postcode	As designed study	Original study recruitment
GP code	As designed study	Original study recruitment
Date of recruitment	As designed study	Original study recruitment
Randomisation	As designed study	Randomisation schedule
Length of postoperative stay	Total LOS following operative procedure	GeneCIS, Cardiff PAS Inpatient system, PEDW, Morriston PAS data
Readmission: 3–6 months 7–12 months up to 12 months	Total no. of inpatient readmissions up to 3 months following operative procedure 7–12 months not available in timescale	GeneCIS, Cardiff PAS Inpatient system, PEDW, Morriston PAS, Singleton PMS, PEDW
Pad test baseline	No appropriate surrogate	No appropriate surrogate
Pertinent voiding difficulties baseline	Urodynamic and residual volume examination, self-catheterisation at baseline	Cardiff Radiology, Singleton Radiology
UDI QoL baseline	Any comparable diagnostic symptom codes at baseline	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
IIQ QoL baseline	Any comparable diagnostic symptom codes at baseline	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Presence and site of pain 24 h postoperation	Any postoperative pain diagnostic code associated with surgical procedure episode Site of pain not available	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Presence and site of pain 4–5 days postoperation	Any postoperative pain diagnostic code associated with surgical procedure episode Site of pain not available	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Pad test 3 months postoperation	No appropriate surrogate	No appropriate surrogate
Presence and site of pain 3 months postoperation	Any postoperative pain diagnostic code up to 1 year Site of pain not available	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Patient satisfaction 3 months	No appropriate surrogate	No appropriate surrogate
UDI QoL 3 months	Any comparable diagnostic symptom codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
IIQ QoL 3 months	Any comparable diagnostic symptom codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Pad test 6 months postoperation	No appropriate surrogate	No appropriate surrogate
Presence and site of pain 6 months postoperation	Any postoperative pain diagnostic code up to 1 year Site of pain not available	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Patient satisfaction 6 months	No appropriate surrogate	No appropriate surrogate
UDI QoL 6 months	Any comparable diagnostic symptom codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
IIQ QoL 6 months	Any comparable diagnostic symptom codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Pad test 12 months postoperation	No appropriate surrogate	No appropriate surrogate

continued

TABLE 12 *Nature and sources of routine data (cont'd)*

Designed data	Routine data	Source of routine data
Pertinent voiding difficulties 12 months postoperation	Urodynamic and residual volume examination, self-catheterisation up to 1 year postoperation	Cardiff Radiology, Singleton Radiology
Presence and site of pain 12 months postoperation	Any postoperative pain diagnostic code up to 1 year	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
	Site of pain not available	
Patient satisfaction 12 months	No appropriate surrogate	No appropriate surrogate
UDI QoL 12 months	Any comparable diagnostic symptom codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
IIQ QoL 12 months	Any comparable diagnostic symptom codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW
Perioperative complications: Bladder injury Bleeding UTI Suprapubic catheter problems Respiratory Cardiac Wound	Any adverse events/complications codes up to 1 year postoperation	GeneCIS, Cardiff PAS Inpatient system, Morriston PAS, Singleton PMS, PEDW, Singleton Theatre
	Length of time under anaesthesia, recovery from operation	

Analysis of designed data

Data were analysed by intention to operate. To counteract the effect of possible differences in baseline HRQoL scores and volume of leakage, changes in individual measurements from baseline were compared with *t* tests. Covariance and regression analysis[38] were performed to see whether any observed significant change was a statistical artefact. Operative time and length of hospital stay were also analysed by *t* tests. Incidence and sites of postoperative pain, urinary urgency, voiding difficulty, patient satisfaction, complications and readmissions were analysed using chi-squared tests.

Results of analysis of designed data

At 12 months there was no significant difference between the two groups in reduction of urine leakage, stress incontinence, urge incontinence, voiding difficulty or patient satisfaction. Long sling patients improved significantly more in UDI scores at 12 months. However, covariance and regression analysis suggested that the observed difference in relative improvement between the study groups could be due to their difference scores at recruitment.

There was no significant difference in pain or site of pain between the two groups at 24 months or at 4 or 5 days postsurgery. No significant difference was found in perioperative complications, postoperative LOS or adverse events not requiring readmission. Operating time for the short sling group was significantly shorter. Patients in the short sling group had significantly less pain at the lateral angles of sling dissection or readmission at 3 months.

Conclusions from designed data

The short sling technique is quicker to perform and causes less long-term pain at the lateral angles of sling dissection. There is no other significant difference between the efficacy and morbidity of the two procedures at 12 months.

Routine data

Sources of routine data

Patient-identifiable, electronically stored routine data were sought from the PEDW, hospital PAS, radiology, pathology and theatre systems. At Morriston, a departmental clinical information system (GeneCIS) was also accessed (*Table 12*).

Analysis of routine data

The two surgical techniques were compared at baseline and 12 months to determine morbidity and clinical effectiveness. IIQ and UDI questionnaires were used as templates to determine the presence of diagnoses relevant to stress incontinence in order to assess clinical effectiveness. Numerical data were analysed by *t* tests and categorical data by chi-squared tests.

TABLE 13 *Validity of routine data*

Data source	No. (%) of study patients identified (*n* = 165 overall: Morriston = 41; Singleton = 69; Cardiff = 55)	Comment
PEDW (all sites)	78 (47)	PEDW was accessed 6 months after the last patient entered the study
GeneCIS (Morriston)	39 (95)	
Cardiff CFIS	47 (85)	
Cardiff Radiology	50 (91)	
Cardiff Pathology	55 (100)	
Morriston PAS	41 (100)	Only 50% of the patients had diagnostic and procedure codes documented
Morriston Pathology	41 (100)	Had to print out results: could not save to disk
Morriston Theatre	0 (0)	Theatre Man system was down for at least 6 months
Singleton PMS	23 (33)	Difficulties obtaining data from the system
Singleton Radiology	66 (96)	
Singleton Pathology	69 (100)	Had to print out results: could not save to disk
Singleton Theatre	67 (97)	

Availability and validity of routine data

The level of data completeness varied with the different data sources as shown in *Table 13*.

Results of analysis of effectiveness using routine data

There were no significant differences between the two groups in postoperative LOS or readmission within 3 months of surgery. The proportion of patients with relevant complications following surgery was similar in the two groups, as was the number of patients who were self-catheterised. The proportion of patients reporting any of the relevant symptoms from the IIQ or UDI questionnaire was similar at baseline and up to 12 months following surgery in the two groups.

The total number of postoperative urodynamic and residual volume procedures and time to first and last urodynamic or residual volume procedures were used as surrogates for successful voiding on the assumption that these tests would not be repeated if voiding were successful. None of these variables showed any differences between the two groups. Urine leakage as measured by the pad test could not be compared, as no surrogate variables were available.

The analysis of routine data found only one significant difference between the two surgical procedures. The operation time was significantly shorter for group B (sling on a string) than group A (long sling) (11.7 minutes, $p < 0.05$; 95% CI, –21.1 to –2.35).

There were no differences in the number of patients with reported 'clean-intermittent self-catheterisation' postoperation. Length of time to first successful void could not be compared owing to a lack of data. Successful voiding up to 1 year was proxied using the length of time to the first and last urodynamic examination, and neither of these showed any differences between the two groups. Patient satisfaction could not be compared with the available data.

Conclusions from routine data

The electronically stored routine data available were able to answer some of the research questions. Clinical information systems that allow reporting of symptoms, signs and diagnoses would have been useful for comparing QoL between the groups. Some capacity for reporting patient comments would also have allowed patient satisfaction to be compared.

Comparison of clinical outcome using designed and routine datasets

Table 14(a, b) shows the difference in operative morbidity and outcome derived from designed and routine data. The results are consistent, with a few exceptions. The designed data show a significant difference in pain at the lateral angles of sling dissection at 3 months. The routine data analysis was able to pick up those patients given a diagnostic code of pain during the operative

TABLE 14(a) *Differences in operative morbidity: comparison between designed and routine data*

Designed data (*n*)	Mean difference between groups (95% CI)	Routine data: surrogate variable (*n*)	Mean difference between groups (95% CI)
Length of postoperative stay (days) (165)	0.25 (–0.71 to 1.20)	LOS postoperation GeneCIS (36) Cardiff CFIS (31) PEDW (70) All systems (86)	 –0.41 (–2.67 to 1.85) 0.75 (–1.99 to 3.50) 0.65 (–1.01 to 2.31) 0.19 (–1.55 to 1.93)
Proportion of patients with any type of pain: 24 h postoperation (159)	0.04 (–0.12 to 0.04)	Not available	
Proportion of patients with pain at the lateral angles of sling dissection: 24 h postoperation (159)	0.02 (–0.13 to 0.17)	Not available	
Proportion of patients with any type of pain: 4–5 days postoperation (165)	0.15 (0 to 0.30)	Not available	
Proportion of patients with pain at the lateral angles of sling dissection: 4–5 days postoperation (165)	0.05 (–0.08 to 0.18)	Not available	
Proportion of patients with any type of pain: 3 months postoperation (160)	–0.15 (–0.30 to 0)	Not available	
Proportion of patients with pain at the lateral angles of sling dissection: 3 months postoperation (160)	–0.19 (–0.34 to –0.05)	Proportion with diagnostic code of pain unspecified R529 or pain localised to other parts of lower abdomen R103: postoperation to 1 year: Cardiff CFIS (47)	0.05 (–0.04 to 0.13)
		Proportion with diagnostic code of lower abdomen pain R103: postoperation to 1 year PEDW (35)	0.01 (–0.14 to 0.17)
Proportion with immediate postoperative complications (165): Bladder injury Bleeding UTI Suprapubic catheter problems Respiratory Cardiac Wound	 0.01 (–0.04 to 0.06) 0.03 (–0.12 to 0.07) 0.05 (–0.14 to 0.04) 0.02 (–0.03 to 0.08) 0.00 (–0.06 to 0.06) 0.03 (–0.03 to 0.10) 0.03 (–0.11 to 0.05)	Proportion with postoperative ITU admission GeneCIS (39)	–0.27 (–0.56 to 0.02)
Proportion with any postoperative complications up to 3 months (162)	0.08 (–0.08 to 0.23)	Proportion with postoperative wound haematoma up to 3 months: GeneCIS (39)	–0.05 (–0.15 to 0.05)
		Proportion with postoperative urinary tract infection up to 3 months GeneCIS (39)	–0.06 (–0.22 to 0.11)

continued

TABLE 14(a) *Differences in operative morbidity: comparison between designed and routine data (cont'd)*

Designed data (*n*)	Mean difference between groups (95% CI)	Routine data: surrogate variable (*n*)	Mean difference between groups (95% CI)
		Proportion having CXR for respiratory complications up to 3 months postoperation	
		Cardiff Radiology (49)	–0.08 (–0.26 to 0.10)
		Singleton Radiology (66)	0.06 (–0.06 to 0.17)
Proportion with any postoperative complications: 3–6 months (163)	0.10 (–0.25 to 0.05)	Not available	
Proportion with any postoperative complications: >6 and up to 12 months (153)	0.09 (–0.05 to 0.23)	Not available	
Proportion with complications related to operation from discharge up to 12 months (165)	–0.09 (–0.23 to 0.04)	Proportion with any relevant postoperative complication up to 12 months	
		GeneCIS (39)	–0.11 (–0.30 to 0.08)
		Cardiff CFIS (47)	–0.04 (–0.17 to 0.10)
		PEDW (34)	–0.02 (–0.33 to 0.30)
		Singleton PMS (23)	–0.28 (–0.60 to 0.05)
		All systems (86)	–0.07 (–0.18 to 0.05)
Proportion of patients readmitted for problems related to surgery up to 3 months (162)	–0.13 (–0.25 to –0.02)	Proportion of patients readmitted postoperation up to 3 months	
		Cardiff CFIS (47)	–0.15 (–0.36 to 0.06)
		PEDW (80)	–0.23 (–0.41 to –0.04)
		All systems (86)	–0.10 (–0.30 to 0.10)
Proportion of patients readmitted for problems related to surgery: 3–6 months (163)	0.04 (–0.05 to 0.13)	Not available	
Proportion of patients readmitted for problems related to surgery: >6 to 12 months (153)	–0.06 (–0.16 to 0.4)	Not available	
Proportion of patients readmitted for problems related to surgery from discharge up to 12 months (165)	–0.13 (–0.28 to 0.01)	Not available	

CXR: chest X-ray.

period, but the site of the pain could not be determined. The number of patients in each group given a pain code was not significantly different. The designed data analysis highlighted a significant difference in favour of sling in readmission at 3 months, but this was not shown with routine data on the number of postoperative readmissions. Neither designed nor routine data showed any significant difference in clinical effectiveness between the two surgical techniques.

Both analyses showed a significant advantage in favour of the sling on a string in the duration of operation.

Table 14(c) shows reasonable consistency between designed data and surrogate variables from GeneCIS in changes in health outcomes for two of the four HRQoL scales. Analysis of the designed data showed significant differences in changes at 12 months for two of the HRQoL subscales

TABLE 14(b) *Differences in clinical outcome: comparison between designed and routine data*

Designed data (*n*)	Mean difference between groups (95% CI)	Routine data: surrogate variable (*n*)	Mean difference between groups (95% CI)
Skin-to-skin operation time (minutes) (165)	–8.43 (–13.34 to –3.52)	Skin-to-skin operation time (minutes) (67)	–11.7 (–21.1 to –2.35)
Proportion with stress incontinence at 12 months (154)	0.00 (–0.12 to 0.12)	Proportion with diagnostic code of stress incontinence of urine N393: postoperation to 1 year	
		GeneCIS (39)	–0.14 (–0.39 to 0.11)
		Cardiff CFIS (47)	–0.08 (–0.35 to 0.19)
Proportion with urge syndrome at 12 months (154)	–0.08 (–0.24 to 0.07)	Proportion with diagnostic code of other specified incontinence of urine N394: postoperation to 1 year	
		GeneCIS (39)	0.10 (–0.03 to 0.23)
		Proportion with diagnostic code of retention of urine R33 or other difficulties with micturition N391: postoperation to 1 year	
		Cardiff CFIS (47)	–0.08 (–0.19 to 0.03)
		Proportion with diagnostic code of retention of urine R33: postoperation to 1 year	
		PEDW (35)	–0.05 (–0.15 to 0.05)
Proportion with pertinent voiding difficulty at 12 months (165)	–0.03 (–0.16 to 0.09)	Proportion with diagnostic code of self-catheterisation	
		GeneCIS (39)	–0.01 (–0.24 to 0.22)
		Total no. of urodynamic procedures performed postoperation to 1 year	
		Cardiff Radiology (49)	–0.20 (–0.67 to 0.27)
		Singleton Radiology (66)	–0.10 (–0.38 to 0.18)
		All systems (115)	–0.15 (–0.42 to 0.11)
		Time postoperation to first urodynamic procedure (months)	
		Cardiff Radiology (42)	0.30 (–2.01 to 2.60)
		Singleton Radiology (41)	–1.67 (–3.71 to 0.37)
		All systems (83)	–0.58 (–2.20 to 1.04)
		Time postoperation to last urodynamic procedure (months)	
		Cardiff Radiology (42)	–0.56 (–4.72 to 3.61)
		Singleton Radiology (41)	–2.02 (–4.13 to 0.09)
		All systems (83)	–1.31 (–3.60 to 0.99)
		Total no. of residual volume tests performed postoperation to 1 year	
		Singleton Radiology (66)	–0.05 (–0.34 to 0.24)
		Time postoperation to first residual volume test (months)	
		Singleton Radiology (17)	0.47 (–2.70 to 3.65)
		Time postoperation to last residual volume test (months)	
		Singleton Radiology (17)	1.44 (–2.09 to 4.97)

TABLE 14(c) *Differences in health status between baseline and 12 months: comparison between designed and routine data*

Designed data (*n*)	Mean difference between groups (95% CI)	Routine data: surrogate variable (*n*)	Mean difference between groups (95% CI)
Changes in UDI Obstructive symptoms score at 12 months (131)	5.34 (–2.45 to 13.12)	Changes in surrogate UDI score for Obstructive symptoms at 12 months (All systems)	–0.08 (–0.26 to 0.09)
Changes in UDI Irritative symptoms score at 12 months (132)	13.59 (1.89 to 25.29)	Changes in surrogate UDI score for Irritative symptoms at 12 months (All systems)	0.01 (–0.12 to 0.13)
Changes in UDI Stress symptoms score at 12 months (132)	8.05 (–4.03 to 20.13)	Changes in surrogate UDI score for Stress symptoms at 12 months (All systems)	–0.16 (–0.35 to 0.03)
Changes in UDI Total score at 12 months (131)	27.05 (1.12 to 52.99)	Changes in surrogate Total UDI score at 12 months (All systems)	–0.23 (–0.50 to 0.03)
Changes in IIQ Physical activity score at 12 months (133)	6.72 (–5.84 to 19.29)	No available surrogate	
Changes in IIQ Travel score at 12 months (134)	10.32 (–2.04 to 22.68)	No available surrogate	
Changes in IIQ Social score at 12 months (133)	11.29 (–0.29 to 22.87)	No available surrogate	
Change in IIQ Emotional score at 12 months (134)	10.85 (–1.11 to 22.80)	No available surrogate	
Change in IIQ Total score at 12 months (133)	40.82 (–4.32 to 85.96)	No available surrogate	

TABLE 15 *Reconciliation by research question*

Research question	Definitive conclusions from analysis of designed data	Conclusions from analysis of routine data	Comments
To compare the morbidity of two techniques of fascial sling for urinary incontinence	Significant difference in favour of sling on a string in terms of pain at the lateral angles of sling dissection and readmissions at 3 months. No difference in terms of postoperative complications and postoperative LOS. No differences recorded in presence or site of pain at 24 h or 4/5 days postsurgery. No significant difference in terms of readmission at 6 or 12 months after surgery	No differences in the number of complications or ITU admissions postoperation between the two groups. No differences in postoperative LOS or number of admissions	The number of individual relevant complications was small, hence they were analysed as all complications. Although not significant, patients undergoing the long sling operation appeared to have more ITU admissions

continued

TABLE 15 *Reconciliation by research question (cont'd)*

To compare the clinical efficacy of two techniques of fascial sling for urinary incontinence	Significant difference in favour of sling on a string in terms of operating time. No significant difference in terms of symptoms of stress incontinence, urge incontinence or voiding difficulties at 12 months	Significant difference in favour of sling on a string in terms of operating time. No difference in self-catheterisation or number of urodynamic or residual volume procedures	Surrogate parameters were generated for a number of comparisons owing to absence of true parameters in routine data sources Self-catheterisation, urodynamic and residual volume tests were used as surrogates for voiding difficulties.
To compare changes in health status using two disease specific questionnaires (IIQ and UDI) of two techniques of fascial sling for urinary incontinence	No significant difference in QoL	No differences in displayed individual symptoms or in composite UDI scores at baseline or 12 months	Only a proportion of the listed diagnoses were comparable to the UDI and IIQ questions. A composite IIQ score could not be generated as only one question had a comparable diagnostic code
To compare patient satisfaction with the procedures	No difference in patient satisfaction between the two groups	No appropriate surrogate for patient satisfaction	

which were not shown using surrogate variables from GeneCIS. Covariance analysis suggested that the observed differences using designed data were caused by differences at baseline (Appendix 3).

Table 15 compares the findings from the two datasets by research question. Similar conclusions were drawn from the routine data for two of the three original research questions. No conclusions about patient satisfaction could be drawn from routine data. However, GeneCIS does have the facility for recording additional data in structured or free-text form, and it is possible that this facility could be used to collect patient satisfaction data prospectively.

Chapter 6

Exemplar D: autologous blood transfusion in total knee replacement surgery[29,30]

Background

Concern over the safety of homologous blood transfusion has led to increased interest in the use of autologous blood collected from the patient by venesection before anticipated need. It can also be collected by cell savage at operation.

Objective

To test the hypothesis that cell salvage and autologous blood transfusion results in better clinical outcomes and is more cost-effective than the use of donor blood in TKR surgery.

Methods

Design
Single-centre RCT.

Setting
Morriston Hospital, Swansea, Wales, UK.

Subjects
Two hundred and thirty-one patients requiring TKR surgery recruited during the period May 1995 to July 1998.

Intervention
Perioperative cell salvage and autologous blood transfusion.

Control
Homologous blood transfusion as required. Patients received this if their condition warranted or if their haemoglobin fell below the preset trigger of 9 g/dl.

Main outcome measures
Averse events, transfusion requirements, wound healing, LOS and readmission within 6 months were documented. Patients' HRQoL (EuroQol) was measured before and 7 days, 4 weeks and 3 months after surgery.

Designed data

Sources of designed data
Data on HRQoL were collected using validated questionnaires completed by the patients before operation, and 7 days, 4 weeks and 3 months after surgery. Details of adverse events, transfusion requirements, wound healing, LOS and readmission were extracted from patient records by a dedicated research nurse.

Analysis of designed data
Analysis was by intention to transfuse. Numerical data were compared using two-sample *t* tests. Categorical data were analysed by the chi-squared test. Baseline demographics were compared to ensure that the groups were comparable.

Results of analysis of designed data
There was no significant difference in clinical outcome when analysed for LOS, wound healing, serious adverse events or mortality, or HRQoL 6 months following surgery. There was a significantly lower incidence of homologous blood transfusion in the autologous group, compared with the controls ($p = 0.001$). There was no difference in postoperative mean haemoglobin levels between the two groups. There were significantly fewer readmissions to hospital ($p = 0.008$) and visits to GPs ($p = 0.043$) among patients in the autologous blood transfusion group. Infection complications were increased in recipients of homologous blood ($p = 0.036$). The cost of autologous blood transfusion was nearly three times as expensive as homologous transfusion. Data on resource use were not bootstrapped as they presented very low skew coefficient.[35]

Conclusions from designed data
Postoperative cell salvage and autologous blood transfusion is a safe and effective method for reducing homologous blood use.

TABLE 16 *Nature and sources of routine data*

Designed data	Routine data	Source of routine data
No. and type of adverse events/immediate and subsequent postoperative infections	Patient episodes with the following diagnostic codes within the study period: T81.7: vascular complications following a procedure (postoperative DVT and postoperative pulmonary embolism) T81 other than T81.7: complication of procedure not elsewhere specified, excluding the vascular complications T84: complications of internal orthopaedic prosthetic device implant and graft E89: postprocedural endocrine and metabolic disorders G97: postprocedural disorders of the nervous system H59: other postprocedural disorders of eye and adenexa H95: other postprocedural disorders of ear and mastoid I97: postprocedural disorder of circulatory system J95: postprocedural respiratory disorder K91: postprocedural disorders of digestive system M96: postprocedural musculoskeletal disorder N99: postprocedural disorders of genitourinary system Y83: postoperative complications not classified in the rest of ICD-10 chapters	PAS, PEDW
Mortality	Method of discharge from a hospital provider spell within 6 months of the study: patient died	PEDW
Time to wound healing	Not available in electronic form	–
Drug profile	Not available in electronic form	–
Haemoglobin levels (days 1, 4 and 7)	Haemoglobin levels at days 1, 4 and 7	Blood bank
End of study EuroQol: mobility	Patients with the ICD-10 code of Z74.0: reduced mobility at last patient episode	PAS, PEDW
End of study EuroQol: self-care	Patients with the ICD-10 codes of Z74.1: need for assistance with personal care Z74.3: need for continuous supervision at last patient episode	PAS, PEDW
End of study EuroQol: usual activities	Patients with the ICD-10 codes of Z73.6: limitation of activities due to disability Z73.9: problem due to life management difficulty, unspecified at last patient episode	PAS, PEDW
End of study EuroQol: pain/discomfort	Patients with the ICD-10 codes of R52: pain, not elsewhere specified R52.0: acute pain R52.1: chronic intractable pain R52.2: other chronic pain R52.9: pain, unspecified at last patient episode	PAS, PEDW
End of study EuroQol: anxiety/depression	Patients with the ICD-10 codes of F32.1: moderate depressive episode F32.2: severe depressive episode without psychotic symptoms F41.1: generalised anxiety disorder F41.2: mixed anxiety and depressive disorder F41.9: anxiety, not elsewhere specified at last patient episode	PAS, PEDW
End of study EuroQol: VAS	Not available	
Transfusion requirements	Units of homologous blood transfused No. of patients transfused with homologous blood	Pathology
	Units of autologous blood transfused No. of patients transfused with autologous blood	Theatre Man[a]

continued

TABLE 16 *Nature and sources of routine data (cont'd)*

Designed data	Routine data	Source of routine data
Cost per unit of homologous blood	As designed study	Hospital finance department
Capital, service and disposable cost of the cell salvage system	As designed study	Hospital finance department
Staff time	Not available in electronic forms	–
Length of hospital stay	Total length of hospital stay within study period	PAS, PEDW
Readmission	No. of inpatient spells within study period minus one	PAS, PEDW
Hospital hotel cost	As designed study	Hospital finance department

[a] No data were available owing to a technical problem with Theatre Man at the time of the query.
DVT: deep vein thrombosis

TABLE 17 *Validity of routine data*

Data source	No. (%) of study patients identified (n = 231)	Comment
PEDW	163 (71)	Data items found varied from patient to patient
PAS	223 (97)	
Blood bank	215 (93)	Had to print out results; could not save to disk. Data items found varied from patient to patient
Theatre data	0 (0)	Theatre Man system was down for at least 6 months
Radiology data	0 (0)	Morriston Radiology department had no electronic system for period of study

Routine data

Sources of routine data

Patient-identifiable, electronically stored routine data were sought from the PEDW, the hospital PAS and the blood bank of the Trust pathology system (*Table 16*). Data from Theatre Man were not available because of unresolved local technical problems.

Analysis of routine data

Routine data were analysed by intention to transfuse. Haemoglobin levels were adjusted for gender differences by the empirical mean difference between males and females before analysis. Numerical data were analysed by two-sample t test. Categorical data were analysed by chi-squared test. The confidence intervals of the total secondary care costs were checked with 1000 replications bootstrap.

Availability and validity of routine data

Electronically stored data were extracted by information managers or technicians, expert with the individual systems according to specified matching criteria. Interpretation of the use of clinical codes was confirmed with professional clinical coders. The level of data completeness varied with the different data sources. *Table 17* illustrates the level of completeness for at least one patient episode of care.

There were doubts regarding the validity of clinical data from PAS and PEDW used as surrogates for HRQoL analysis. However, there was no reason to suspect any difference between the two study groups in the accuracy of recording of clinical data on these systems. It was not possible to estimate from routine data the total cost of the study groups. There was no available information on staff time, theatre time or time to wound healing.

Results of analysis of effectiveness using routine data

On PAS, there were 306 consultant episodes (autologous: 162; homologous: 144) for 223 patients (autologous: 111; homologous: 112) out of a total of 231 in the study. There were 44 patients with more than one inpatient spell (autologous: 27; homologous: 17). On PEDW, there were 219 consultant episodes (autologous: 106; homologous: 113) for 163/231 study patients

TABLE 18 *Results from analysis of routine data*

Routine data (*n* by source and group)	Autologous	Homologous	Mean difference between groups (95% CI)
Proportion of patients requiring transfusion (Blood bank: autol = 104, homol = 111)	0.13	0.31	–0.18 (–0.29 to –0.07)
Mean no. of units of homologous blood replaced per patient [SD] (Blood bank: autol = 104, homol = 111)	0.34 [0.95]	0.72 [1.27]	–0.38 (–0.69 to –0.08)
Mean patient length of stay (days) [SD]			
(PAS: autol = 111, homol = 112)	17.20 [11.82]	16.16 [11.94]	1.04 (–2.10 to 4.17)
(PEDW: autol = 77, homol = 85)	18.75 [12.31]	18.20 [8.81]	0.55 (–2.75 to 3.85)
Mean no. of readmissions per patient [SD]			
(PAS: autol = 111, homol = 112)	0.36 [0.73]	0.21 [0.76]	0.15 (–0.05 to 3.44)
(PEDW: autol = 77, homol = 84)	0.29 [0.57]	0.21 [0.48]	0.07 (–0.11 to 0.25)
Proportion of patients died within 6 months			
(PEDW: autol = 77, homol = 86)	0.01	0	0.01 (–0.01 to 0.04)
Proportion of patients with infection (ICD-10 codes T814, T845–7)			
(PAS: autol = 111, homol = 112)	0.05	0.05	0 (–0.05 to 0.05)
(PEDW: autol = 75, homol = 83)	0.03	0.04	0.01 (–0.06 to 0.04)
Proportion of patients with vascular complications following a procedure (T817)			
(PAS: autol = 111, homol = 112)	0.03	0.05	0.02 (–0.07 to 0.03)
(PEDW: autol = 75, homol = 83)	0.03	0.01	0.02 (–0.03 to 0.06)
Proportion of patients with postprocedural disorders of the circulatory system (197)			
(PAS: autol = 111, homol = 112)	0.05	0	0.05 (0.01 to 0.10)
(PEDW: autol = 75, homol = 83)	0.04	0.01	0.03 (–0.02 to 0.08)
Proportion of patients with postprocedural complications other than infection, DVT, pulmonary embolism or postprocedural circulatory disorders			
(PAS: autol = 111, homol = 112)	0.13	0.13	0.01(–0.09 to 0.09)
(PEDW: autol = 75, homol = 83)	0.11	0.15	–0.04 (–0.14 to 0.07)
Proportion of patients with at least one postprocedural disorder			
(PAS: autol = 111, homol = 112)	0.25	0.21	0.05 (–0.06 to 0.16)
(PEDW: autol = 75, homol = 83)	0.20	0.21	–0.01 (–0.13 to 0.12)
Mean end of study surrogate EuroQol Health State: mobility			
(PAS: autol = 111, homol = 112)	1	1	0[a]
(PEDW: autol = 75, homol = 83)	1	1	0[a]
Mean end of study surrogate EuroQol: Health State: self-care			
(PAS: autol = 111, homol = 112)	1	1	0[a]
(PEDW: autol = 75, homol = 83)	1	1	0[a]
Mean end of study surrogate EuroQol Health State: usual activities			
(PAS: autol = 111, homol = 112)	1	1	0[a]
(PEDW: autol = 75, homol = 83)	1	1	0[a]
Mean end of study surrogate EuroQol Health State: pain/discomfort			
(PAS: autol = 111, homol = 112)	1	1	0[a]
(PEDW: autol = 75, homol = 83)	1	1	0[a]

continued

TABLE 18 *Results from analysis of routine data (cont'd)*

Routine data (*n* by source and group)	Autologous	Homologous	Mean difference between groups (95% CI)
Mean end of study surrogate EuroQol Health State: anxiety/depression			
(PAS: autol = 111, homol = 112)	1	1	0[a]
(PEDW: autol = 75, homol = 83)	1	1	0[a]
Mean haemoglobin level from blood bank, adjusted for gender difference			
Day 1 (autol = 96, homol = 97)	11.88	11.56	0.32 (–0.04 to 0.67)
Day 4 (autol = 89, homol = 91)	10.96	10.72	0.24 (–0.12 to 0.61)
Day 7 (autol = 86, homol = 88)	11.00	10.81	0.19 (–0.17 to 0.56)
Mean cost per patient (£)			
Homologous blood [SD] (Blood bank: autol = 104, homol = 111)	18.23 [51.54]	39.04 [68.99]	–20.82 (–37.27 to –4.36)
Capital and servicing (PEDW: autol = 77, homol = 86)	25.51	0	25.51[a]
Disposables [SD] (PEDW: autol = 77, homol = 86)	86.92 [1.10]	0.99 [1.32]	85.92 (85.55 to 86.30)
Inpatient stay [SD]			
(PAS: autol = 111, homol = 112)	1146 [787]	1077 [795]	69.12 (–138.82 to 278.08)
(PEDW: autol = 77, homol = 85)	1250 [820]	1213 [587]	36.86 (–182.92 to 256.65)
Readmission [SD]			
(PAS: autol = 111, homol = 111)	24.01 [49.03]	14.41 [51.14]	9.60 (–3.65 to 22.86)
(PEDW: autol = 77, homol = 84)	18.17 [38.42]	13.48 [32.35]	4.68 (–6.34 to 15.71)
Total NHS Secondary Care [SD]			
(PAS: autol = 101, homol = 103)	1322 [870]	1154 [866]	168 (–71.55 to 408.00)
(PEDW: autol = 76, homol = 83)	1395 [851]	1274 [626]	121 (–111.01 to 354.78)

[a] CI not appropriate as no SD.
autol: autologous; homol: homologous.

(autologous: 77; homologous: 86) within the study period. According to PEDW, there were 31 patients with more than one inpatient spell (autologous: 17; homologous: 14) within the study period. The variables compared and the results using data from these episodes and transfusion data from the blood bank, are shown in *Table 18*.

The differences between findings of the different routine data sources were small. There was a significant difference in the number of patients with postprocedural disorder favouring the homologous group according to PAS. This difference did not reach statistical significance according to PEDW. No significant differences between the study groups were found in mean LOS, number of readmissions, mortality or other adverse events.

There were no significant differences between the study groups in HRQL scores using ICD-10 terms as surrogates for EuroQol dimensions. In the absence of ICD-10 terms to the contrary, all patients were categorised as having perfect HRQoL as measured on five EuroQol scales (EuroQol state of 1,1,1,1,1). This would not be expected, given that these patients had TKR and could be inferred as having a history of mobility problems and some pain or discomfort, which might remain unresolved even after the operation. Furthermore, since over 20% of patients had at least one postprocedural disorder, it was unlikely that the operation restored patients to perfect health immediately. It was likely that symptoms and signs related to the five EuroQol dimensions were grossly under-recorded.

According to blood bank data, there was no significant difference in haemoglobin levels on day 1, 4 or 7 after the operation between the study groups, but there was a significantly lower homologous blood transfusion requirement in the autologous group. There were no other available routine data to triangulate this finding.

There was a significantly lower cost for donor blood for the autologous group, but the costs of capital servicing and disposables were higher. There was no significant difference in the cost of hospital stay between the study groups.

TABLE 19(a) *Differences in morbidity: comparison between designed and routine data*

	Mean difference between groups (95% CI)			
Outcome variable	**Designed data**	**Source (*n*)**	**Routine data**	**Source (*n*)**
LOS from operation (days)	0.66 (–1.39 to 2.71)	Directly monitored (231)	1.04 (–2.10 to 4.17)[a] 0.55 (–2.80 to 3.91)[a]	PAS (223) PEDW (162)
Postoperative wound healing time (days)	0.49 (–1.38 to 2.35)	Directly monitored (230)	None	Not available
Units of homologous blood replaced	–0.34 (–0.57 to –0.12)	Directly monitored (231)	–0.38 (–0.69 to -0.08)	Blood bank (215)
Proportion of patients with relevant postoperative	–0.44 (–0.76 to –0.12)	Directly monitored (231)	0.15 (–0.05 to 3.44)[b] 0.07 (–0.11 to 0.25)[b]	PAS (223) PEDW (161)
Proportion of patients died within 3 months	0.009 (–0.04 to 0.06)	Directly monitored (231)	0.01 (–0.01 to 0.04)	PEDW (163)
Proportion of patients with immediate postoperative adverse events	–0.09 (–0.22 to 0.04)	Directly monitored (231)	0.05 (–0.06 to 0.16) 0.01 (–0.13 to 0.12)	PAS (223) PEDW (158)
Proportion of patients with infective complications (based on original assigned group)	–0.13 (–0.24 to –0.02)	Directly monitored (231)	0 (–0.05 to 0.05) 0.01 (–0.06 to 0.04)	PAS (223) PEDW (158)
Proportion of patients with infective complications (all patients receiving homologous blood placed in homologous blood group)	–0.15 (–0.26 to –0.04)	Directly monitored (219)	Not evaluated because of principle of pragmatic trial	
Proportion of patients with relevant GP contacts	–0.12 (–0.22 to –0.03)	Directly monitored (106)	None	No appropriate surrogate
Postoperative day 1 haemoglobin level (g/dl)	0.38 (–0.003 to 0.77)	Directly monitored (194)	0.32 (–0.04 to 0.67)[c]	Blood bank (193)
Postoperative day 4 haemoglobin level (g/dl)	0.27 (–0.17 to 0.71)	Directly monitored (143)	0.24 (–0.12 to 0.91)[c]	Blood bank (180)
Postoperative day 7 haemoglobin level (g/dl)	0.33 (-0.06 to 0.72)	Directly monitored (171)	0.19 (–0.17 to 0.56)[c]	Blood bank (174)

[a] The routine study included the period before the operation.
[b] The routine study included all readmissions, not just those considered to be operation related.
[c] Adjusted by gender differences.

TABLE 19(b) *Differences in HRQoL: comparison between designed data and routine data*

	Mean difference between groups (95% CI)			
Outcome variable	**Designed data**	**Source (*n*)**	**Routine data**	**Source (*n*)**
EuroQol preoperatively	–0.015 (–0.09 to 0.07)	Patient reported (228)	No appropriate surrogate	
EuroQol at 1 week	0.032 (–0.05 to 0.12)	Patient reported (229)	No appropriate surrogate	
EuroQol at 4 weeks	–0.003 (–0.08 to 0.08)	Patient reported (196)	No appropriate surrogate	
EuroQol at 3 months	0.02 (–0.05 to 0.09)	Patient reported (161)	0 (CI not appropriate)[a]	PAS (223) PEDW (158)
3 months health improvement: EuroQol score	0.02 (–0.08 to 0.12)	Patient reported (160)	No appropriate surrogate	

continued

TABLE 19(b) *Differences in HRQoL: comparison between designed data and routine data (cont'd)*

	Mean difference between groups (95% CI)			
Outcome variable	**Designed data**	**Source (*n*)**	**Routine data**	**Source (*n*)**
VAS preoperatively	0.01 (–0.04 to 0.06)	Patient reported (228)	No appropriate surrogate	
VAS at 1 week	0.02 (–0.02 to 0.07)	Patient reported (222)	No appropriate surrogate	
VAS at 4 weeks	0.02 (–0.03 to 0.07)	Patient reported (200)	No appropriate surrogate	
VAS at 3 months	–0.01 (–0.06 to 0.04)	Patient reported (157)	No appropriate surrogate	
3 months health improvement: VAS EuroQol score	–0.03 (–0.08 to 0.03)	Patient reported (156)	No appropriate surrogate	

[a] Surrogate EuroQol score for the last patient episode within 6 months of the operation.

TABLE 19(c) *Differences in resource use: comparison between designed and routine data*

	Mean difference between groups (95% CI)			
Costs £	**Designed data**	**Source (*n*)**	**Routine data**	**Source (*n*)**
Homologous blood	–18.58 (–30.70 to –6.45)	NHS Trust (231)	–20.81 (–37.27 to –4.36)	Blood bank (215)
Staff time	52.19 (CI not appropriate as no SD)	Directly monitored (231)	None	
Capital and servicing	25.51 (CI not appropriate as no SD)	NHS Trust (231)	As designed study	
Disposables	83.21 (80.69 to 85.73)	NHS Trust (216)	86.05 (85.74 to 86.36) 88.95 (85.57 to 86.32)	PAS (223) PEDW (162)
Total cost of transfusion	130.06 (119.20 to 140.92)	(216)	93.98 (76.71 to 111.26)[a]	Blood bank data matched with PAS (204)
			84.36 (66.29 to 102.43)[a]	Blood bank data matched with PEDW (162)
Postoperative stay in hospital	44.00 (–92.60 to 180.60)	Directly monitored (231)	69.13 (–139.82 to 278.08)[b] 36.86 (–182.92 to 256.65)[b]	PAS (223) PEDW (160)
Readmission	–24.29 (–51.37 to 2.78)	Directly monitored (231)	9.60 (–3.65 to 22.86)[c] 4.68 (–6.34 to 15.71)[c]	PAS (222) PEDW (161)
Total secondary care cost[d]	124.70 (–3.30 to 252.71)A	(216)	168.23 (–71.54 to 408.00)[a]B	Blood bank data matched with PAS (204)
			121.89 (–111.01 to 354.78)[a]C	Blood bank data matched with PEDW (159)

The routine data assume that all autologous patients had their blood reinfused.
[a] Incomplete estimate; [b] included hospital stay before the operation; [c] included all readmissions, not just those considered to be operation related.
[d] The CI of the total secondary care costs were checked with 1000 replication bootstrap. They were A = –24.00 to 229.32, B = –39.38 to 446.35 and C = –114.11 to 353.02.

TABLE 20 *Reconciliation by research question*

Research question	Definitive conclusions from analysis of designed data	Conclusions from analysis of routine data	Comments
To compare morbidity in terms of LOS, wound healing, readmissions, serious adverse events and mortality	No significant differences in clinical outcome when analysed for LOS, wound healing, serious adverse events or mortality between the two groups 3 months following surgery. Significantly fewer readmissions to hospital and visits to the GP among patients in the autologous blood transfusion group. Infective complications were increased in the homologous blood group	No significant differences between the two study groups in clinical outcome when analysed in terms of postprocedural infection, vascular disorders (pulmonary embolism, DVT), LOS and number of readmissions.	Significantly more patients in the autologous group with postprocedural circulatory disorders according to PAS; not confirmed according to PEDW data
To compare HRQoL using a generic QoL measure (EuroQol)	No significant differences in HRQoL using EuroQol 6 months after surgery	No significant differences in HRQoL between the study groups as assessed by ICD-10 terms surrogating EuroQol dimensions	All patients reported as having perfect HRQoL (EuroQol state of 1,1,1,1,1)
To compare NHS resource use	Total patient costs were £113 higher in the autologous blood group	Patients in the autologous group used significantly fewer NHS resources in terms of donor blood; however, autologous blood cost significantly more in terms of capital, service and disposable cost. No significant difference in hospital LOS	Cost comparisons were incomplete. No available information on staff time, theatre time and time to wound healing

Conclusions from routine data

There were no significant differences between the two study groups in clinical outcome when analysed in terms of postprocedural infection, vascular disorders (pulmonary embolism and DVT), LOS and number of readmissions. There were significantly more patients in the autologous group with postprocedural circulatory disorders according to PAS, which was not confirmed using PEDW data. There were no significant differences in HRQoL between the study groups as assessed by ICD-10 terms as surrogates for EuroQol dimensions. Patients in the autologous group used significantly fewer NHS resources in terms of donor blood but more in capital servicing and disposable costs.

Comparison of designed and routine datasets

Table 19(a) shows good consistency of findings from designed and routine data for most of the variables. There were differences in two variables. The original study found significantly lower postoperative infection rates in the autologous group, but this was not replicated by analysis of the routine data. However, routine data identified significantly more postprocedural circulatory disorders in the homologous group. This could be a Type I error.

Table 19(b) shows that analysis of routine and designed data reached the same conclusions for HRQoL. There were no significant differences in patients' QoL between the two study groups. Findings of the routine data had to be interpreted with care. There was probably under-reporting of the signs and symptoms associated with the EuroQol dimensions. *Table 19(c)* shows that routine data led to similar conclusions to those of designed data for resource use.

Table 20 compares the conclusions reached in answer to the three original research questions. The conclusions were similar for two of the questions, but were at variance with respect to postoperative morbidity. This could be a Type I error. In summary, used with caution, routine data could be used to address the research questions.

Chapter 7

Proxy outcome measures from routine data

Introduction

Changes in HRQoL have become important in assessing clinical effectiveness following interventions. The exemplars have addressed both generic (SF-36,[39] EuroQol[40]) and disease-specific (UKIBDQ,[41] IIQ, UDI[42]) outcomes. The SF-36 has been shown to be valid in measuring QoL in such conditions as back pain, menorrhagia and suspected peptic ulcer.[43] The UKIBDQ[41] has been validated for IBD, and the IIQ and UDI for stress incontinence.[42]

HRQoL measures are not normally routinely administered to patients within the clinical setting and are usually used within the context of designed research studies. These QoL questionnaires are based on determining the presence or severity of patient symptoms and in most instances are completed by the patients. The completed questionnaires are scored to generate a single scale or a number of subscales that can give an indication of the severity of the condition.

Most routine data sources have the provision to record diagnostic codes. Some systems also allow recording of patient symptoms and signs. To assess the potential effectiveness of routine data systems in measuring HRQoL, an attempt was made to develop routine data surrogates for one disease-specific (UKIBDQ) and one generic questionnaire (SF-36) that were used as part of exemplar A (open-access follow-up for IBD, Chapter 3).

The ability to develop a surrogate outcome measure is dependent on the completeness of data kept by the various systems. Because detailed clinical data (including symptoms and signs) have been collected on GeneCIS, these data could be analysed as surrogates for patient-reported outcome. The open-access follow-up study (exemplar A) recorded information on GeneCIS that provided the opportunity to assess the reliability and validity of the surrogate QoL measures. This chapter therefore concentrates on this exemplar only.

Open-access follow-up for IBD

Objective

To evaluate the potential of electronic routine data systems for measuring HRQoL.

Questionnaires

Disease-specific (UKIBDQ) and generic (SF-36) questionnaires were used to measure QoL. These are reproduced in Appendices 5 and 6.

Methods

The data items in the UKIBDQ and SF-36 questionnaires were coded by a clinical coding specialist. They were then used as templates to identify relevant symptoms, signs and diagnostic codes in the routine systems available (see *Tables 21* and *22*). Surrogate subscale and total UKIBDQ scores and subscale SF-36 scores were generated by adding up the presence of relevant symptoms. The calculated baseline surrogate scores were transformed to make them comparable with the original UKIBDQ and SF-36 scores obtained.[44] Cronbach alpha or Kuder–Richardson values were generated for the surrogate UKIBDQ and SF-36 scores to determine the reliability of the new scales.

Data sources

Electronic routine data sources included the PEDW and local patient administration (PAS, PIMS), radiology, pathology, theatre and clinical information systems. Following data extraction it was found that those patients studied at the Morriston site (without the clinical information system GeneCIS) had limited information available, thus making comparisons meaningless. The results are therefore based on 93 patients from the Neath arm of the study with complete data from GeneCIS as well as PEDW and local patient administrative data.

Data analysis

Following transformation of the surrogate item scores, Pearson rank correlation tests were performed to determine whether any relationship existed between the primary and secondary scores. Kuder–Richardson (KR-20) or Cronbach's alpha was calculated for the surrogate UKIBDQ and SF-36 scores to determine the reliability of the new scales.

TABLE 21 *Original UKIBDQ questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
Q1. On how many days over the last two weeks have you had loose or runny bowel movements? (a) None (b) on one or two days (c) on three to seven days (d) on eight to fourteen days (Bowel I)	(b–d) R19.4: Change in bowel habit	(b–d) 19F2 Y30c8: Diarrhoea, 19EA.Y7EfC: Change in bowel habit
Q2. On how many days over the last two weeks have you felt tired? (a) None (b) on one or two days (c) on three to seven days (d) on eight to fourteen days (Systemic)	(b–d) R53: Malaise and fatigue (includes tiredness)	(b–d) Xa96S Y7DO0: Tiredness; R0075: Tiredness; 168Z: Tiredness symptom NOS; 1682: Fatigue; XE0qk Y7DO7: Tired all the time
Q3. In the last two weeks have you felt frustrated? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	None	(b–d) Ua164 Ya5sa: Feeling frustrated
Q4. In the last two weeks, has your bowel condition prevented you from carrying out your work or other normal activities? (a) No, not at all (b) Yes, for one or two days (c) Yes, for three to seven days (d) Yes, for eight to fourteen days (Social)	(b–d) Z73.8: Other problems related to life management difficulty	(b–d) ZV4KB Ym9n2(0){V}: Other problems related to life management difficulty
Q5. On how many days over the last two weeks have you opened your bowels more than three times a day? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel I)	None	(b–d) X76dG Y7Efc: Frequency of bowel action
Q6. On how many days over the last two weeks have you felt full of energy? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Systemic)	None	(b–d) Xa029 Ya0v0: Increased energy
Q7. In the last two weeks did your bowel condition prevent you from going out socially? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d Yes, all of the time (e) Does not apply to me (Social)	(b–d) Z73.2: Lack of relaxation and leisure	(b–d) ZV4K8 YM9mz(O){V}: Lack of relaxation and leisure

continued

TABLE 21 *Original UKIBDQ questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
Q8. On how many days over the last two weeks have your bowels opened accidentally? (a) None (b) on one or two days (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel I)	(b–d) R15: Faecal incontinence; R19.8: Other specified symptoms and signs involving the digestive system and abdomen	(b–d) 3931.Y7GLm: Bowels; occasional accident; XE0rG Y30cX: Incontinent of faeces
Q9. On how many days over the last two weeks have you felt generally unwell? (a) None (b) on one or two days (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Systemic)	(b–d) R53: Malaise and fatigue (includes general physical deterioration)	(b–d) XM06M Y7DNd: Generally unwell; 212B YM005(O){V}: O/E looks ill
Q10. In the last two weeks have you felt the need to keep close to a toilet? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	(b–d) R19.8: Other specified symptoms and signs involving the digestive system and abdomen	(b–d) 397..Y7GM3: Toilet dependency
Q11. In the last two weeks has your bowel condition affected your leisure or sports activities? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (e) Does not apply to me (Social)	(b–d) Z72.3: Lack of physical exercise; Z73.2: Lack of relaxation and leisure	(b–d) ZV4K2 YM9mt (O)[V}: Lack of physical exercise; ZV4K8 YM9mz (O){V}: Lack of relaxation and leisure
Q12. On how many days over the last two weeks have you felt pain in your abdomen? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel II)	(b–d) R10.0: Severe abdominal pain; R10.1: Upper abdomen pain; R10.2: Pelvic/perineal pain; R10.3: Lower abdominal pain; R10.4: Other and unspecified abdominal pain	(b–d) 1969.Y7Cmg: Abdominal pain; 1968 (O): Abdominal discomfort; X75rQ Y7C19: Abdominal tenderness; 1972.Y7CnC: Epigastric pain; X76dT Y7can: Abdominal cramps; IDC5 Y7Ckh (O): Griping pain; 1962.Y7Cne: Abdominal colic; 197B.Y7Cn2: Upper abdominal pain; 197C.Y7Cn1: Lower abdominal pain; 197A Y7Cn4: Generalised abdominal pain
Q13. On how many nights over the last two weeks have you been unable to sleep well (days if you are a shift worker)? (a) None (b) on one or two night only (c) on three to seven nights	(b–d) G47.3: Sleep apnoea; G47.9: Sleep disorder unspecified	(b–d) X76AF Y7DNA: Unable to sleep; XE2Pv Y00RD: Insomnia

continued

TABLE 21 *Original UKIBDQ questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
(d) on eight to fourteen nights (i.e. more than every other night) (Systemic)		
Q14. In the last two weeks have you felt depressed? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	(b–d) F32.9: Depressive episode, unspecified (includes Depression NOS); F32.0, F32.1, F32.2 (if degree of depression known)	(b–d) X00sO YM3yH: Depressed
Q15. In the last two weeks have you had to avoid attending events where there was no toilet close at hand? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Social)	None	(b–d) Xa327 Ya5x3: Tends to avoid group social interactions
Q16. On how many days over the last two weeks have you had a problem with large amounts of wind? (a) None (b) on one or two days (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel II)	(b–d) R14: Flatulence and related conditions	(b–d) 19B2.Y7Eed: Excessive flatulence; 19B2.YMG6m: Full of wind; 19BZ (O): Wind NOS
Q17. On how many days over the last two weeks have you felt off your food? (a) None (b) on one or two days (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Systemic)	(b–d) R63.0: Anorexia (includes loss of appetite)	(b–d) XM07X Y7Ece: Off food; XM07Y Y7Eck: Loss of appetite
Q18. Many patients with bowel problems have worries about their illness. How often during the last two weeks have you felt worried? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	(b–d) R45.2: Unhappiness (includes Worries NOS)	(b–d) 1BK..Yaa2mJ: Worried; XE0rb YM1eG: Anxiousness; 2259 (O): O/E nervous
Q19. On how many days over the last two weeks has your abdomen felt bloated? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel II)	(b–d) R14: Flatulence and related conditions (includes Bloating)	(b–d) Xa1c1 Y7Dyu: Bloated abdomen; 19A3. Y7Cjt : Abdomen feels distended; 1968 (O): Abdominal discomfort; Xa1c1 Y7Dbu: Abdominal swelling; 19A2 (O): Abdominal distension symptom; 19A4.Y7Cjs: Abdomen feels swollen

continued

TABLE 21 *Original UKIBDQ questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
Q20. In the last two weeks have you felt relaxed? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	None	(b–d) Ua16A YMGTF: Feeling relaxed
Q21. In the last two weeks have you been embarrassed by your bowel problem? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	None	(b–d) XM03v Y7DCW: Embarrassing behaviour
Q22. On how many days over the last two weeks have you wanted to go back to the toilet immediately after you thought you had emptied your bowels? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel I)	None	(b–d) 397. Y7GM3: Toilet dependency
Q23. In the last two weeks have you felt upset? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	(b–d) R45.2: Unhappiness	(b–d) Ua168 Ya5sp: Feeling upset
Q24. On how many days over the last two weeks have you had to rush to the toilet? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel)	None	(b–d) 1A25.YMGid: Urgency to pass urine
Q25. In the last two weeks have you felt angry as a result of your bowel problem? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	(b–d) R45.4: Irritability and anger	(b–d) XM016.Ya5sd: Feeling angry
Q26. In the last two weeks, has your sex life been affected by your bowel problem? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (e) Does not apply to me (Social)	(b)-d) F52.0: Lack or loss of sexual desire; F52.1: Sexual aversion and lack of sexual enjoyment	(b–d) X766u Y7Dfn: Sexual aversion; Ua1sq YabUS: Decreased sexual function

continued

TABLE 21 *Original UKIBDQ questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
Q27. On how many days over the last two weeks have you felt sick? (a) None (b) on one or two days only (c) on three to seven days (d) on eight to fourteen days (i.e. more than every other day) (Bowel II)	(b–d) R11: Nausea and vomiting	(b–d) Xa1pJ Ya3Wx: Nausea and vomiting
Q28. In the last two weeks have you felt irritable? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	(b–d) R45.4: Irritability and anger	(b–d) XE0rc Ya5sf: Feeling irritable
Q29. In the last two weeks have you felt lack of sympathy from others? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	None	None
Q30. In the last two weeks have you felt happy? (a) No, not at all (b) Yes, some of the time (c) Yes, most of the time (d) Yes, all of the time (Emotion)	None	(b–d) Ua1X9 YMEBI: Happy[a]

For the above questions 0 = symptom not recorded, 1 = symptom recorded, apart from [a] 1 = symptom not recorded, 0 = symptom recorded.
A lower UKIBDQ score indicates worse health. The surrogate measures adopt the same convention, i.e. a lower score indicating worse health.

TABLE 22 *Original SF-36 questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
Q1. In general would you say your health is:		
excellent	None	
very good	None	XE0sP Y7DNa (O): Patient feels well; Xa96k Y7DNX: Fit and well (Score 1)
good	None	None
fair	Symptom not recorded (Score 2)	Symptom not recorded (Score 2)
poor	R53: Malaise and fatigue (includes General physical deterioration) (Score 3)	Xa96i Ya36i: Looks ill (Score 3)
(General Health)		
Q2. Compared to one year ago how would you rate your general health now?		
much better now than one year ago	None	None
somewhat better now than one year ago	None	XE1gy YM0Zd: Patient's condition improved (Score 1)
about the same as one year ago	Symptom not recorded (Score 2)	2128.YM0Zk: Patient's condition the same, Symptom not recorded (Score 2)
somewhat worse than one year ago	R53: Malaise and fatigue (includes General physical deterioration) (Score 3)	Xa35q Ya5Jq: General health deterioration (Score 3)
much worse than one year ago	None	None
(Health Change)		
Q3. The following items are about activities you might do in a typical day. Does your health now limit you in these activities. If so, how much? (1 = yes limited a lot, 2 = yes limited a little, 3 = no not limited at all)		
(a) vigorous activities, such as running, lifting heavy objects, participating in strenuous sports	(a–f) Z73.6: Limitation of activities due to disability (Score 1) Symptom not recorded (Score 0)	(a) Xa26n Ya405: Difficulty running (Score 1) Symptom not recorded (Score 0)
(b) moderate activities, such as moving a table, pushing a vacuum cleaner, bowling or playing golf		(b) Xa3Ng Ya5iV: Difficulty cleaning room (Score 1) Symptom not recorded (Score 0)
(c) lifting or carrying groceries		(c) Xa45A Ya6Ur: Difficulty lifting; Xa7n4 YaV52: Difficulty performing shopping activities (Score 1) Symptom not recorded (Score 0)
(d) climbing several flights of stairs		(d, e) Xa21W Ya3tC: Difficulty climbing stairs (Score 1) Symptom not recorded (Score 0)
(e) climbing one flight of stairs		
(f) bending, kneeling or stooping		(f) X76mV Y7Ex2: Difficulty bending (Score 1) Symptom not recorded (Score 0)
(g) walking more than a mile	(g–i) R26.2: Difficulty in walking, not elsewhere classified (Score 1) Symptom not recorded (Score 0)	(g–i) N097.Y7Deb: Difficulty in walking (Score 1) Symptom not recorded (Score 0)
(h) walking half a mile		
(i) walking 100 yards		
(j) bathing and dressing yourself	(j) Z74.1: Need for assistance with personal care (Score 1) Symptom not recorded (Score 0)	(j) Xa22ud Ya54F: Difficulty bathing self; Xa2xi Ya57R: Difficulty dressing (Score 1) Symptom not recorded (Score 0)
(Physical Functioning)		

continued

TABLE 22 *Original SF-36 questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
Q4. During the past 4 weeks have you had any of the following problems with your work or other regular daily activities as a result of your physical health (1 = yes, 2 = no)		
(a) cut down on the amount of time you spent on work or other activities (b) accomplished less than you would like (c) were limited in the kind of work or other activities (d) had difficulties performing the work or other activities (for example, it took extra effort) (Role – Physical)	(a–d) Z73.6: Limitation of activities due to disability (Score 1) Symptom not recorded (Score 0)	(a–d) Xa844 YaVcu: Difficulty performing labouring activities (Score 1) Symptom not recorded (Score 0)
Q5. During the past 4 weeks have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious) (1 = yes, 2 = no)		
(a) cut down on the amount of time you spent on work or other activities (b) accomplished less than you would like (c) didn't do work or other activities as carefully as usual (Role – Emotional)	(a–c) R45.8: Other symptoms and signs involving emotional state (Score 1) Symptom not recorded (Score 0)	(a–c) Xa844 YaVcu: Difficulty performing labouring activities; X00SO YM3yH: Depressed; XE0rb Ya5sj: Feeling anxious (Score 1) Symptom not recorded (Score 0)
Q6. During the past 4 weeks to what extent has your physical or emotional problems interfered with your normal social activities with family, friends, neighbours or groups? not at all slightly moderately quite a bit extremely (Social Functioning)	Z73.3: Stress, not elsewhere classified (includes Physical and mental strain NOS) (Score 1) Symptom not recorded (Score 0)	Ua18k YMFTA: Emotional problems (Score 1) Symptom not recorded (Score 0)
Q7. How much bodily pain have you had in the past 4 weeks? none very mild moderate severe very severe (Bodily Pain)	R52.9: Pain, unspecified (includes generalised pain NOS); R52: Pain, not elsewhere classified; R52.0: Acute pain; R52.1: Chronic intractable pain; R52.2: Other chronic pain (Score 1) Symptom not recorded (Score 0)	1971.Y7CnD: Central abdominal pain; 1962.Y7Cne: Abdominal colic; 1972.Y7CnC: Epigastric pain; 197A.Y7Cn4: Generalised abdominal pain; 197C.Y7Cn1: Lower abdominal pain; 1976(O): Right flank pain; 1977.Y7Cn7: Right iliac fossa pain; 1978.Y7Cn6: Left iliac fossa pain; 1979.Y7Cn5: Suprapubic pain; 197B.Y7Cn2: Upper abdominal pain; XE1F4 Y7Coy: Cervicalgia; 2118(O): O/E tenderness/pain; 2252(O): O/E in pain; 197Z(O): Site of GIT pain NOS;

continued

TABLE 22 *Original SF-36 questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
		1D13(O): C/O a pain; 1DC5(O): Griping pain; XE1Fp bY7CqA: Pain in limb (Score 1) Symptom not recorded (Score 0)
Q8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? not at all a little bit moderately quite a bit extremely (Bodily Pain)	Z73.6: Limitation of activities due to disability; Z73.8: Other problems related to life-management difficulty (Score 1) Symptom not recorded (Score 0)	Xa844 YaVcu: Difficulty performing labouring activities (Score 1) Symptom not recorded (Score 0)
Q9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes the closest to the way you have been feeling. How much of the time during the past 4 weeks (1 = all the time, 2 = most of the time, 3 = a good bit of the time, 4 = some of the time, 5 = a little of the time, 6 = none of the time)		
(a) did you feel full of life (Vitality)	(a) None	(a) Xa029 Ya0v0: Increased energy (Score 0) Symptom not recorded (Score 1)
(b) have you been a very nervous person (Social Function)	(b) R45.0: Nervousness (Score 1) Symptom not recorded (Score 0)	(b) XE0ra Y7D7v: Nervousness; XE0rb YM1eG: Anxiousness (Score 1) Symptom not recorded (Score 0)
(c) have you felt so down in the dumps that nothing could cheer you up (Mental Health)	(c) R45.2: Unhappiness (Score 1) Symptom not recorded (Score 0)	(c) X00sO YM3yH: Depressed (Score 1) Symptom not recorded (Score 0)
(d) have you felt calm and peaceful (Mental Health)	(d) None	(d) Ua16A YMFR0: Feeling calm (Score 0) Symptom not recorded (Score 1)
(e) did you have a lot of energy (Vitality)	(e) None	(e) Xa029 Ya0v0: Increased energy (Score 0) Symptom not recorded (Score 1)
(f) have you felt downhearted and low (Mental Health)	(f) R45.2: Unhappiness; F32.9: Depressive episode, unspecified (includes Depression NOS) (Score 1) Symptom not recorded (Score 0)	(f) X00sO YM3yH: Depressed (Score 1) Symptom not recorded (Score 0)
(g) did you feel worn out (Vitality)	(g) R53: Malaise and fatigue (includes Tiredness and General physical deterioration) (Score 1) Symptom not recorded (Score 0)	(g) X76Ae YMGTP: Worn out; 1682.YM1eB: Fatigue (Score 1) Symptom not recorded (Score 0)
(h) have you been a happy person (Mental Health)	(h) None	(h) None

continued

TABLE 22 *Original SF-36 questions with comparable coded symptoms, signs and diagnoses from all electronic routine data sources (cont'd)*

Original question (subscale)	Comparable ICD-10 code	Comparable READ v3 code
(i) did you feel tired (Vitality)	(i) R53: Malaise and fatigue (includes Tiredness) (Score 1) Symptom not recorded (Score 0)	(i) 1682.YM1eB: Fatigue (Score 1) Symptom not recorded (Score 0)
Q10. During the past 4 weeks, how much of the time have your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)? all of the time most of the time some of the time a little of the time none of the time (Social Functioning)	Z73.3: Stress, not elsewhere classified (includes Physical and Mental Strain NOS); Z73.9: Problem related to life management difficulty, unspecified) (Score 1) Symptom not recorded (Score 0)	Ua18K YMFTA: Emotional problems (Score 1) Symptom not recorded (Score 0)
Q11. Please choose the answer that best describes how true or false each of the following statements is for you (1 = definitely true, 2 = mostly true, 3 = not sure, 4 = mostly false, 5 = definitely false)		
(a) I seem to get ill more easily than other people (General Health)	(a) R69: Unknown and unspecified causes of morbidity (includes Illness NOS) (Score 1) Symptom not recorded (Score 0)	(a) X76At Ya1hG: Feeling ill (Score 1) Symptom not recorded (Score 0)
(b) I am as healthy as anyone I know (General Health)	(b) None	(b) Xa96k Y7DNX: Fit and well (Score 0) Symptom not recorded (Score 1)
(c) I expect my health to get worse (General Health)	(c) R53: Malaise and fatigue (Includes General physical deterioration) (Score 1) Symptom not recorded (Score 0)	(c) Xa35q Ya5Jq: General health deterioration (Score 1) Symptom not recorded (Score 0)
(d) My health is excellent (General Health)	(d) None	(d) None

A lower SF-36 score indicates worse health.
The surrogate measures were reversed and transformed to adopt the same convention, i.e. a lower score indicating worse health.

Results

Twenty-nine of the 30 UKIBDQ and 34 of the 36 SF-36 questions had a comparable symptom, sign or diagnostic codes in the routine data systems. Although 29 questions were codable on the UKIBDQ, in practice routine clinical recording only captured ten of the potential 29 symptoms. Similarly, only eight of the 34 potentially codable SF-36 questions had recorded symptoms.

Tables 23 and *24* illustrate the descriptive statistics for the original and surrogate UKIBDQ and SF-36 scores. The surrogate UKIBDQ scores were consistently higher than the original UKIBDQ subscales and total scores. The surrogate SF-36 scores were generally higher than the original SF-36 subscale scores, with the exception of vitality and mental health.

Table 25 illustrates the level of correlation between the primary and secondary item scores for the UKIBDQ and SF-36 questionnaires.

Three of the four calculated surrogate UKIBDQ subscales were significantly correlated with the corresponding primary subscale scores and the total IBDQ score was significantly related to the

TABLE 23 *Descriptive statistics for the original and surrogate subscale and total UKIBDQ scores*

	Surrogate	**Original**
Emotional Function	Mean = 97.3 95% CI = 96.1 to 98.5 SD = 5.8 Min = 75 Max = 100 No. of items with recorded symptoms = 3	Mean = 72 95% CI = 68.7 to 75.3 SD = 16.1 Min = 33.3 Max = 100 No. of items with recorded symptoms = 11
Bowel Function I	Mean = 97.1 95% CI = 95.6 to 98.6 SD = 7.2 Min = 66.7 Max = 100 No. of items with recorded symptoms = 2	Mean = 66 95% CI = 61.1 to 70.8 SD = 23.9 Min = 11.1 Max = 100 No. of items with recorded symptoms = 5
Bowel Function II	Mean = 82.5 95% CI = 78 to 87 SD = 21.7 Min = 0 Max = 100 No. of items with recorded symptoms = 4	Mean = 65.4 95% CI = 59.9 to 70.9 SD = 26.6 Min = 0 Max = 100 No. of items with recorded symptoms = 4
Systemic Function	Mean = 97.6 95% CI = 96.1 to 99.1 SD = 7.4 Min = 75 Max = 100 No. of items with recorded symptoms = 1	Mean = 52.7 95% CI = 47 to 58.3 SD = 27.4 Min = 0 Max = 100 No. of items with recorded symptoms =5
Social Function	No surrogate available	Mean = 83 95% CI = 78.6 to 87.4 SD = 21.4 Min = 14.3 Max = 100 No. of items with recorded symptoms = 5
Total UKIBDQ	Mean = 94.9 95% CI = 93.7 to 96.1 SD = 5.9 Min = 71.7 Max = 100 No. of items with recorded symptoms = 10	Mean = 69.9 95% CI = 66.1 to 73.7 SD = 18.2 Min = 30 Max = 96.7 No. of items with recorded symptoms = 30

primary total score. A standardised item KR-20 score of 0.51 was calculated for the surrogate total scale based on the 29 comparable scored questions.

Of the six surrogate SF-36 subscale items, three showed a significant correlation with the primary subscale items. Cronbach alpha or Kuder–Richardson scores could not be calculated for any of the SF-36 surrogate subscales because few surrogate symptoms were recorded. Further analysis is needed to examine the correlations between the surrogate and original scores.

There were no significant differences between the routine and open-access appointment groups with respect to the surrogate UKIBDQ and SF-36 scores. This mirrors the results of the original study.

Discussion

This chapter explores a unique approach to the measurement of patient outcome. This was achieved in the context of this study because of the availability of detailed clinical data routinely collected and stored in an operational clinical information system.

Clinical gastroenterology at Neath is supported by GeneCIS, which is used routinely to record

TABLE 24 *Descriptive statistics for the original and surrogate SF-36 subscales*

	Surrogate	**Original**
General Health	Mean = 64.2 95% CI = 63.3 to 65.2 SD = 4.7 Min = 50 Max = 75 No. of items with recorded symptoms = 1	Mean = 51 95% CI = 45.7 to 56.4 SD = 25.9 Min = 5 Max = 100 No. of items with recorded symptoms = 5
Physical Function	No surrogate available	Mean = 72.6 95% CI = 67 to 78.2 SD = 27.4 Min = 0 Max = 100 No. of items with recorded symptoms = 10
Role Physical	No surrogate available	Mean = 54.8 95% CI = 45.7 to 63.9 SD = 44.3 Min = 0 Max = 100 No. of items with recorded symptoms = 4
Role Emotional	Mean = 92.6 95% CI = 87.1 to 98 SD = 26.4 Min = 0 Max = 100 No. of items with recorded symptoms = 1	Mean = 70.9 95% CI = 62.4 to 79.4 SD = 41.5 Min = 0 Max = 100 No. of items with recorded symptoms = 3
Social Function	Mean = 98.2 95% CI = 96.7 to 99.8 SD = 7.52 Min = 66.7 Max = 100 No. of items with recorded symptoms = 1	Mean = 73.8 95% CI = 67.9 to 79.7 SD = 28.6 Min = 0 Max = 100 No. of items with recorded symptoms = 3
Bodily Pain	Mean = 76.1 95% CI = 70.9 to 81.2 SD = 25.1 Min = 50 Max = 100 No. of items with recorded symptoms = 1	Mean = 61.2 95% CI = 56.2 to 66.1 SD = 24.3 Min = 22 Max = 90 No. of items with recorded symptoms = 2
Vitality	Mean = 44.7 95% CI = 41.5 to 47.9 SD = 15.5 Min = 0 Max = 50 No. of items with recorded symptoms = 2	Mean = 47.1 95% CI = 41 to 52.2 SD = 24.8 Min = 0 Max = 90 No. of items with recorded symptoms = 4
Mental Health	Mean = 63.8 95% CI = 61.1 to 66.6 SD = 13.5 Min = 0 Max = 66.7 No. of items with recorded symptoms = 2	Mean = 66.4 95% CI = 62.3 to 70.5 SD = 20 Min = 4 Max = 96 No. of items with recorded symptoms = 4

TABLE 25 *Level of correlation between primary and secondary item scores*

Item		Pearson correlation (*r*)	*p* Value
UKIBDQ	Bowel I	0.16	0.12
	Bowel II	0.26	0.01*
	Emotion	0.25	0.02*
	Systemic	0.24	0.02*
	Social	No surrogate	
	Total	0.38	0.00*
SF-36	General Health	0.01	0.95
	Social Function	0.14	0.19
	Vitality	0.28	0.01*
	Mental Health	0.30	0.00*
	Bodily Pain	0.31	0.00*
	Physical Function	No surrogate	
	Role Physical	No surrogate	
	Role Emotional	0.10	0.37
	Health Change	No surrogate	

* $p < 0.05$.

symptoms and signs in coded form, as well as diagnoses and procedures. The system uses a clinically rich terminology for coding (Clinical Terms version 3[45]), which includes many terms that correspond to the questions in the UKIBDQ and SF-36. For the patients in exemplar A about one-third of the symptoms identified in these QoL questionnaires had been collected routinely. This reflects clinical practice and it is clear that a more complete dataset could be collected in the course of routine clinical practice if required in the context of an RCT. More standard hospital administrative systems record diagnoses only, using ICD-10. Although many of the questions in the UKIBDQ and SF-36 questionnaires have a corresponding term and code in ICD-10, these are clinically unwieldly and rarely recorded in coded form. Routine systems that do not use a clinically rich terminology will therefore miss information that could be useful in measuring QoL.

Questions that had a corresponding term and code were generally scored as present if the patient exhibited this code or absent if the code was not present. One flaw with this approach is that absence of evidence does not equate to evidence of absence. In other words, it cannot be assumed that a patient did not have a particular symptom just because it was not recorded in the system. Only ten of the 30 UKIBDQ and eight of the 36 SF-36 questions had symptoms recorded as being present in the patients in the study. This reflects routine clinical practice and the unstructured collection of data in the follow-up outpatient setting. Routine data will therefore consistently under-record symptoms that might contribute to the assessment of HRQoL, even when captured in a clinically rich information system. In spite of this, the surrogate UKIBDQ total score was shown to be significantly correlated with the original primary UKIBDQ score. The surrogate scale was also shown to be fairly reliable, with a Kuder–Richardson value of 0.51. This compares reasonably with the UK validation study.[41] Although a Cronbach alpha of 0.94 was reported, this is arguably too high, suggesting some duplication of items. Although symptoms were only recorded as present or absent, it appears that the clinical information system GeneCIS may be useful in measuring QoL changes in disease-specific conditions such as IBD.

In contrast, although three of the five transformed surrogate subscales were significantly correlated with the original SF-36 items, the calculated Cronbach alpha values could not be calculated, as symptoms were absent for the large majority of questions. As the SF-36 questionnaire is a generic questionnaire, this may account for the limited number of questions with symptoms recorded as present. Used retrospectively, routine data systems appear to have less value in assessing generic QoL changes following interventions than condition-specific changes.

Conclusions

This chapter has shown that where clinical data are collected in detail in structured form (such as

in clinical information systems, e.g. GeneCIS) it is possible to develop surrogate measures for patient outcomes. Where only diagnoses and procedures are collected (such as on hospital administrative and PEDW data), this does not appear to be possible. This is due not to the lack of available codes for generic or disease-specific questions, but to the under-utilisation of the terms and codes within routine data systems. Both the surrogate UKIBDQ and SF-36 scores were consistently milder than the original scores, indicating an under-reporting of symptoms. This under-reporting had little effect within this study as it was designed as an RCT and under-reporting occurred across both groups. However, future prospective studies with non-random controlled designs could be biased by the use of routine data to measure HRQoL.

In a prospective trial it would be possible to identify and record routinely the presence or absence of all symptoms and signs that are normally recorded in disease-specific or generic questionnaires. This could eliminate the need for patient-administered questionnaires and the associated difficulties such as patient comprehension and poor response rates.

The danger is that not only are routine data collectors not blind to treatment received, but also they often have a vested interest in the success of the experimental regimen. Thus, a randomised trial using routine data might fail to comply with the CONSORT guidelines.[46] The validity of such adaptations in measuring QoL needs to be explored further. It is also possible that validated measures (e.g. the UKIBDQ or SF-36) could be routinely completed by patients when attending for consultation or admission and held on a clinical system such as GeneCIS. The feasibility and cost of this should be explored.

Chapter 8

Health economics: resource implications of using routine data in health technology assessment

Aims

1 To estimate the costs of data collection and analysis in the four exemplars using designed data.
2 To estimate the cost of data collection and analysis in the four exemplars using routine data.
3 To compare the difference in costs between designed and routine data with differences in their ability to answer the study questions.
4 To estimate the marginal value of the benefits of designed data, as shown by 'willingness to pay' (WTP).

Principles

The present analysis addresses **differences** in costs between the two methods, that is, the extra cost or cost saving of conducting each study using routine rather than designed data. In theory, the cost of all activities from initial design of the study to writing up results could be affected by the data method used. For practical purposes, however, all study costs apart from those directly related to data handling have been ignored. For both types of data there are five steps.

Step 1: determining which data can be obtained from routine sources

The four exemplars used designed data only where the required data were not available routinely. Accordingly, the cost of identifying routine data should not differ between methods.

Step 2: designing data or modifying routine data

Where routine data were not available, each study designed *ad hoc* data to be collected within the trial. Although this activity is specifically excluded from routine data studies, it was nevertheless often possible to manipulate routine data to produce surrogates for designed data. As long as these were developed solely from routine sources they conform to the definition of a routine data health technology assessment. Costs again depend on specific circumstances, but it is likely that these will be lower for the routine data studies.

Step 3: extracting data

Costs vary with circumstances.

Step 4: validation of data

Designed data can be regarded as the gold standard in the sense that they can capture the most appropriate data relevant to the specific study question and make use of tools that have already been validated for use in particular circumstances. Accordingly, their validity is likely to be greater than data collected for other (routine) purposes. Routine data need more time and effort to validate the data and ensure internal and external consistency. Costs are thus likely to be higher for routine data studies.

Step 5: data input and analysis

The costs of entering and analysing each data item are likely to be similar providing that the same protocols are followed, for example, the same statistical tests are applied. However, the number of data items is likely to be higher for the primary analysis (e.g. certain patient-borne costs can be obtained only from patients as designed data). Hence, the costs of secondary analysis are likely to be lower. Thus, the total costs of routine data health technology assessments are generally lower than designed data health technology assessments. *Table 26* summarises the anticipated direction of costs for each element.

Method for aim 1: estimating the costs of designed data

A top-down approach was used. The research staff involved in each study were asked to provide best

TABLE 26 *Relative burden of research activity*

Stage	Routine data studies
1. Determining sources of routine data	Same
2. Designing data	Lower
3. Extraction	Generally lower
4. Validation	Generally higher
5. Input and analysis	Lower

guesses of what proportion of the total time spent on the study had been devoted to each of the following activities:

- project management
- literature review
- data extraction
- data validation
- statistical analysis
- reporting
- meetings
- travel
- correspondence
- unassigned.

The cost of each activity was then estimated by apportioning the total time that each researcher had been funded to work on the project between these activities (assuming a 37.5-hour working week and a 42-week working year). However, the total time apportioned to the various activities did not include the time of some staff who contributed to the projects but who were funded from other sources. Accordingly, to reflect the cost of the designed data studies (i.e. the value of the resources used) more accurately, the hours of these non-funded staff were added where possible. To allow comparison with the routine data costs, staff time was valued at current hourly gross employment costs for the relevant grade.

In exemplar A, however, a detailed record of the time spent on different activities had been kept. This permitted a more accurate bottom-up approach.

Two sets of sensitivity analyses were performed on these data:

- *Varying designed data task times.* The time spent on the data aspects of the designed data studies was based on the length of time that individuals had been funded for each study. Research proposals cannot estimate these costs precisely and so estimates are likely to be rounded. For example, an estimated 11 months of work is likely to lead to a request for funding for a full year. Moreover, staff employed on any project come in discrete units (whole time equivalents or fractions thereof) rather than as hourly paid staff. Sensitivity analyses have thus been performed on the time required for designed data analyses.
- *Varying the grade of staff employed.* In the base case analyses, staff costs were determined using the actual staff involved in the studies (*Table 27*). As some of these tasks could be done by less costly staff, a series of sensitivity analyses was performed.

Method for aim 2: estimating the costs of routine data

A bottom-up approach was used. The time spent in replicating the exemplars using routine data was monitored prospectively via timing sheets. The time of the analysts plus that of any other professionals involved was valued using the hourly gross employment costs (*Table 27*).

TABLE 27 *Unit costs*

Staff	Cost/hour (£)	Source
GP	33.87	Netten *et al.* (1999)[47]
Research assistant		University Personnel Department
Junior	10.14	
Senior	18.37	
Research nurse		Department of Health Internet website
E grade	12.34	
G grade	16.59	
H grade	20.61	
Project manager	as H grade nurse	
Statistician		University Personnel Department
Junior	20.27	
Senior	27.88	
Doctor		Bro Morgannwg NHS Trust Finance Department
Senior house officer	16.00	
Registrar	21.97	
Consultant		Bro Morgannwg NHS Trust Finance Department
Junior	36.33	
Senior	47.27	

As with aim 1 sensitivity analyses were used, which varied the grade of staff employed.

Method for aim 3: assessment of costs against effectiveness

Differences in costs were compared with differences in results, as reported in Chapters 3–6, to identify dominance. Dominance occurs when one alternative is both more costly and less effective than another, or vice versa, in which case it is unambiguously more (less) cost-effective. If a more costly version is also more effective, economic evaluations often calculate an incremental cost-effectiveness ratio for comparison with other studies that measure effectiveness in the same way (e.g. life years gained). In the present example, however, effectiveness was not measured in a way that would allow comparison with other studies and calculation of incremental cost-effectiveness ratios would be of little value. Instead, a more complex cost–benefit approach was adopted.

Method for aim 4: estimating the marginal value of designed data

There are several difficulties in assessing whether either analysis 'answered the study question'. For example, the IBD study (exemplar A), using designed data, assessed health outcomes using two measures that had both been previously validated for patients of this type. The routine data counterpart constructed a surrogate health outcome measure with similarities to SF-36. In the event, neither analysis showed statistically significant differences between intervention and control groups. Thus, both 'answered the study question'. However, the designed measures were clearly of greater validity. They could well have identified significant differences not captured by the surrogate measure.

Another problem is completeness. Routine data were often available only for some study patients. For example, in the case of the IBD study (exemplar A), GeneCIS data were available at only one of the two study sites (representing 54% of the total sample), while PAS data were available only at the other sites (46% of the total sample), raising concerns about lack of power. In other studies summary variables such as 'total cost' included more items in the designed study (e.g. time off work) than in the routine data study.

Decisions on whether or not to pay for better studies are made by research funding bodies which, in reality, often do have to choose between proposals that address the same study questions but vary in terms of both quality and costs. Such choices must be made *ex ante*, on predictions of how much better the higher cost proposals will answer the study questions, and whether this can justify the higher level of funding requested, given the opportunity cost in terms of forgone outputs from other research.

On the assumption that a funding body can be regarded as having the goal of maximising the **value** of research output from finite research money, an attempt was made to estimate the values that such funding bodies would place on better research output, by setting up a mock grants committee with the following composition:

- Chair: Professor of Radiology and former Chair of Grants Committee, Wales Office for Research and Development in Health and Social Care
- Members: Professor of Surgery, Professor of Medicine, Consultant in Public Health Medicine, Professor of Nursing, two Senior Lecturers in General Practice, Senior Lecturer in Health Care Research, Lecturer in Statistics, Senior Lecturer in Health Economics.

The meeting took 4 hours. Members of the research team (lead applicant, two statisticians and two health economists) attended to provide clarification as needed, but did not participate in the valuation exercise. All members of the research group took notes and a tape recorder was switched on at relevant points in the discussion.

On arrival, members were presented with two versions of a proposal: a more costly version [option (a)] using designed data, and a cheaper version [option (b)] using routine data. Sources of data in both versions and estimates of completeness were given. (An example of the information presented is given in Appendix 7.) The exercise was repeated four times, once for each exemplar study.

Since collective decisions are influenced by a number of factors, including the knowledge, eloquence or power of individual members of the group, the choices made do not necessarily reflect the mean of its members. Accordingly, the group was viewed as a single entity making a single (group) decision. However, since information on variations between group members could also be revealing, individual valuations were elicited before the group discussions.

Initially, therefore, we asked members to record their own preferences about which version to fund.

Choosing the more costly option implied that the marginal value of the anticipated better output was at least equal to the difference in costs. Choosing the less costly option implied that the marginal value of the better output was less than this. In an attempt to refine these valuations, members who chose option (a) were also asked to indicate how much cheaper option (b) would have to be to persuade them to change their mind (this would have to be more than the difference in cost between versions), while those who chose option (b) were asked how much extra they would be willing to pay to get the extra output of option (a) (this would have to be less than the difference in costs).

Group discussion then took place. The steps undertaken by individuals were repeated to obtain the group choices and values.

In the second phase of the exercise, members were informed that the more costly option (a) had been funded, but that both analyses had in any case been undertaken. Results of both versions were presented, and the committee was asked (again individually and then collectively) whether, on the basis of these results, the decision to fund the designed data study had been justified and how much extra they would now be willing to pay for the better output of the designed data study. This was done by asking them to assume that the results of the routine data study had been published and were now in the public domain, and asking how much they would be willing to spend to gain the (known) better outputs of the designed data version.

This exercise thus contained elements of both the 'implied value' method of eliciting values of intangible benefits[48] and 'contingent valuation' methods.[49,50]

Results for aims 1–3: designed data versus routine data studies

The costs of the types of data handling for each of the four exemplars are compared in *Tables 28–31*.

The costs of all aspects of data design, collection and analysis were more than three times higher for the designed study. Routine data or surrogates were available for most health outcomes and aspects of resource use. The designed data costs associated with these specific aspects of the study were more than double those for routine data (£19,247 versus

TABLE 28 *Exemplar A (IBD): cost of designed and routine data*

Designed data[a] (Bottom-up approach)				Routine data (Bottom-up approach)			
Task	**Who**	**Hours**	**Cost (£)**	**Task**	**Who**	**Hours**	**Cost (£)**
Data design	Senior statistician	90	2509	Data extraction 1	Senior statistician	102.25	2851
Data validation 1	Senior statistician	18.5	516	Data extraction 2	Junior research assistant	17.5	177
Data validation 2	Senior consultant	15	709	Data validation	Senior statistician	23.25	648
Data input and analysis 1	Senior statistician	324.5	9047	Data input and analysis 1	Senior statistician	65	1812
Data input and analysis 2	Senior consultant	10	473	Data input and analysis 2	Junior research assistant	17.5	177
Qualitative input and analysis 1	Senior statistician	47.5	1324				
Qualitative input and analysis 2	GP	50	1694				
Data management	Research nurse (G)	114	1891				
Interviewing	GP	32	1084				
Total		**701.5**	**£19,247**	**Total**		**225.5**	**£5665**

[a] These include only those costs in the original study that would be affected by the use of routine in place of designed data.

TABLE 29 *Exemplar B (OSA): cost of designed and routine data*

Designed data[a] (Top-down approach)				Routine data (Bottom-up approach)			
Task	**Who**	**Assigned Hours**	**Cost (£)**	**Task**	**Who**	**Hours**	**Cost (£)**
Data design 1	Project manager	246	5070	Data management	Senior statistician	42	1172
Data design 2	Economist	95	1324	Validation	Senior statistician	37	1027
Data input and analysis 1	Senior statistician	71	2000	Data extraction 1	Senior statistician	9	251
Data input and analysis 2	Research nurse (H)	1379	28,421	Data extraction 2	Research assistant	38	386
Data input and analysis 3	Economist	220	3067	Data input and analysis	Senior statistician	72.5	2021
Data input and analysis 4	Junior doctor	638	10,208				
Data management	Project manager	17	350				
Total		**2,666**	**£50,440**	**Total**		**198.5**	**£4857**

[a] These include only those costs in the original study that would be affected by the use of routine in place of designed data.

TABLE 30 *Exemplar C (sling): cost of designed and routine data*

Designed data[a] (Top-down approach)				Routine data (Bottom-up approach)			
Task	**Who**	**Assigned Hours**	**Cost (£)**	**Task**	**Who**	**Hours**	**Cost (£)**
Data design 1	Project manager	131.25	2705	Patient identification	Research Assistant	4	56
Data design 2	Research nurse (G)	262.5	4355	Data management	Senior statistician	3.75	105
Validation	Senior statistician	53.5	1492	Data extraction 1	Senior statistician	5	139
Data extraction	Research nurse (G)	262.5	4355	Data extraction 2	Research assistant	13.5	188
Data input and analysis 1	Senior statistician	105.75	2948	Data input and analysis 1	Research assistant	42	585
Data input and analysis 2	Project manager	393.75	8115	Data input and analysis 2	Senior statistician	52.75	1471
Data input and analysis 3	Research nurse (G)	157.5	2613				
Data management	Project manager	105	2164				
Total		**1471.75**	**£28,747**	**Total**		**121**	**£2544**

[a] These include only those costs in the original study that would be affected by the use of routine in place of designed data.

TABLE 31 *Exemplar D (autologous blood): cost of designed and routine data*

Designed data[a] (Top-down approach)				Routine data (Bottom-up approach)			
Task	**Who**	**Assigned Hours**	**Cost (£)**	**Task**	**Who**	**Hours**	**Cost (£)**
Data design 1	Project manager	63	1298	Patient identification	Research assistant	4	41
Data design 2	Research nurse (G)	79	1311	Data management	Research assistant	12.75	129
Data design 3	Economist	95	1324	Validation	Senior statistician	6.5	181
Validation	Senior statistician	22.5	627	Data extraction	Research assistant	58.25	590
Data input and analysis 1	Senior statistician	55	1533	Data input and analysis 1	Research assistant	5	51
Data input and analysis 2	Project manager	360	7420	Data input and analysis 2	Senior statistician	81	2259
Data input and analysis 3	Research nurse	408.5	6777				
Data input and analysis 4	Economist	220	3067				
Total		**1303**	**£23,357**	**Total**		**167.5**	**£3251**

[a] These include only those costs in the original study that would be affected by the use of routine in place of designed data.

£5665). Both studies provided similar results, but this may have been due at least in part to the fact that designed data showed no difference between groups. It would have been interesting to see whether the routine data could have replicated the results of the designed data had the latter shown significant differences between groups.

Patient and doctor preferences were assessed in the designed data study by means of interviews with GPs and questionnaires to patients. No comparable routine data could be found. The cost of this part of the designed data study (£5993) made up approximately one-third of the total costs associated with data handling.

The aim of the original study was to evaluate whether open-access follow-up of patients with IBD is better than follow-up by routine, booked appointments. In this case, 'better' can be construed as better health outcomes, lower costs, or being preferred by patients or doctors. As the study was concerned with alternative ways of managing patients, as opposed to alternative clinical treatments, patient and doctor preferences are of considerable importance. They become paramount when health outcomes and costs are similar for the two methods of follow-up.

Designed data costs were here estimated using a top-down approach. The costs of designed data were considerably higher than for any of the other studies, both absolutely and in comparison with the routine data. The large difference between the two methods was largely due to the nature of this study, which allowed little substitution of routine for designed data.

Both analyses concluded that home and hospital monitoring were similar in terms of their ability to detect OSA. While these conclusions were the same, the routine data sources consistently overestimated the number of OSA patients detected by either method. The difficulty in distinguishing suspected from confirmed cases is a consistent problem with routine data.

Both analyses concluded that home monitoring is less costly than hospital monitoring, but the routine study underestimated the difference between the two methods by approximately 40%. This was due in part to the fact that the routine data study could not capture all the costs included in the designed study including many elements of staff time and patients' lost productivity due to time off work. Whether or not lost productivity should be included in cost-

TABLE 32 *Sensitivity analyses*

Study	Sensitivity analysis	Revised cost of designed data (£)	Estimated cost of designed data (£)	Cost of routine data (£)
A: IBD	Data input and analysis 1 done by junior statistician	16,778	19,247	5665
	Assigned hours reduced by 20%	15,398		
B: OSA	Data input and analysis 1 done by junior statistician	49,879	50,440	4857
	Data input and analysis 2 done by research assistant	41,252		
	Date input and analysis 4 done by research assistant	40,926		
	Assigned hours reduced by 20%	40,352		
C: Urethral sling	Data input and analysis 1 done by junior statistician	27,943	28,747	2544
	Data extraction, input and analysis 3 done by E grade nurse	26,962		
	Assigned costs reduced by 20%	22,998		
D: Autologous blood	Data input and analysis 1 done by junior statistician	22,939	23,357	3251
	Data input and analysis 3 done by E grade research nurse	21,621		
	Assigned hours reduced by 20%	18,686		

effectiveness studies, however, is a controversial methodological issue.[51]

The cost of the routine data analysis was lower here than for any of the other studies. This study, however, raises doubts as to whether a routine data study can produce meaningful results. It failed to provide surrogates on a wide number of designed data variables, and did not capture the significant difference in pain 3 months after the operation demonstrated with designed data. The difference in skin-to-skin operation time between the two operations was significant in both analyses, at the 5% level for routine and 0.1% for designed data.

The large difference in cost between the two methods was again largely due to the inability of the routine data study to provide surrogates. The primary outcome in terms of wound healing rates could not be obtained from routine data sources.

Similarly, QoL measured using EuroQol (EQ-5D) in the designed data study could not be replicated in the routine data study. Although there were no significant differences in EuroQol scores according to the transfusion method used, the measure did capture the continual improvement in both groups over time following surgery.

The designed data study was a cost–utility analysis calculating differences in quality-adjusted life years (QALYs) between trial groups. Because the QALY captures both length and quality of life in a single utility index it is valuable in informing resource allocation decisions across different healthcare interventions. It is becoming commonplace in economic evaluations carried out alongside clinical trials. As the EQ-5D consists of only a short questionnaire on five health dimensions (three levels each) and a visual analogue scale (VAS) (thermometer scale), it could be routinely completed by patients.

The routine data study also performed poorly in identifying differences in resource use between groups.

Results of the sensitivity analyses are shown in *Table 32*.

Many tasks undertaken by research or clinical staff could have been done by less costly staff. *Table 32* shows that for all studies, the effect of substituting cheaper research staff had little effect on differences in total costs. The greatest potential saving, representing an overall cost reduction of 18%, was made by substituting an E for a G grade nurse for data input and analysis in study B. The other duties of the G grade nurse in that study could not realistically be done by a lower grade nurse. In all studies a 20% reduction in assigned hours would have been associated with a greater fall in overall costs than any staff substitution.

TABLE 33 *Summary of results of the mock grants committee's valuation exercise*

Study	Group decision			Individual decisions	
	Option chosen (£)	Implied value (£)	WTP	Option chosen (*n* = 9)	Median WTP (range) (£)
A: IBD (*ex ante*)	a	>14,282	40,000	8a, 1b	73,886 (2500–97,772)
A: IBD (*ex post*)	a	>14,282	55,000	9a	95,000 (30,000–100,000)
B: OSA (*ex ante*)	b	<45,583	2,500	2a, 7b	10,000 (0–102,664)
B: OSA (*ex post*)	a	>45,583	65,000	7a, 2b	65,000 (22,500–101,495)
C: Sling (*ex ante*)	a	>26,203	70,000	6a, 3b	70,000 (0–110,642)
C: Sling (*ex post*)	a	>26,203	30,000	8a, 1b	26,000 (22,500–37,500)
D: Autologous blood (*ex ante*)[a]	a	>20,106	42,500	4a, 4b[b]	30,000 (0–89,143)
D: Autologous blood (*ex post*)[a]	a	>20,106	30,000	8a	30,000 (10,000–75,000)

[a] Only eight members participated in this round.
[b] Deciding vote [for (a)] cast by chairman.

Results for aim 4: marginal value of better studies

The results of the mock grants committee are shown in *Table 33*.

Study A: IBD

Ex ante *choices and valuations*

The group opted to fund option (a), implying that the value of the anticipated extra benefit from designed data was at least £14,282 (actual cost of £97,772 for option (a) less estimated cost of £83,490 for option (b)). They indicated that they would have chosen option (b) only if it were at least £40,000 cheaper than option (a). Eight of nine members individually chose option (a). Four indicated that they would never fund option (b) regardless of how much cheaper it was, with one stating in a written comment that option (b) would be unethical. Median marginal WTP for option (a) was £73,886 (range £2500–£97,772). (Note that refusal to fund option (b) at any price implies that the anticipated marginal value of output from option (a) is at least equal to the total cost of funding that version.)

Key reasons for choosing option (a) included concerns that (1) changes in symptoms of patients with IBD would not be captured by routine data, (2) one routine data system was in use in only one of the two study centres, which could lead to serious problems of bias and loss of power due to reduced sample size, (3) routine sources were unable to provide data on professional or patient preferences and (4) routine sources may be unable to provide valid data on QoL.

Ex post *choices and valuations:*

Having been presented with the results of both analyses, the group agreed that the decision to fund option (a) had been justified. Group valuation of the extra benefits of option (a) was increased to £55,000. The results had largely confirmed expectations of the quality of routine data. In particular, the absence of significant differences in outcomes and costs increased the value attached to patient and doctor preferences that could not be gleaned by routine data.

Individually, all felt that the decision to fund option (a) had been justified. Median WTP for option (a) was now £95,000 (range £30,000–£100,000).

Study B: OSA

Ex ante *choices and valuations*

The group opted to fund option (b) implying that anticipated marginal value of option (a) was less than its extra cost of £45,583 [actual cost of

£102,664 for option (a) less estimated cost of £57,081 for option (b)]. It would have been willing to fund option (a) only if its cost were no more than £2500 more than option (b). In terms of individual choices, seven of nine members chose option (b), with a median WTP of £10,000 (range 0–£102,664) for the extra benefit from option (a). This was less than its incremental cost. This relatively high median value was due to skewed data caused by the two members who chose option (a), indicating that they would not fund option (b) at any price. These two individuals clearly had little influence in the group valuation.

While recognising that option (a) would provide better data, the collective (and in most cases individual) feeling was that the study questions could be reliably answered using routine data. In particular, it was felt that enough information would be available in the coding systems to compare the diagnostic accuracy of the two systems. Since the focus of the studies was on diagnostic accuracy, the absence of preference data was not considered to be a great loss.

Ex post *choices and valuations*

Having been presented with the results of both analyses, the group felt that the decision to fund option (a), that is, to pay an extra £45,583 (which they did not support) had been justified. They now felt that the extra benefit of option (a) was worth at least £65,000. Individually, seven of nine members agreed with this, with a median WTP of £65,000 (range £22,500–£101,495).

This result was largely due to disappointment with what the routine data study had been able to achieve. Specifically, there was disappointment over the inability of routine data to provide a surrogate for QoL, recognition of the importance of lost productivity, which could not be gleaned from routine data, and serious concerns about the validity of much of the routine data, for example, that the same code was used for both suspected OSA and confirmed OSA.

Study C: urethral sling

Ex ante *choices and valuations*

The group opted to fund option (a), implying that the marginal value of the better study was at least £26,203 (actual cost of £110,642 for option (a) less estimated cost of £84,439 for option (b)). It would have been willing to fund option (a) up to an incremental cost of £70,000. Individually, six of nine members agreed with this decision and median WTP was £70,000 (range 0–£110,642). There was considerable disagreement over the value of preference data, which could only be gleaned from option (a), and the extent to which QoL could be estimated from data on the presence or absence or clinical symptoms.

Ex post *choices and valuations*

Having been presented with the results of both analyses, the group felt that the decision to fund option (a), that is, to pay an extra £26,203, had been justified. They now felt that the extra benefits of option (a) were worth at least £30,000. This lower *ex post* valuation was generally due to a feeling that option (a) had not produced as impressive results as had been anticipated. Individually, eight of nine members agreed that the decision to fund option (a) had been justified, with a median WTP of £26,000 (range £22,500–£37,500).

Study D: autologous blood

Ex ante *choices and valuations*

The group opted to fund option (a), implying that the marginal value of the better study was at least £20,106 (actual cost of £89,143 for option (a) less estimated cost of £69,037 for option (b)). It would have been willing to fund option (a) up to an incremental cost of £42,500. Individually, four of eight members (one had now left) agreed with this decision and the chairman had to cast the deciding vote. Median WTP was £30,000 (range 0–£89,143). This study was the only one where there was no clear majority for either option. Those opting for option (a) were concerned that an anticipated low incidence of postoperative complications was unlikely to be picked up by routine data and were concerned by the absence of health status measures. Those opting for option (b) tended to feel that major outcomes such as differences in length of hospital stay and wound infection rate would be captured by routine systems.

Ex post *choices and valuations*

Having been presented with the results of both analyses, the group unanimously felt that the decision to fund option (a), that is, to pay an extra £20,106, had been justified. They now felt that the extra benefits of option (a) were worth at least £30,000. This was largely due to the poor quality of the routine data, including the absence of any health outcome data that could replace EuroQol, the inability to obtain data on wound healing and the fact that different routine sources were providing conflicting data. In particular, technical problems had prevented access to theatre data. Individually, all chose option (a), with median WTP of £30,000 (range £10,000–£75,000).

Discussion

It is evident from these results that routine data studies offer considerable scope to reduce the costs of RCTs. The costs associated with all aspects of data handling for the routine data studies were between 8.8% (study C) and 29.4% (study A) of the corresponding cost for designed data.

Differences in the costs of the two approaches should, however, be interpreted with a degree of caution. The designed data studies had to adopt a top-down approach, which is likely to have exaggerated costs relative to the routine data studies, which used a bottom-up approach. These studies recorded every hour actually worked. Thus, while a day may have been devoted solely to a single activity, a late arrival, extended lunch or early leaving would have recorded fewer than 7.5 hours, whereas a top-down approach would have recorded the whole working day. Evidence of this is seen in the fact that study A (IBD), which was costed 'bottom-up', had the lowest cost of any of the designed data studies.

It would also be misleading to subtract the data-handling costs of the designed data studies from the total funding secured for each project, and to deduce that the difference represents the 'other' costs of conducting the trial (including patient recruitment and general project management). In many cases the reported data costs include considerable input by others not funded on the projects. In the case of the study on OSA, for example, a junior doctor who was working towards a degree in this clinical area made a major contribution (638 hours = £10,208) despite not having been mentioned (or costed) in the study proposal. Moreover, all costs reported here are in constant (i.e. current) prices rather than those pertaining at the time of the studies.

Nevertheless the routine data studies were considerably less costly. However, the results of Chapters 3–6 demonstrate that the quality of the output of all these studies was inferior to their designed data counterparts. So, from a cost-effectiveness perspective, no method dominated. Whether or not routine data studies are an efficient way of conducting RCTs thus depends on the value attached to the better research output.

In the mock grants committee exercises, the *ex ante* exercises represent choices between proposals and valuations of anticipated outputs. In three of four cases, the choice was made to fund option (a), that is, the implied value of anticipated extra benefits was at least equal to the differences in costs. For these three proposals, the group WTP was roughly double the implied values. For the single decision not to fund option (a), the group WTP was only one-twentieth of the implied value. While implied values can only represent either minima (opting for (a) implies that the value of extra benefit is **at least** equal to the cost of producing them) or maxima (opting for (b) implies that the value of extra benefit is **less than** the cost of producing them), their results suggest that implied values are poor proxies for actual values.

While there are no rules regarding how grants committees reach decisions, the one-member, one-vote system used here reflects the experiences of the three members of the research team who have sat on such bodies. The collective valuations can vary and, in some cases, do vary considerably from the median values reported by the individual members before discussion.

The *ex post* exercises represent choices between known outcomes with reasons why these results emerged, for example, technical problems with routine data systems. In all cases the extra cost of funding the better designed data studies was judged to have been worthwhile. Differences between *ex ante* and *ex post* WTP, however, were evenly split, with two *ex post* valuations being higher than their original *ex ante* values and two being lower. The reasons for this varied according to the circumstances of the study.

These exercises were designed as preliminary and exploratory investigations into ways of valuing the intangible benefits associated with 'better' research outputs. Accordingly, they had several limitations:

- *'Mock' grants committee.* All members of the mock grants committee were senior representatives of their disciplines, several had experience of sitting on grants committees and the chairman had experience of chairing a grants committee. Nevertheless, these exercises were inevitably artificial.
- *Knowledge of opportunity cost.* A real grants committee will have before it all proposals being considered in the funding round. The opportunity cost of funding more costly studies will thus be apparent. While members of the mock committee were told that the number of proposals worthy of funding far exceeded available funds (to emphasise that there would be opportunity costs to their decisions), it would have been impractical to provide details of other studies.

- *The learning curve.* It was inevitable that as the exercises progressed members would become increasingly familiar with what routine data could and could not provide. In the case of OSA, for example (the second exercise), the committee found that their high expectations of what data routine sources could provide was not met. It is possible, therefore, that this influenced their behaviour in subsequent rounds. However, the option of scheduling the four *ex ante* choices between proposals before the four *ex post* choices between results was rejected on the grounds that it would be less confusing to deal with each study in its entirety before moving on to the next.
- *Availability of information.* For practical reasons the amount of information given to members on methods and results was limited. A real grants committee would inevitably have more information, including comments from referees.
- *Prejudices.* There is considerable concern among clinicians regarding the validity of routine data, and the discussion reflected this. The views expressed might have been different if there were greater trust in the data.

Chapter 9

Health technology assessment through routine data

Since the early 1990s health technology assessment has grown in importance, rigour and cost. This is especially true in England and Wales, where the HTA programme began in 1993. From then the annual budget has almost tripled and the number of completed assessments has grown exponentially. Nevertheless, the flow of new health technologies continues to increase in breadth and depth. So there is clearly an argument for introducing technology assessment methods that might reduce the costs of some assessments, if that were possible without reducing their validity, both internal and external.

This project was designed to assess the scope for a new model for health technology assessment by RCTs. Although this model follows traditional design principles and recruitment methods, it adopts a radical approach to data collection. Rather than designing and collecting data for the sole purpose of health technology assessment, it makes prospective use of existing electronic databases. While this novel approach has the potential to reduce the costs of health technology assessment, it is highly dependent on the scope, accessibility and quality of existing electronic data. Hence, a prospective feasibility study of such an assessment would be risky.

Instead, a retrospective form of feasibility study was undertaken. Four exemplar RCTs were identified, all externally funded and in progress within the School of Postgraduate Studies in Medical and Health Care, University of Wales Swansea. Taking the original trials as designed, and the trial population as randomised, a search was made for existing electronic data systems capable of substituting for the four designed datasets. Primary responsibility for two of the original trials lay with one statistician, while that for the other two trials lay with another statistician.

The four exemplars chosen represent four different health technologies and the results have shown how routine data fared in the context of RCTs, which involved a total of five hospitals in south Wales. However, without further work to assess the data required to support a broader spectrum of assessments, it is difficult to be sure how generalisable these results are.

The retrospective feasibility study was conducted as a cross-over design. One statistician undertook the primary analysis of two trials; and the secondary analysis of the other two, using only the newly identified and assimilated routine electronic data. The other statistician undertook the secondary analysis of the first two of these trials, using only routine electronic data, and the primary analysis of the other two. By comparing the four pairs of analyses, a preliminary assessment could be made of the potential value of this new model for health technology assessment. The advantage of this retrospective study is that we were able to identify many of the strengths and weaknesses of the new model at minimal cost.

However, a major disadvantage is that the routine data that were identified and analysed were less rigorous and less comprehensive than they would have been if all four trials had started life as routine health technology assessments. Thus, the following assessment of the new model of health technology assessment represents the worst case. Therefore, there is a need to identify opportunities to improve electronic systems that would make the new model a real candidate for future assessments.

Each of the four exemplar RCTs has yielded conclusions about the feasibility, utility and resource implications of doing health technology assessment through routine data. Some of the research questions posed by health technology assessment can indeed be answered using routinely collected data within RCTs (*Table 34*).

With the exception of the autologous blood study, the two approaches yielded the same conclusion for most research questions. It was concluded that it is usually possible to answer questions about NHS resource use. Provided clinical symptoms and signs are collected in sufficient detail, one may also be able to assess clinical effectiveness using proxy measures for patient outcomes. There is a need for research to assess the validity and reliability of such proxy outcomes.[52] Although

TABLE 34 *Comparison of conclusions from the two analyses for each exemplar study*

Study A: IBD		Study B: OSA		Study C: Sling		Study D: Autologous blood	
Research questions	**Conclusion**	**Research question**	**Conclusion**	**Research question**	**Conclusion**	**Research question**	**Conclusion**
Specific measure of HRQoL: UKIBDQ (Table 5a)	Same	Diagnostic scores (Table 10a)	Same	Specific measure of HRQoL: IIQ and UDI (Table 14c)	Same for UDI; no surrogate for IIQ	Clinical outcomes (Table 19a)	Different conclusions for infections and circulatory disorder[b]
Generic measure of HRQoL: SF-36 (Table 5a)	Same		Same	Clinical effectiveness (Table 14b)	Same	Generic measure of HRQoL: EuroQol (Table 19b)	Same
NHS resource use and patient-borne costs (Table 5b, c)	Same	NHS resource use and patient-borne costs (Table 10b)	Same	Morbidity scores (Table 10a)	Different conclusion at 3 months.[a] Same conclusion for the remaining follow-up period	NHS resource use (Table 19c)	Same
Patients' and GPs' preferences	Not available from routine data			Patient satisfaction	Not available from routine data		

[a] The designed data analysis showed significantly less pain at the lateral angles of sling dissection and readmission at 3 months in the sling-on-a-string group.
[b] The designed data analysis showed a significantly fewer infections in the autologous group. The routine data analysis showed significantly more circulatory disorders in the homologous group.

patient and professional preferences rarely appear in routine data, they could be collected routinely by adapting existing instruments.

Availability of routine data

The potential for routine data to be used for health technology assessment is summarised in *Table 35*. This potential will be fully realised if the information strategies for the NHS in England[53,54] and Wales[55,56] deliver the infrastructure, culture change, systems and standards that will facilitate the capture of detailed clinical data in structured form in the course of day-to-day clinical practice. The generic clinical information system used to record data on patients recruited into two of the studies (GeneCIS,[21] used in exemplars A and C) can and does capture clinical data in such depth. It also has the facility for the local set-up of structured questionnaires, which facilitate the capture of multi-item scores (HRQoL, patient preferences, disease severity, disability, activities of daily living, etc.). Such systems are not presently widely available, but are embraced in the strategies and described in a specification of requirements published by the Academy of Medical Royal Colleges.[57]

There are thus many potential strengths in routine data. There are also, however, major weaknesses.

Many of the data categories listed in *Table 35* are available at present, and some have been put to good use for prospective health technology assessment in the past. The ISIS study used death as an absolute outcome measure,[58] and the outcome of the management of severe sepsis using a new drug in intensive care patients has been assessed using a high-quality intensive care clinical database, a situation where the results from a conventional RCT may not be generalisable.[59] Further work is required to classify research data items and map these to possible routine sources.

Confidentiality of routine data

During the lifetime of this study there were major changes in the requirements for the

TABLE 35 *Potential of routine data for health technology assessment*

Data category	Source	Potential use for HTA	Current availability	Example
Demography	PAS	Baseline	Wide	Exemplars A–D
Administrative (contacts, dates)	PAS	Resource use	Wide	Exemplars A–D
Clinical Problems Symptoms Signs	Clinical systems	Outcome	Limited	Exemplar A
Clinical Multi-item scores	Clinical systems	Outcome	Very limited	Exemplar A
Clinical Diagnosis Procedures Investigations	HES, CIS HES, CIS, theatre Pathology, X-ray	Outcome and resource use	Wide	Exemplars A–D
Therapy	CIS	Outcome	Limited	Recombinant human activated protein C[59]
Patient preference	CIS	Preferences	Not available	–
Socio-economic	No source	Patient-borne costs	Not available	–
Death	HES	Outcome	Wide	ISIS[58]

confidentiality and security of patient-identifiable data.[60,61] Although ethical approval was obtained at the beginning of the study, the rules changed. At one stage, difficulty was encountered in accessing clinical databases. In particular, routinely collected data could not be obtained from the majority of general practice systems in Exemplar B (OSA), because patient consent for this purpose had not been sought at recruitment. In future, the consent obtained from patients when they enter healthcare should include the possibility of their data being used for research. Confidentiality is unlikely to be an issue for prospective randomised trials, providing consent is obtained at recruitment.

Validity of routine data

Table 36 summarises the percentage of patients for whom data were available within the different systems in each study.

In Wales, HES are held in the PEDW. PEDW included between 29 and 78% of the patients participating in the four studies. The lowest percentage, for the IBD study, was largely due to most patients being seen only as outpatients. However, the remaining three percentages give greater cause for concern, as all of the participants became inpatients or day patients.

These findings are in accord with those of Lewsey and co-workers,[62] who used a different approach to assess the potential of routine data to complement and enhance the results of RCTs. The present authors agree with their conclusions about the validity of centrally held routine data and ways to improve this. The quality of data regularly collected for central use from the NHS has long been in question.[63–65] This study has also shown that central returns may contain diagnoses that have been listed only as queries in clinical notes, there is underuse of symptom codes when the diagnosis is uncertain, some episodes are missing entirely, and there are major differences in data identified by different routine sources.

In contrast, data were more available on local Trust systems. The exceptions were two patient administration systems from which great difficulty in obtaining data was experienced, and one theatre system, which was inaccessible for the duration of the study. This experience suggests that data routinely collected on departmental systems under clinical ownership are more reliable than those collected through processes under managerial control. This evidence is consistent with the literature.[66–69] However, even clinical systems may inappropriately record as definitive, clinical facts that are only suspected. Hence, there is a clear need for standards for the recording of uncertainty.

TABLE 36 *Percentage of study patients identified*

Routine data source	Study A: IBD (two centres)	Study B: OSA (single centre)	Study C: Sling (three centres)	Study D: Autologous blood (single centre)
All Wales: PEDW	29	83	47	71
Trust: Patient administration systems	0; 100	88	85; 33; 100	96
Pathology	100; 100	NA	100; 100; 100	NA
Theatre Man	100; system down	NA	No system; 97; system down	System down
GeneCIS	100; no system	NA	No system; no system; 95	NA
Radiology	100; no system	NA	91; 96; no system	No system
Hospital ECG data	NA	100	NA	NA
Blood bank	NA	NA	NA	84
Primary care: GP data	86	33	NA	NA

NA: not applicable.

The evidence suggests that the lack of clinical ownership of the process of collecting data for central returns primarily is responsible for poor validity. Validation of these data at source would improve their quality. A major driver towards improved data quality is the likelihood that HES data will underpin consultant appraisal and perhaps the new consultant contract.[70] The vision embodied in the information strategies for the NHS in England[53,54] and Wales[55,56] describes electronic records to support the primary purpose of healthcare delivery. If the data are accurate enough to underpin individual care, the accuracy for secondary purposes will improve. Even if there is an improvement, however, the use of routine data for health technology assessment will still require mechanisms to assess the validity of data and permit the identification of missing items. Previous experience of using routine data for health technology assessment needed considerable effort in seeking and then editing errors, inconsistencies and omissions.[71]

Identification of relevant systems

Difficulties were encountered in this study through unfamiliarity with data systems within the organisations with which we were dealing. They lacked expertise in the workings of the systems and relevance of the data they contained. This made it difficult to decide which systems to include. The difficulties were aggravated by changes in databases and the process of data collection. It is imperative that changes to technical specifications and upgrades are logged for every system in use within an organisation.

This study did not use systems outside the health service, such as the Driver and Vehicle Licensing Agency and the Office of National Statistics. However, it is likely that these databases contain information of relevance. For instance, patients with OSA will be more prone to somnolence during the day and this may be manifested in an increased number of vehicle accidents. There is a need for explicit procedures for accessing non-NHS systems.

Access to identified systems

Difficulties were encountered in accessing some systems because staff were not available, trained or cooperative. These are major impediments to making best use of data held in a routine system. The development of local information laboratories that bring together technical, managerial and clinical expertise would be a major step towards

maximising the availability of information from a wide variety of systems in response to properly defined questions.

The concept of Public Health Observatories, which are to be established under the NHS Plan,[72] does not embrace the local access to information systems in primary, community and secondary care that will be needed if clinicians, managers and researchers are to make full use of the data that are stored. This will not only provide a facility to support research, audit, monitoring and planning, but also have educational value, both in promoting understanding of data processes and availability, and in raising awareness of the importance of data quality. The valuation exercise highlighted the mistrust of academics and clinicians in routine data. It is hoped that this will improve as validity improves. Access to data will require formal proposals, which should be scrutinised by the appropriate Caldicott Guardian and Ethics Committee.

Case identification and matching

There was sometimes difficulty in matching patients in different systems through a lack of unique identifiers. The widespread use of the NHS number would ameliorate this problem. In the meantime, loss of data increases with the number of identifying matches required. Thus, if date of birth and postcode are both required to match patients, many will be lost when fields are missing.

Some systems have limited ability to accept electronic queries and extraction was laborious. There was often a need to match case by case. There was also difficulty in ascertaining when data existed on given systems.

Utility of routine data

Most of the systems encountered used different data structures and coding systems. This aggravates the difficulty of extracting and analysing data, particularly when the coding system is local. Data definitions also differed across systems. Standards to ensure the uniformity and validity of data collected in different systems are essential. The need for common, unambiguous definitions was highlighted by McKee and colleagues.[73] No attempt was made to use any data held in free-text format. Nevertheless, this feature provides scope for qualitative data and merits exploration.

Potential of routine data

The strengths of routine data would be enhanced if the weaknesses were overcome. Since the data have already been compiled and stored in electronic form, proponents of the use of routine data believe that studies can be completed more quickly and more cheaply.[74] This may apply especially to health technology assessments that require a large sample size, such as those that aim to detect significant but rare events, small differences in effectiveness or differences that take a long time to manifest themselves fully.[19]

The effectiveness of health technology assessment depends on the size of the effect of the technology being assessed.[75] Thus, assessments intended to identify small to moderate differences between technologies (one of the supposed advantages of health technology assessment by routine data) are at risk from small biases,[71] especially recording bias. Because this risk is greater for retrospective studies,[76] this cannot be advocated as a method of health technology assessment.

Because routine data are not intended for health technology assessment, there is greater danger of multiple analyses of multiple end-points. A simulated clinical trial[77] illustrated the dangers of this approach by allocating 1000 patient records into two 'treatment' groups at random. By subdividing patients in different ways, the authors were able to create significant differences between the two identical hypothetical treatments. This illustrates why the proposal for a new model for health technology assessment requires traditional design principles, especially a rigorous protocol and an explicit analysis plan that prevents multiple comparisons. In other words, health technology assessments using routine data must still fulfil the highest standards for randomised trials. These standards include the CONSORT guidelines for reporting RCTs.[46] These guidelines will need updating to take account of the new model. For example, the principle of blind assessment will give additional responsibilities to staff who manage electronic data symptoms.

Enhancing routine data for RCTs

If routine data are to be used prospectively for RCTs, all of these issues need attention (*Table 37*).

The findings that accurate patient-focused data can act as a proxy for HRQoL are encouraging. Further research on this issue would be valuable.

TABLE 37 *Improving routine data*

Problem	Proposed solution
Case identification	Universal use of NHS number
Validity of routine data	Define appropriate ownership of and responsibility for data collection, with procedures for validation at source, and continuous feedback
Access to routine data	Improve internal knowledge of data systems Create local information laboratories that bring together local expertise Define procedures for scrutiny and approval of data queries
Scope of routine data	Review central requirements for routine data in the light of National Service Frameworks Eliminate unused data items Develop proxy measures for QoL, including routine capture and/or explore the feasibility and cost of routine capture of HRQoL Explore methods of routine capture of patient preferences

TABLE 38 *Conducting health technology assessment by RCTs: practical steps designed or routine data*

Designed data studies	Routine data studies
Conceive study	Same
Set up study team	Similar but include expertise in NHS IT
Define research questions	Same
Generate hypotheses	Same
Develop outcome measures	Consider strengthening routine systems
Estimate sample size	Consider bigger sample to exploit potential and compensate for reduced responsiveness[52]
Define data needs	Similar but outcome measures may change needs
Determine whether data available routinely	Same
Design data where necessary	Identify or develop surrogates
Recruit trial centres	Consider more centres
Gain ethics approval	Same
Recruit patients (informed consent)	Potentially more
Collect and monitor designed data	Not needed
Extract and monitor routine data	More effort needed
Validate data	More effort needed
Analyse data	More scepticism needed
Interpretation	Consider weaknesses and strengths of routine data
Report	Use modified CONSORT guidelines[46]

There is great scope for improving existing systems, especially through a common data structure and coding system. Many unused routine data create 'noise' in systems and these are serious candidates for culling.

Design and conduct of health technology assessment by RCTs through routine data

This experience suggests that the initial steps in the design of RCTs will be similar whether they use designed or routine data. Thereafter, the planning processes will diverge (*Table 38*). Identification of routine data sources and items, possible surrogates, and strengths and weaknesses of these data will require specialist knowledge of information systems and processes. Witting or unwitting subversion of data by healthcare staff who have knowledge that a patient is in a trial is a potential bias in unblinded studies. Routine data identification and capture will need adapted CONSORT guidelines to ensure that data collection and assessment is blind, and analysis and reporting are unbiased.

Conclusions and recommendations for further research

In conclusion, routine data have the potential to support health technology assessment by RCTs. The cost of data collection and analysis is likely to fall, although further work is required to improve

the validity of routine data, particularly in central returns. Better knowledge of the capability of local systems, and access to the data held on them is also essential. Routinely captured clinical data have real potential to measure patient outcomes, if the data were collected in detail and with precision.

There is a need for further research to:

- test prospectively the feasibility of health technology assessment by RCTs through routine data
- classify the research data needed for health technology assessment, and to map these data to potential routine sources
- assess the feasibility, cost and effects of greater clinical ownership and responsibility for HES
- explore the feasibility and cost of local information laboratories aimed at maximising access to, and the utility of, routine data
- understand and change clinicians' and researchers' attitudes to routine data, particularly as validity and availability improve
- define standards to ensure the uniformity and validity of data collected by different local and national systems
- explore the use of surrogate clinical data for measuring patient-focused outcomes
- explore the feasibility and cost of routine completion of HRQoL questionnaires in clinical practice
- explore the feasibility and cost of routine capture of patient preference data.

Acknowledgements

In addition to the authors many others have contributed to the work of this study. We are very grateful to the following: Mrs D Coleman, Project Support Assistant, School of Postgraduate Studies, Swansea; Dr M Cotter, Consultant in Public Health Medicine, Health Solutions Wales, Cardiff; Mrs G Demery, Personal Assistant, School of Postgraduate Studies, Swansea; Miss JC Demery, Research Support Assistant, School of Postgraduate Studies, Swansea; Ms H Dickinson, GeneCIS Technical Manager, School of Postgraduate Studies, Swansea; Dr PD Duane, Consultant Gastroenterologist, Morriston Hospital, Swansea; Dr P Ebden, Consultant Respiratory Physician, Prince Philip Hospital, Llanelli; Mr SJ Emery, Consultant Obstetrician and Gynaecologist, Singleton Hospital, Swansea; Mr DV Ford, Business Manager, School of Postgraduate Studies, Swansea; Mrs T Hughes, Clinical Information Projects Assistant, School of Postgraduate Studies, Swansea; Dr JGC Kingham, Consultant Gastroenterologist, Singleton Hospital, Swansea; Mr MG Lucas, Consultant Urologist, Morriston Hospital, Swansea; Mrs S Lucas, GeneCIS Implementation Officer, School of Postgraduate Studies, Swansea; Ms JM Morgan, Clinical Information Scientist, School of Postgraduate Studies, Swansea; Mr T Stephenson, Consultant Urologist, University Hospital of Wales, Cardiff; Mrs T Stoneman, GeneCIS Information Manager, Neath General Hospital; Dr DW Thomas, Consultant Anaesthetist, Morriston Hospital, Swansea; Ms K Wareham, Clinical Research Scientist, School of Postgraduate Studies, Swansea. Staff of the Clinical Research Unit, School of Postgraduate Studies, Swansea. The staff in many departments of participating hospitals.

References

1. McKee M. Routine data: a resource for clinical audit? *Quality in Health Care* 1993;**2**:104–11.
2. NHS. Central data collections from the NHS. Health Service Circular 1998/054. Leeds: NHS Executive, 1998.
3. Yudkin PL, Redman CWG. Obstetric audit using routinely collected computerised data. *BMJ* 1990;**301**:1371–3.
4. Williams JG. The use of clinical information to help develop new services in a district general hospital. *Int J Med Inf* 1999;**56**:151–9.
5. Carstairs V. Multiple deprivation and health state. *Community Medicine* 1981;**3**:4–13.
6. Carstairs V, Morris R. Deprivation and health in Scotland. Aberdeen: Aberdeen University Press; 1991.
7. Townsend P, Phillimore P, Beattie A. Health and deprivation: inequality and the north. London: Croom Helm, 1988.
8. Reading R, Jarvis S, Openshaw S. Measurement of social inequalities in health and use of health services among children in Northumberland. *Arch Dis Child* 1993;**68**:626–31.
9. Walsh SS, Jarvis S. Measuring the frequency of 'severe' accidental injury in childhood. *J Epidemiol Community Health* 1992;**46**:26–32.
10. Morgan M, Chinn W. ACORN group, social class and health. *J Epidemiol Community Health* 1983;**37**:196–203.
11. Elliot P, Westlake AJ, Hills M, Kleinschmidt I, Rodrigues L, McGale P, *et al.* The Small Area Health Statistics Unit: a national facility for investigating health around point sources of environmental pollution in the United Kingdom. *J Epidemiol Community Health* 1992;**46**:345–9.
12. Majeed FA, Chaturvedi N, Reading R, Ben-Shlomo Y. Monitoring and promoting equity in primary and secondary care. *BMJ* 1994;**308**:1426–9.
13. Colver AF. Health surveillance of pre-school children: four years experience. *BMJ* 1990; **300**:1246–8.
14. Chambers M, Clarke A. Measuring readmission rates. *BMJ* 1990;**301**:1134–6.
15. Roos NP, Wennberg JE, Malenka DJ, Fisher ES, McPherson K, Andersen TF, *et al.* Mortality and reoperation after open and transurethral hyperplasia. *N Engl J Med* 1989;**320**:1120–4.
16. Bernstein SJ, Kosecoff J, Gray D, Hampton JR, Brook RH. The appropriateness of the use of cardiovascular procedures. British versus US perspectives. *Int J Technol Assess Health Care* 1993;**9**:3–10.
17. Scott EA, Black N. Appropriateness of cholecystectomy: the public and the private sectors compared. *Ann R Coll Surg Engl* 1992;**74**:97–101.
18. Flood AB. Peaks and pits of using large databases to measure quality of care. *Int J Technol Assess Health Care* 1990;**6**:253–62.
19. Lange LL, Jacox A. Using large databases in nursing and health policy research. *J Prof Nurs* 1993;**9**:204–11.
20. Sheldon T. Please bypass the PORT. *BMJ* 1994;**309**:142–3.
21. Williams JG, Morgan JM, Howlett PJ, Severs MP. Let there be light. *Br J Healthcare Computing* 1993;**10**:30–2.
22. Williams JG, Morgan JM, Greenway SC, Dickinson H, Cheung WY. Clinical information: the key to improving clinical services. In Richards B, editor. Current perspectives in health care computing. London: BJHC; 1995. pp. 524–8.
23. McCormick A, Fleming D, Charlton J. Morbidity statistics from general practice. Fourth National Study 1991–92. London: HMSO; 1995.
24. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *BMJ* 1991;**302**:766–8.
25. Williams JG, Cheung WY, Russell IT, Cohen DR, Longo M, Lervy B. Open access follow up for inflammatory bowel disease: pragmatic randomised trial and cost effectiveness study. *BMJ* 2000; **320**:544–8.
26. Williams J, Cheung WY, Russell I, Cohen D, Lervy B. Towards appropriate outpatient follow-up of patients with chronic disease. Report to NHS Executive National R&D Programme on the Primary–Secondary Care Interface. Swansea: SPSMHC; 1999.
27. Hutchings H, Evans E, Ebden P. Assessing health gains of community diagnosis of obstructive sleep apnoea. Report to Wales Office of Research and Development for Health and Social Care. Swansea: SPSMHC; 1997.

28. Lucas M, Emery S, Stephenson T, Wareham K, Cheung I, Russell I, Williams J. A randomised study to assess and compare the clinical effectiveness of two surgical techniques for the treatment of stress incontinence in women. Report to Wales Office of Research and Development for Health and Social Care. Swansea: SPSMHC; 2000.

29. Thomas D, Wareham K, Cohen D, Hutchings H. Autologous blood transfusion in total knee replacement surgery. *Br J Anaesth* 2001;**86**:669–73.

30. Thomas D, Newington D, Shenolikar A, Wareham K, Jones L. A study to evaluate the potential benefits and viability of autologous blood transfusion following total knee replacement compared with standard homologous blood transfusion. Report to Wales Office of Research and Development for Health and Social Care. Swansea: SPSMHC; 1998.

31. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman and Hall; 1993.

32. WHO. International statistical classification of diseases and related health problems. 10th rev. Geneva: World Health Organization; 1992.

33. Office of Population Censuses and Surveys. Tabular list of the classification of surgical operations and procedures. 4th revision. London: HMSO; 1990.

34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.

35. Kendall MG, Stuart A. The advanced theory of statistics. Vol. I: Distribution theory. 3rd ed. London: Charles Griffin; 1969.

36. Lawrenson R, Williams T, Farmer R. Clinical information for research: the use of general practice databases. *J Public Health Med* 1999;**21**:299–304.

37. Jarvis GJ. Stress incontinence. In Mundy AR, Stephenson TP, Wein AJ, editors. Urodynamics – principles, practice and application. Edinburgh: Churchill Livingstone, 1994. pp. 299–326.

38. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford: Blackwell; 1994.

39. Brazier JE, Harper R, Jones NMB, O'Cathain A, Thomas KJ, Usherwood T, *et al.* Validating the SF-36 health survey questionnaire: a new outcome measure for primary care. *BMJ* 1992:**305**:160–4.

40. The EuroQol Group. EuroQol – a new facility for measurement of health related quality of life. *Health Policy* 1990;**16**:199–208.

41. Cheung W-Y, Garratt AM, Russell IT, Williams JG. The UK IBDQ – a British version of the inflammatory bowel disease questionnaire: development and validation. *J Clin Epidemiol* 2000; **53**:297–306.

42. Shumaker SA, Wyman JF, Uebersax JS, McClish D, Fanti JA. Health-related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. *Qual Life Res* 1994;**3**:291–306.

43. Garratt AM, Ruta DA, Abdalla MI, Buckingham KJ, Russell IT. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 1993;**306**:1440–4.

44. SF-36 Health Survey Scoring Manual for English Language Adaptations: Australia/New Zealand, Canada, United Kingdom. Boston, MA: Medical Outcome Trust; 1994.

45. NHS Information Authority. Clinical terminology and classification services. URL: http://www.coding.nhsia.nhs.uk

46. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, *et al.* Improving the quality of reporting of randomised controlled trials: the CONSORT statement. *JAMA* 1996;**276**:637–9.

47. Netten A, Dennet J, Knight J. Unit costs of health and social care. Canterbury: Personal Social Service Unit, University of Kent; 1999.

48. Diener A, O'Brien B, Gafni A. Health care contingent valuation studies: a review and classification of the literature. *Health Econ* 1998; **7**:313–26.

49. Mooney G. Economics, medicine and health care. 2nd ed. London: Harvester; 1992.

50. Shackley P, Donaldson C. Willingness to pay for publicly financed health care: how should we use the numbers? *Appl Econ* 2000;**32**:2015–21.

51. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. Oxford: Oxford University Press; 1996.

52. Streiner GL, Norman RD. Health measurement scales: a practical guide to their development and use. Oxford: Oxford University Press; 1995.

53. NHS. Information for health. An information strategy for the modern NHS 1998–2005. A national strategy for local implementation. NHS Executive; 1998.

54. Department of Health. Building the information core: implementing the NHS plan. London: Department of Health; 2001.

55. Welsh Office. Better information better health. Information management and technology for health care and health improvement in Wales. A strategic framework 1998–2005. Cardiff: Welsh Office; 1998.

56. Informing healthcare – towards a knowledge-based NHS. Cardiff: Welsh Assembly Government; 2002.

57. Academy of Colleges' Information Group. Specification of core requirements for clinical information systems in support of secondary care. URL: http://www.aomrc.org.uk. 2001.

58. ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. ISIS-3: a randomised trial of streptokinase vs tissue plasminogen activator vs anistreplase and of aspirin plus heparin vs aspirin alone among 41,299 cases of suspected acute myocardial infarction. *Lancet* 1992; **339**:753–70.

59. Padkin A, Rowan K, Black N. Using high quality clinical databases to complement the results of randomised controlled trials: the case of recombinant human activated protein C. *BMJ* 2001;**323**:923–6.

60. France E. The impact of the law on patient confidentiality: likely consequences. In British Computer Society Health Informatics Specialist Group, editors. Current perspectives in healthcare computing, Part II. London: BJHC; 1997. pp. 21–8.

61. General Medical Council. Confidentiality: protecting and providing information. London: General Medical Council; 2000.

62. Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess* 2000;**4**(22).

63. Körner E. First Report to the Secretary of State of the Steering Group on Health Services Information. London: HMSO; 1987.

64. Williams JG. Information from practice: current position. In Pusey C, editor. Horizons in medicine. Vol. 11. London: Royal College of Physicians; 1999. pp. 221–8.

65. Williams JG, Mann RY. Hospital episode statistics: time for clinicians to get involved? *Clin Med* 2002; **2**:34–7.

66. Kenney N, Macfarlane A. Identifying problems with data collection at a local level: survey of NHS maternity units in England. *BMJ* 1999;**319**:619–22.

67. Pollock AM, Vickers N. Reducing DCO registration through electronic matching of cancer registry data and routine hospital data. *Br J Cancer* 2000; **82**:712–17.

68. Galland RB, Magee TR, Berridge DC, Hopkinson GB, Lewis MH, Shiralkar S, *et al.* Accuracy of centrally recorded OPCS codes for vascular surgery in the United Kingdom. *Eur J Vasc Endovasc Surg* 1998;**16**:415–18.

69. Galland RB, Whatling PJ, Crook TJ, Magee TR. Regional variation in varicose vein operations in England 1989–1996. *Ann R Coll Surg Engl* 2000; **82**:275–9.

70. Maynard A, Bloor K. Reforming the contract of UK consultants. *BMJ* 2001;**322**:541–3.

71. Moses LE. Framework for considering the role of databases in technology assessment. *Int J Technol Assess Health Care* 1990;**6**:183–93.

72. Department of Health. The NHS Plan. A plan for investment. A plan for reform. London: Stationery Office; July 2000.

73. McKee M, Dixon J, Chenet L. Making routine data adequate to support clinical audit: unambiguous definitions are needed. *BMJ* 1994;**309**:1246–7.

74. Sisk JE. Introduction to measuring health care effectiveness. *Int J Technol Assess Health Care* 1990; **6**:181–2.

75. Russell IT. The evaluation of computerised tomography: a review of research methods. In Culyer AJ, Horisberger B, editors. The economic and medical evaluation of health care technologies. Berlin: Springer; 1983.

76. Temple R. Problems in the use of large datasets to assess effectiveness. *Int J Technol Assess Health Care* 1990;**6**:211–19.

77. Lee KL, McNeer JF, Starner FC, Harris PJ, Rosai RA. Clinical judgement and statistics: lessons from a simulated randomised trial in coronary artery disease. *Circulation* 1980;**61**: 508–15.

Appendix 1

Shared care of inflammatory bowel disease

From: Williams JG, Cheung WY, Russell IT, Cohen DR, Longo M, Lervy B. Open access follow-up for inflammatory bowel disease: pragmatic randomised trial and cost-effectiveness study. *BMJ* 2000;**320**:544–8.

Reproduced from the *BMJ* by permission of the BMJ Publishing Group.

Abstract

Objective: To evaluate whether follow-up of patients with IBD is better through open access than by routine booked appointments.

Design: Pragmatic RCT.

Setting: Two district general hospitals in Swansea and Neath, Wales, UK.

Participants: 180 adults (78 with Crohn's disease, 77 ulcerative or indeterminate colitis, 25 ulcerative or idiopathic proctitis) recruited from outpatient clinics during October 1995 to November 1996.

Intervention: Open access follow-up according to patient need.

Main outcome measures: Generic (SF-36) and disease-specific (UKIBDQ) QoL, number of primary and secondary care contacts, total resource use, and views of patients and GPs.

Results: There were no differences in generic or disease-specific quality of life. Open-access patients had fewer day visits (0.21 versus 0.42, $p < 0.05$) and fewer outpatient visits (4.12 versus 4.64, $p < 0.01$), but some patients had difficulty obtaining an urgent appointment. There were no significant differences in specific investigations undertaken, inpatient days, GP surgery or home visits, drugs prescribed or total patient-borne costs. Mean total cost in secondary care was lower for open access patients ($p < 0.05$), but when primary care and patient-borne costs were added there were no significant differences in total costs to the NHS or to society. GPs and patients preferred open access.

Conclusions: Open-access follow-up delivers the same quality of care as routine outpatient care and is preferred by patients and GPs. It uses fewer resources in secondary care but total resource use is similar. Better methods of ensuring urgent access to outpatient clinics are needed.

Introduction

Gastroenterology is a busy medical speciality with a large and expanding outpatient workload.[1] Many patients with gastrointestinal disorders have chronic relapsing disease and some, particularly those with IBD, are traditionally kept under continuing follow-up. This reflects the wishes of GPs[2] as well as specialists, who feel that the unpredictable course, complications and treatment of IBD merit specialist care.[3] However, this traditional approach puts increasing pressure on outpatient clinics.

The aim of the study was to evaluate open access rather than routine booked appointments as a means of following up patients with IBD. The null hypothesis was that outpatient follow-up of patients with IBD through open access is no worse than by routine booked appointments, as judged by HRQoL, total resource use, and patient and GP preference.

Participants and methods

The study was undertaken at two neighbouring hospitals that differ in organisation and management. Morriston is a large district general hospital, which provides most regional specialties. Neath is a smaller hospital with a busy medical intake but no acute surgical services. The hospitals are 14.5 km apart and between them serve a local population of about 250,000 in a predominantly urban area. Gastroenterology clinics at Neath are dedicated to the speciality, whereas at Morriston the clinics also cover general medicine. Neath has a comprehensive clinical information system supporting clinical and service management, which facilitates monitoring and review of patient progress.[4,5] This was not available at Morriston. The study was approved by the West Glamorgan Local Research Ethics Committee, and all patients gave written consent after an oral and written explanation.

Protocol

Comprehensive guidelines for the shared management of IBD were distributed to all local GPs before the study started. These covered diagnosis, medical treatment of mild to severe

disease, laboratory monitoring, the place of surgery, stoma care, follow-up and surveillance, communication, documentation and audit. For patients due for follow-up by open access responsibility for care was transferred back to the GP and routine appointments at outpatient clinics were stopped. In return, the patient was guaranteed rapid access to specialist care when necessary. The normal recall system continued for patients needing regular surveillance by colonoscopy because of the increased risk of colorectal cancer.

Patients were recruited by three consultant gastroenterologists, one staff doctor, two senior registrars and four registrars from outpatient clinics at the two hospitals during October 1995 to November 1996. Patients aged over 18 years, with inactive or mildly active but stable IBD, were invited to take part. Those with active disease requiring treatment, a stoma or other disease that required regular follow-up, or who were thought unable to comply with data collection were excluded. Three GPs declined to collaborate, and their patients were also excluded.

Patients randomised to routine follow-up made their next appointment at the end of each hospital visit as usual. The practice in both hospitals has been to see patients at short notice between appointments if requested by the patient or GP. It was made clear to patients in the routine arm that this policy still applied. Those randomised to follow-up through open access were asked to contact their GP about problems or to contact the hospital directly if they were unable or unwilling to see the GP first. Appointments were made by telephoning outpatient clerks or gastroenterology secretaries, who were made aware of the need to offer an early appointment.

Patients were reviewed in the outpatient clinic in the normal way. A relapse did not require withdrawal, but the patient was seen and treated as appropriate and remained in the same study group. All patients were called for review 24 months after entry into the study.

Primary outcome was measured by the generic and disease-specific QoL questionnaires SF-36[6,7] and UKIBDQ.[8] The questionnaires were completed in clinic at recruitment and at the end of the study and by post at 6-monthly intervals between. Two reminders were sent to non-respondents. Those who failed to attend the end of study appointment after two reminders were sent the final questionnaire by post.

Resource use was estimated from patient questionnaires and case notes. Medical staff abstracted data from hospital notes, and practice staff abstracted them from GP records. Questions on patient-borne costs within the QoL questionnaires covered travel, parking, time off work for the patient and any accompanying person, and other costs such as babysitting. Patient and GP satisfaction, preferences and views were assessed by postal questionnaires at the end of the study, supplemented by semistructured interviews with a sample of GPs. To minimise bias these were undertaken by GPs during audit visits.

When the trial was designed no disease-specific HRQoL scale had been validated for use in the UK. As no previous study of IBD that used SF-36 as an outcome was found, sample size was estimated using SF-36 scores from patients with suspected peptic ulcer,[7] which showed a standard deviation of 20 for most subscales. A difference of 10 points was considered to be clinically important, equivalent to a standardised difference of 0.5. Hence, 170 patients (85 per group) would yield 90% power to detect a significant difference with a significance level of 0.05.[9] A target of 180 was set to allow for loss to follow-up. Limited information on the distribution of key resource variables meant that the sample size could not reflect likely differences in costs.[10]

Analysis was by intention to follow-up. To counteract the effect of possible differences in baseline HRQoL scores, changes in individual scores from baseline were analysed using *t* tests. Preference data were analysed by chi-squared tests.

Because data on use of resources tend to be highly skewed, routine parametric statistics are not appropriate. Therefore significance was assessed by the Mann–Whitney *U* test. As economic analysis is mainly concerned with a comparison of means, however, means and standard deviation are reported for each variable.

Valuation of hospital resources was based on estimates provided by the trusts. Costs of outpatient and GP home visits were derived from Netten and co-workers,[11] drug costs from the British National Formulary,[12] and costs of GP surgery visits from Graham and McGregor.[13] Patients' lost work time was valued by using average wages[14] and their motoring costs were estimated from Automobile Association figures.[15] Total costs to society were derived by summing primary care, secondary care and patient-borne costs.

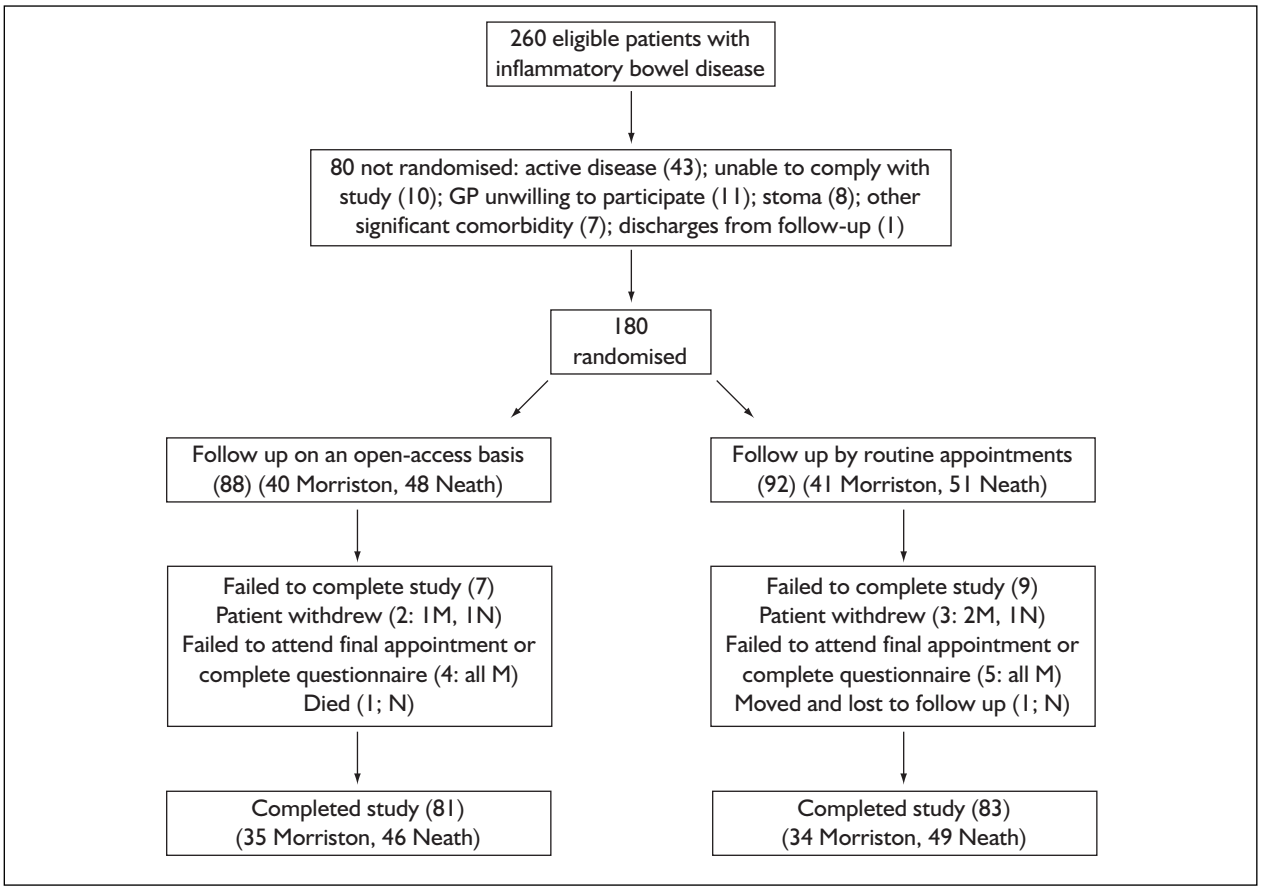

FIGURE 1 *Trial profile*

Random allocation

To ensure balance in type of follow-up, patients were first stratified by centre and between four diagnostic groups: ulcerative or idiopathic proctitis; ulcerative or indeterminate colitis affecting more than the rectum; Crohn's disease of the small or large bowel; and Crohn's disease of the small and large bowel. The computer-generated allocation lists were securely held by one independent researcher in each centre. When the clinician had established the eligibility of the patient and received informed consent, the local researcher was contacted for the random assignment and the patient immediately informed of the follow-up arrangements.

Results

Participant flow and follow-up

Figure 1 shows the progress of the trial. No patients refused to participate, although five subsequently withdrew. QoL questionnaires were completed by 170 patients at 6 months (94%), 160 at 12 months (89%), 159 at 18 months (88%) and 164 at 2 years (91%). The number of patients who failed to complete the study differed significantly between the two hospitals (12 in Morriston versus four in Neath; $p < 0.05$). There was no significant difference between groups at baseline in age, gender, diagnostic group or QoL.

Quality of life

There was no significant change in mean HRQoL scores in either group over the two years of the study, although there was some deterioration in both groups in most subscales. There were no significant differences between groups in changes in HRQoL scores at 6, 12, 18 or 24 months compared with baseline (*Table 1*).

Patients' preferences

Patients had a strong preference for open-access follow-up (103/164, $p < 0.01$); 69/81 (85%) in the open-access follow-up group preferred open-access follow-up, and 34/83 (41%) in the routine group would have preferred open-access follow-up. The main reason given for this preference was the appropriateness of attending only when ill. The reason most commonly given for keeping routine appointments was for reassurance. Some patients had difficulty arranging open-access appointments, and a few would probably have

TABLE 1 *Differences between patients allocated to open access and routine follow-up in changes in QoL from baseline*

	Mean difference (95% CI)[a]			
	At 6 months	**At 12 months**	**At 18 months**	**At 24 months**
UKIBDQ				
Bowel movements and use of facilities	–0.3	4.5	–0.5	0.3 (7.7 to –7.1)
General bowel symptoms	0.2	–1.3	–1.6	–3.5 (4.0 to –10.9)
Emotional function	–1.1	–0.0	–1.6	–1.3 (3.4 to –5.9)
Social function	–1.3	0.5	–2.4	0.4 (6.5 to –5.7)
Systemic function	–2.8	2.6	–0.3	2.2 (9.9 to –5.6)
SF-36				
Vitality	0.1	–2.5	–6.1	–3.7 (3.3 to –10.7)
General health perception	–0.0	–4.6	–2.2	–3.5 (2.0 to –8.9)
Mental health	–2.2	–0.4	–4.1	–3.7 (2.4 to –9.9)
Bodily pain	0.8	0.8	–0.5	–2.5 (5.0 to –10.0)
Physical functioning	–0.5	2.0	–6.0	–3.7 (3.2 to –10.5)
Role limitations due to physical problems	–8.3	–2.7	–2.6	–2.7 (11.4 to –16.8)
Social functioning	–2.9	–2.1	–4.2	–0.4 (8.1 to –8.9)
Role limitations due to physical problems	–13.3	–1.2	–3.4	–5.3 (11.4 to –22.0)

[a] Positive differences denote a better change in open-access patients than in routine outpatients; negative differences denote a better change in routine outpatients than in open-access patients.

been lost to follow-up if they had not been called for the end-of-study visit.

GPs' preferences

Study patients were registered with 53 practices. Forty practices returned postal questionnaires relating to 155 patients, including 12 patients who did not complete the final patient questionnaire (86% response rate). Sixty-nine GPs indicated their preferred method of follow-up for 143 patients (including eight who did not complete the final patient questionnaire). The GPs preferred open-access follow-up for 108 patients (55 in open-access follow-up, 53 routine) and routine follow-up for 35 patients (15 open access, 20 routine). This difference was highly significant even after potential correlation between multiple responses from individual GPs was allowed for ($p < 0.001$). Preference for open-access follow-up was associated with sensible patients, stable disease and the effective booking of urgent review. Forty-four GPs (64%) favoured a gastrointestinal nurse practitioner as point of contact; ten were opposed to this, eight wanted further discussion of the role and seven did not express a view.

Use of resources

Comprehensive data on resource use in both primary and secondary care were available for 155 patients. *Table 2* shows use of hospital facilities. Open-access patients had fewer day visits ($p = 0.019$) and fewer outpatient visits ($p = 0.002$), and cost less in total investigations ($p = 0.032$). There were no significant differences in numbers of inpatient days or specific investigations. Patient-borne costs were lower for open-access patients ($p = 0.002$). Mean total cost for hospital care was significantly lower for open-access patients than routine outpatients (£582 versus £611, $p = 0.012$).

Analysis of resource use in primary care showed no significant differences in GP visits, patient-borne costs or drugs prescribed (*Table 3*). Although more maintenance drugs (5-amino-salicylates) were prescribed in the open-access follow-up group, this did not reach significance. Primary care costs were higher for open-access patients, but not significantly so. When primary and secondary care costs were considered together there was no difference in total costs to society between study groups (*Table 4*).

Discussion

For patients with quiescent or mild, stable IBD, open-access follow-up is preferred by patients and GPs and allows less resource-intensive follow-up in outpatient clinics without deterioration in QoL. QoL questionnaires were completed in the clinic at the beginning and end of the study, and by post at 6-monthly intervals in between. Completion in the clinic tends to underestimate the effect of disease on QoL,[16] but this would not affect the present comparisons between groups.

TABLE 2 *Mean (SD) resources used per patient in hospitals over 24 months*

Resource variable	Open access (*n* = 77)	Routine visit (*n* = 78)	*p* Value for difference[a]
No. of outpatient visits	4.12 (3.41)	4.64 (2.38)	0.002
No. of day cases	0.21 (0.47)	0.42 (0.66)	0.019
No. of inpatient days	0.83 (3.53)	0.41 (1.74)	0.71
No. of investigations			
Full blood count	2.44 (2.43)	2.79 (2.25)	0.09
Erythrocyte sedimentation rate and C-reactive protein	2.27 (2.29)	2.47 (2.23)	0.48
Biochemical profile	2.12 (3.21)	2.46 (2.41)	0.10
Colonoscopy	0.17 (0.38)	0.31 (0.54)	0.09
Rigid sigmoidoscopy	0.08 (0.39)	0.12 (0.46)	0.53
Flexible sigmoidoscopy	0.31 (0.57)	0.32 (0.55)	0.81
Biopsy	0.22 (0.62)	0.24 (0.54)	0.45
Vitamin B_{12}, folate and ferritin	0.49 (0.91)	0.57 (0.99)	0.58
Total cost of investigations (£)	198 (278.99)	257 (276.10)	0.032
Total cost of secondary care (£)	582 (807.94)	611 (475.47)	0.012
Patient-borne cost (£)	74 (61.72)	87 (47.67)	0.002
Total cost to society (£)	656 (859.74)	699 (516.17)	0.011

[a] Mann–Whitney *U* test.

TABLE 3 *Mean (SD) resources used per patient in primary care over 24 months*

Resource variable	Open access (*n* = 77)	Routine visit (*n* = 78)	*p* Value for difference[a]
No. of surgery visits	9.23 (7.76)	7.73 (5.77)	0.47
No. of home visits	0.36 (1.15)	0.41 (1.23)	0.69
No. of tablets or enemas:			
Immunosuppressive drugs[b]	248 (521)	167 (633)	0.70
Maintenance drugs[c]	1287 (1475)	855 (1234)	0.14
Antidiarrhoeal drugs[d]	298 (838)	283 (1138)	0.95
Steroid enemas[e]	9 (31.6)	7 (33.2)	0.42
Total cost of drugs (£)	376 (464)	263 (404)	0.17
Total primary care costs (£)	464 (467)	340 (431)	0.07
Patient-borne cost (£)	40 (33)	35 (27)	0.53
Total costs to society (£)	504 (472)	375 (438)	0.06

[a] Mann–Whitney *U* test.
[b] Prednisolone, azathioprine.
[c] Mesalazine, olsalazine, balsalazide.
[d] Diphenoxylate, codeine phosphate, loperamide.
[e] Hydrocortisone, prednisolone.

TABLE 4 *Summary of mean (SD) costs for open access and routine follow up (£)*

Resource variable	Open access (*n* = 77)	Routine visit (*n* = 78)	*p* Value for difference[a]
No. of surgery visits	9.23 (7.76)	7.73 (5.77)	0.47
Secondary care	582 (808)	611 (475)	0.01
Primary care	464 (467)	340 (431)	0.07
Total NHS cost	1046 (948)	951 (680)	0.89
Patient-borne cost	115 (82)	122 (64)	0.07
Cost to society	1160 (1007)	1074 (724)	0.78

[a] Mann–Whitney *U* test.

Because resource use was skewed in both groups, a larger sample size would have been needed to detect all true differences in costs. However, trends were found towards lower secondary care costs and higher primary care costs for open-access patients. In secondary care these differences were significant, even though the cost of open-access follow-up showed greater variability.

Problems with open access

Despite the strong preference for open-access follow-up, some patients experienced difficulties in making urgent appointments, largely because of pressure on clinics and the inexperience of clerical staff in managing open-access follow-up. Letters from GPs were effective but took time. The best way to overcome this problem may be to have a single telephone point of contact for patients, which is staffed by a specialist gastrointestinal nurse practitioner. GPs were generally supportive of this proposal, and the authors intend to introduce and evaluate this approach. Chronic inflammatory disease is a well-recognised risk factor for the development of gastrointestinal malignancy.[17,18] As well as managing open access, a nurse practitioner could ensure that patients are called back at appropriate intervals for assessment and colonoscopy if necessary.

There were no significant differences in patient characteristics between the two hospital sites. However, significantly more patients at Neath completed the study. The dedicated clinics and computerised clinical information system at Neath may have contributed to this.

To the authors' knowledge, this is the first randomised trial comparing open-access and routine follow-up for patients with IBD, although Probert and colleagues recommended such reorganisation in 1993.[3] Their survey reported that most gastroenterologists in Britain cared for at least 100 patients with IBD, and nearly a quarter of them for 200 or more. Thus the reduction in outpatient attendances that is documented here could save each consultant 25–50 visits a year.

Wider applicability

A study of shared care of patients with moderately severe asthma also found that it was equally effective as hospital care and produced cost savings in secondary care without a significant increase in primary care workload.[19] Similarly, a randomised trial of patients with breast cancer showed that follow-up of patients in remission by GPs was not associated with increased time to diagnosis of relapse, increased anxiety or deterioration in HRQoL. However, resource use and preferences were not evaluated.[20]

Although much has been written about shared care,[19–24] further studies are needed to evaluate whether the present findings can be extrapolated to conditions such as arthritis, epilepsy, heart failure and multiple sclerosis. Much would depend on the ability and willingness of GPs to shoulder the increased responsibility.

What is already known on this topic?

- Routine follow-up of patients with IBD is putting increasing pressure on outpatient clinics.
- Transferring the responsibility for care of patients with asthma saves resources in secondary care without increasing primary care workload or affecting patients' HRQoL.

What this study adds

- Open-access follow-up for patients with IBD does not affect patient care but saves secondary care resources.
- Most patients prefer follow-up through open access.
- GPs think that open-access follow-up is more appropriate for most patients.
- Effective methods are needed for making urgent appointments.

References

1. Sellu D. Have we reached crisis management in outpatient clinics? *BMJ* 1998;**316**:635–6.
2. Moody GA, Mann R, Gay S, Wicks ACB, Mayberry JF. The gastroenterology service: a survey of general practitioners' requirements. *J R Soc Med* 1993;**86**:26–7.
3. Probert CS, Jayanthi V, Mayberry JF. British gastroenterologists' care profile for patients with inflammatory bowel disease: the need for a patients' charter. *J R Soc Med* 1993;**86**:271–2.
4. Williams JG, Morgan JM, Severs MP, Howlett P. Let there be light. *Br J Health Care Computing* 1993; **10**:30–2.
5. Williams JG. The use of clinical information to help develop new services in a district general hospital. *Int J Med Inf* 1999;**56**:151–9.

6. Jenkinson C, Coulter A, Wright L. Short Form 36 (SF36) health survey questionnaire: normative data for adults of working age. *BMJ* l993;**306**:1437–40.
7. Garratt AM, Ruta DA, Abdalla MI, Buckingham KJ, Russell IT. The SF36 health survey questionnaire: an outcome measure suitable for routine use within the NHS? *BMJ* 1993;**306**:1440–4.
8. Cheung WY, Garratt A, Russell IT, Williams JG. The UKIBDQ: a British version of the inflammatory bowel disease questionnaire. *J Clin Epidemiol* 2000;**53**:279–306.
9. Altman DG. Statistics and ethics in medical research III. How large a sample? *BMJ* l979;**282**:1336–8.
10. Drummond MF. Economic analysis alongside controlled trials. London: Department of Health; 1994.
11. Netten A, Dennett J, Knight J. Unit costs of health and social care. Canterbury: Personal Social Services Research Unit, University of Kent; 1998.
12. British Medical Association, Royal Pharmaceutical Society of Great Britain. British national formulary. London: BMA, RPS; 1998 (No. 35).
13. Graham B, McGregor K. What does a GP consultation cost? *Br J Gen Pract* 1997;**47**: 170–2.
14. Office for National Statistics. The new earnings survey. London: Stationery Office; 1998.
15. Automobile Association. Motoring costs. London: AA; 1998.
16. Lyons RA, Wareham K, Lucas M, Price D, Williams JG, Hutchings HA. SF-36 scores vary by method of administration: implications for study design. *J Public Health Med* l999;**21**:41–5.
17. Stewenius J, Adnerhill I, Anderson H, Ekelund GR, Floren CH, Fork FT, *et al.* Incidence of colorectal cancer and all cause mortality in non-selected patients with ulcerative colitis and indeterminate colitis in Malmo, Sweden. *Int J Colorectal Dis* 1995;**10**:117–22.
18. Bernstein D, Rogers A. Malignancy in Crohn's disease. *Am J Gastroenterol* 1996;**91**:434–40.
19. Grampian Asthma Study of Integrated Care (GRASSIC). Integrated care for asthma: a clinical, social, and economic evaluation. *BMJ* 1994;**308**: 559–64.
20. Grunfeld E, Mant D, Yudkin P, Adewuyi-Dalton R, Cole D, Stewart J, *et al.* Routine follow-up of breast cancer in primary care: randomised trial. *BMJ* 1996;**313**:665–9.
21. Diabetes Integrated Care Evaluation Team. Integrated care for diabetes: clinical, psychosocial and economic evaluation. *BMJ* 1994;**308**:1208–12.
22. Petrie JC, Robb OJ, Webster J, Scott AK, Jeffers TA, Park MD. Computer-assisted shared care in hypertension. *BMJ* 1985;**290**:1960–2.
23. Van Damme RAE, Drummond NA, Beattie JAG, Douglas JG. Integrated care for patients with asthma: views of general practitioners. *Br J Gen Pract* 1994;**44**:9–13.
24. Hickman M, Drummond N, Grimshaw J. The operation of shared care for chronic disease. *Health Bull* 1994;**52**:118–26.

Appendix 2

Diagnosis of obstructive sleep apnoea

From: Hutching HA, Evans EN, McKell-Redwood D, Cohen D, Russell IT, Ebden P. Home versus inpatient investigation of obstructive sleep apnoea: diagnostic and cost comparison. Final Report to the Wales Office of Research and Development.

Abstract

Objective: To evaluate the diagnostic validity and costs of home monitoring compared with inpatient investigation of OSA.

Design: Pragmatic, single-blind, crossover, RCT.

Setting: A small district general hospital in Llanelli, Wales, UK.

Participants: 102 patients referred with suspected OSA.

Intervention: Synectics MicroDigitrapper S home sleep system.

Method: The AHI derived from home and hospital diagnosis, and the cost of performing each arm of the study were compared.

Results: The median AHI was similar for inpatient (1.4; range 0–77.0) and home (1.6; range 0–45.4) diagnosis. The inpatient system diagnosed three more cases (25) than the home system (22) and there was an 83% level of agreement between the two systems, with low failure rates. Using the Bland and Altman method, the mean difference was 2.2 (range 25.8 to –21.4) with 95% confidence intervals of 4.8 to –0.4. The mean cost of home monitoring was £89.99 per patient compared with £301.07 for inpatient diagnosis.

Conclusions: Monitoring for OSA at home compared well with inpatient diagnosis. Since home monitoring was considerably cheaper it may provide a useful alternative to inpatient investigation in a large proportion of cases.

Introduction

OSA is caused by upper airway collapse during sleep leading to recurrent arousals, which may be associated with hypoxia and which are terminated with increased respiratory effort. The disorder presents with repetitive apnoeas and hypopnoeas during sleep, loud snoring and pathological daytime sleepiness.

Polysomnography has traditionally been thought of as the 'gold standard' for investigation of sleep-related breathing disorders. The usefulness of recording electrophysiological variables, however, is now debated. The recording of respiratory indices can be calculated with equally sufficient accuracy from the total time in bed as from the total sleep time, and the additional expense of recording sleep quality and quantity therefore appears to be unjustified.[1] Current electroencephalography scoring for arousals is insensitive[2] and also poorly predicts daytime dysfunction compared with the AHI.[3] There is therefore no need to assess the quality and quantity of sleep in order to diagnose OSA.

The number of patients referred with suspected OSA is increasing, resulting in growing demands for diagnostic studies. The majority of centres lack the time, money and experienced staff to perform full polysomnography and, as such, the use of more limited sleep studies is being advocated for the diagnosis of patients.[1,4,5]

Home monitoring has an advantage over limited inpatient sleep studies, with patients able to sleep in their usual environment. There are also likely to be cost advantages of home monitoring with the potential for significantly reduced costs. However, these systems must have proven diagnostic capabilities and be physically robust, otherwise the cost savings could be negated by increased failure rates at home and difficulties in setting up the system by the patients.

Many studies comparing home sleep studies with polysomnography have performed these studies within the laboratory setting where they are partially or fully attended.[6–8] There have been few studies that have compared these unattended devices in the home setting and the authors are aware of only one validation study in the UK.[9] The aims of this prospective cross-over study therefore were to compare the diagnostic efficacy of home versus inpatient diagnosis of OSA.

Methods

Patients

The study population consisted of 102 consecutive patients (88 male) of median (range) age 47 (17–76) years, with a median body mass index of

29.2 (19.5–52.2) kg/m^2, referred to the hospital with suspected OSA. Patients were examined and interviewed (using a clinical symptom questionnaire) and, after full explanation of the nature of the study, were asked for their consent to participate in a study comparing home and inpatient diagnosis of OSA.

Patients were randomised to receive either inpatient or home monitoring first, in a cross-over design. The alternative investigation was performed on the following night. An AHI of greater than five events per hour was considered diagnostic of OSA and patients were treated on the basis of the first study result. This study was approved by the Local Research Ethics Committee.

Inpatient investigation

Patients were monitored overnight with the Compumedics P-Series Sleep System (Compumedics, Windsor, Australia).

The patients were studied over a single night. Abdominal and thoracic movements were detected via two respiratory inductance bands, one placed at the level of the axillae and the other at the widest abdominal girth. Leg electromyograms were recorded using sensors placed on the tibialis anterior muscle of each leg (in order to exclude periodic leg movements). Nasal/oral airflow was detected by a thermistor placed between the nose and the mouth. Pulse rate and arterial oxygen saturation were monitored via an in-built pulse oximeter connected to the patient by a finger probe. Body position was detected by a mercury switch and a light sensor detected background light. An in-built microphone detected any sounds above a preset threshold. In addition an overnight video recording of the patient with a timed overlay (using a monochrome camera sensitive to infra-red) was made. This picture was transmitted to the ward for care of the patient to be maintained, but the investigator was not present during the overnight assessment. The sensors for the system were attached to the patient by the same investigator in the late afternoon.

Each patient was instructed to attach the finger probe and connect him or herself to the patient interface box when retiring to bed. The following morning the information was downloaded to the computer, and the recording was inspected and analysed.

The analysis program automatically identified apnoeas and hypopnoeas, and further manual inspection of the record was made to edit periods of artefact. The Compumedics system defined apnoeas and hypopnoeas by an 80% and 50% reduction in airflow, respectively, within 30 seconds of a 4% dip in SpO_2 (oxygen saturation by pulse oximetry).

Home investigation

Home monitoring was performed with the Synectics MicroDigitrapper S Sleep System (Synectics Medical, Stockholm, Sweden).

Pulse rate and oxygen saturation were recorded via an in-built oximeter, which was attached to the patient by an adhesive finger probe. Snoring sound was recorded using a surface encased skin-throat microphone. Thoracoabdominal movements were detected by piezoelectric sensors incorporated into an effort sensitive detector and abdominal band. Nasal/oral airflow was monitored via a thermistor and body position via a five-way encapsulated mercury switch.

The same sampling rates and constants were used for all patients and were set up to correspond as closely as possible to the settings used on the Compumedics system.

Patients for home monitoring attended the hospital to collect the equipment, and were given a verbal and written explanation of how to use the monitor at home and how to attach the sensors. The monitor was set up to start recording when next switched on by the patient when they were ready to sleep. The equipment was returned on the next working day and the information retrieved by means of a computer interface linked to the Microdigitrapper S software. Analysis of the channels of information was performed automatically. After visual analysis of the automated scores for each channel, editing and further analysis was performed manually. The Synectics system defined apnoeas and hypopnoeas by a 70% and 50% reduction in airflow, respectively, within 30 seconds of a 4% dip in SpO_2.

Economic comparison

The patients were questioned regarding their mode of travel, time and distance travelled to the hospital. Using the Automobile Association motoring costs,[10] the average cost per journey was calculated. Lost productivity, attributed to days lost from work, was calculated using national average earnings.[11] The time required to set up, offload and analyse the data from the home and inpatient systems was recorded. The costs of the equipment used in the study were calculated using the Equivalent Annual Cost Method.[12]

Data analysis

Data were loaded onto an SPSS program (SPSS, Chicago, IL, USA) and statistical analyses were performed using this program. Data were expressed as medians (ranges), apart from the economic data, which were expressed as means (± SD) where appropriate. The strength of agreement between the AHIs derived using the inpatient and home sleep systems was assessed using the Bland and Altman method.[13] The sensitivity, specificity, positive predictive value and negative predictive value of home monitoring, compared with inpatient monitoring were calculated. A paired samples *t* test was used to determine whether there was any significant difference between the AHI obtained during inpatient and home monitoring.

Results

Inpatient and home results

In 18 cases data were missing for either the inpatient (five) or home (13) monitoring system. Comparison of the systems was therefore made on the remaining 84 patients. The median AHI obtained on inpatient investigation was 1.4 (0–77.0), compared with a value of 1.6 (0–45.4) obtained at home. The inpatient system diagnosed 25 cases positive for OSA, with the home system diagnosing 22. There was no significant difference between the AHI values obtained with home and hospital monitoring ($p > 0.05$). The level of agreement between the home monitoring system and the inpatient system using the Bland and Altman method is illustrated in *Figure 1*. The mean bias between the two systems (mean of the differences between inpatient and home monitoring) was 2.2, with limits of agreement (mean bias ± 2 S.D) of –21.4 to 25.8 and 95% confidence intervals of –0.4 to 4.8. Only six values fell outside the limits of agreement and there was less agreement with increasing values of AHI.

Diagnostic value of home monitoring

The sensitivity of home monitoring compared with inpatient monitoring (taken as the gold standard) was moderate (63%), with a corresponding high specificity of 92%. The percentage positive and negative predictive values were 75% and 86%, respectively. There was an 83% level of agreement between inpatient and home monitoring. Inpatient and home results showed no significant difference whether they were performed as the first or second limb of the investigation ($p > 0.05$).

Economic analysis

Table 1 illustrates the costs of inpatient and home diagnosis of OSA. The mean overall cost for overnight inpatient investigation, including costs for repeat monitoring for failed tests, was £301.07. This was more than three times the cost of

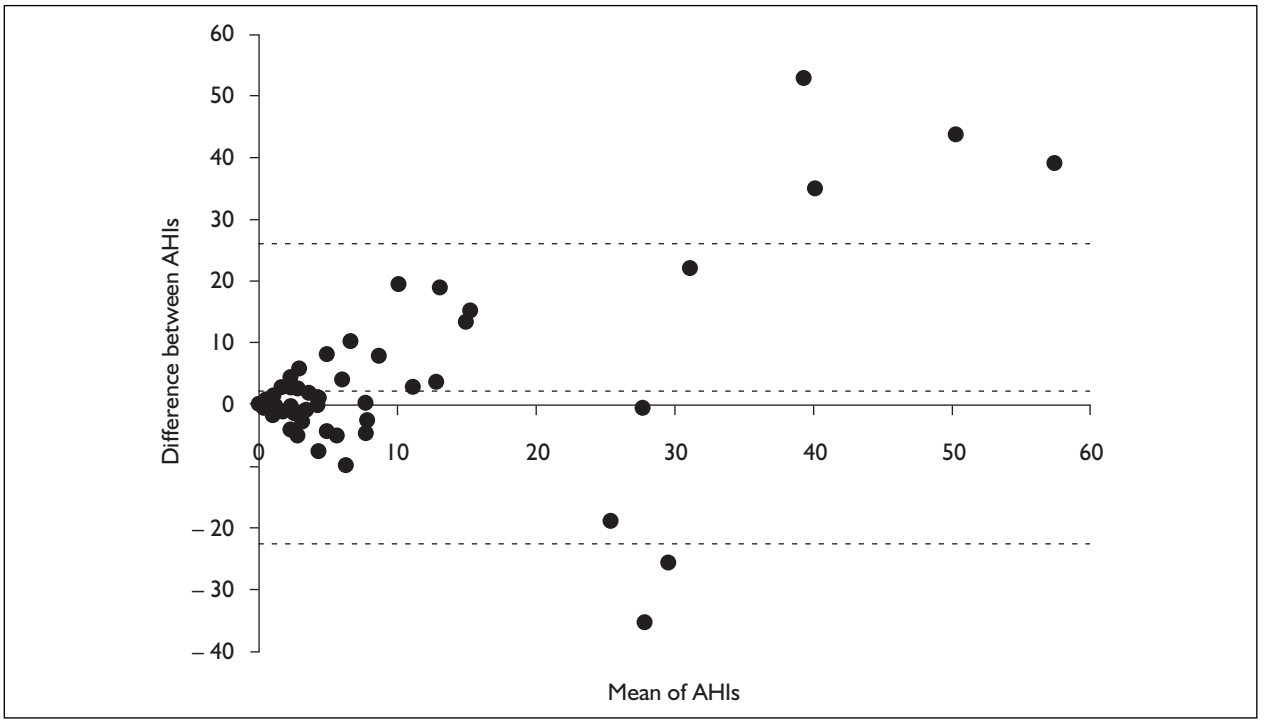

FIGURE 1 *Level of agreement between the home and inpatient AHI calculated using the Bland and Altman method*

TABLE 1 *Home versus hospital diagnosis; cost per patient*

	Cost per patient (£)	
Resource	**Hospital**	**Home**
Patient travel	4.26	8.52
Lost productivity	50.62	28.12
H grade nurse	31.67	16.84
House officer	3.50	0.00
Hotel costs	156.00	0.00
Equipment and servicing	35.74	17.39
Disposables	7.92	13.34
Total costs of monitoring	289.71	84.21
Total costs including repeats	301.07	89.99

monitoring at home (£89.99), which included an increased cost component due to more failed tests. The higher hospital costs comprised hotel costs for an overnight stay, higher patient lost productivity costs due to time off from work, more nurse/technician time to analyse and set up the system, and higher equipment costs.

Discussion

In the home setting, up to 63% of patients can be accurately diagnosed with OSA using an unattended sleep system, with excellent specificity. Eighty-three per cent of patients were given an accurate diagnosis compared with inpatient investigation. The inpatient and home systems diagnosed 25 and 22 positive cases, respectively, at an AHI of 5, and all patients recording an AHI greater than 20 as an inpatient were given the same diagnosis at home.

The median AHI value at home was slightly but not significantly higher than that in hospital. This may occur because patients are able to sleep more comfortably within their usual environment and achieve a greater proportion of rapid eye movement sleep, which is known to aggravate OSA. The range of individual AHI values obtained with the inpatient system was larger than with the home system. The reason for this may be that the total sleep time is less accurate at home and this modifies the AHI. The only indicators of wakefulness were large changes in posture recordings and the presence of artefact. No indication of exact sleep time in terms of lights on/off, etc., were made. The inpatient system recorded lights on/off, which, together with a video recording, allowed a more accurate determination of sleep time and may account for the wider range of AHI values. The increased ranges of values with the inpatient system could, however, point to an increased variability of recordings in hospital with patients unable to sleep as well as in their home environment.

The level of agreement between the inpatient and home systems obtained using the Bland and Altman technique was good, with only six values falling outside the limits of agreement. However, all six patients recorded an AHI of greater than 20 with both systems and were subsequently treated. The values compare well with two recent studies with similar limits of agreement and mean differences obtained.[8,14] It is difficult to compare this study directly with these two studies as they both used attended polysomnography as the inpatient diagnostic system.

With regard to the sensitivity and specificity values obtained, a moderate sensitivity was obtained in this study, which was coupled with a high specificity and good negative and positive predictive values. These calculations assume the inpatient system to be the gold standard, with 100% accuracy in the diagnosis of OSA. The inpatient system would, however, be subject to error and more accurate results may be produced at home because the patients are able to sleep better. The results confirm those of Zucconi and co-workers,[8] who illustrated that the Synectics system was a useful system for diagnosing suspected cases of OSA.

The home system had a higher number of failures than the inpatient system. However, the home system showed good patient compliance, with only two failures of the system due to patient error.

The systems used here performed better than those tested in another study,[14] where the home system showed a bias to lower AHI scores in the home study with higher inpatient AHI values. A number of methodological differences exist between the Edinburgh group and the present study. First, the Edinburgh group performed a validation study with only 20 patients compared with 84 in this own study. Secondly, there was no randomisation between the inpatient and home limbs of the study, with polysomnography always performed first. If this first study was unpleasant it may create apprehension coupled with poor sleep quality with the subsequent underestimation of the AHI in the home study. Lastly, the comparison was made between polysomnography and limited home studies, whereas the present study compared two systems that primarily recorded respiratory variables, which may account for the differing results.

A number of economic findings were highlighted. The mean cost of inpatient monitoring was over £300, with most of this accounted for by hotel and patient loss productivity costs. The mean cost of home monitoring was under £90 per patient, including costs for repeat failed studies. This was more than three times less than inpatient monitoring. This difference is similar to that documented by Parra and co-workers,[14] who performed a similar study in Spain. Less time off was required by the patient for home monitoring, and the time required to analyse and set up the system was much less than for inpatient investigation. Since home monitoring compared well with inpatient monitoring for diagnosis of OSA it is substantially more cost-effective than inpatient monitoring.

In conclusion, this study found that home monitoring compared well with inpatient investigation and, since it represents a substantial reduction in costs per patient, that it could provide a useful alternative to inpatient studies in a large proportion of cases.

References

1. Douglas NJ, Thomas S, Jan MA. Clinical value of polysomnography. *Lancet* 1992;**339**:347–50.
2. Stradling JR, Davies RJO, Pitson DJ. New approaches to monitoring sleep-related breathing disorders. *Sleep* 1996;**19**(9):S77–84.
3. Douglas NJ, Martin SE. Arousals and the sleep apnoea/hypopnoea syndrome. *Sleep* 1996;**19**(10):S196–7.
4. Ferber R, Millman R, Coppola M, Fleetham J, Murray CF, Iber C, *et al.* Portable recording in the assessment of obstructive sleep apnoea. *Sleep* 1994;**17**:378–92.
5. Standards of Practice Committee of the American Sleep Disorders Association. Practice parameters for the use of portable recordings in the assessment of obstructive sleep apnoea. *Sleep* 1994;**17**:372–7.
6. Lloberes P, Montserratt JM, Ascaso A, Parra O, Grandados A, Alonso P, *et al.* Comparison of partially attended night time respiratory recordings with full polysomnography in patients with suspected sleep apnoea/hypopnoea syndrome. *Thorax* 1996;**51**:1043–7.
7. Esnaola S, Dura'n J, Infante-Rivard C, Rubio R, Ferna'ndez A. Diagnostic accuracy of a portable recording device (MESAM IV) in suspected obstructive sleep apnoea. *Eur Respir J* 1996;**9**:2597–605.
8. Zucconi M, Ferini-Strambi L, Castronovo V, Oldani A, Smirne S. An unattended device for sleep-related breathing disorders: validation study in suspected obstructive sleep apnoea syndrome. *Eur Respir J* 1996;**9**:1251–6.
9. Whittle AT, Finch SP, Mortimore IL, MacKay TW, Douglas NJ. Use of home sleep studies for the diagnosis of the sleep apnoea/hypopnoea syndrome. *Thorax* 1997;**52**:1068–73.
10. AA motoring costs. 1996.
11. National Office of Statistics. Labour force survey. June 1996.
12. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 1987.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**i**:307–10.
14. Parra O, García-Esclasans N, Monserratt JM, García Eroles L, Ruíz J, López JA, *et al.* Should patients with sleep apnoea/hypopnoea syndrome be diagnosed and managed on the basis of home sleep studies? *Eur Respir J* 1997;**10**:1720–4.

Appendix 3

Surgery for stress incontinence

From: Lucas MG, Emery SJ, Stephenson TP, Wareham K, Cheung WY. Tension free autologous fascial sling; a randomised controlled trial to compare two surgical techniques for the treatment of stress urinary incontinence in women. Final Report to the Wales Office of Research and Development.

Abstract

Objective: To compare the relative morbidity and efficacy of two techniques using an autologous fascial sling for the treatment of genuine stress incontinence in women.

Design: RCT.

Patients and methods: 165 patients with urodynamically confirmed genuine stress urinary incontinence were randomised to receive tension free autologous fascial sling utilising either a pubovaginal sling (group A) or a short sling-on-a-string procedure (group B). The study was undertaken at three centres in south Wales. Patients were followed up during hospitalisation and at 3, 6 and 12 months following surgery.

Main outcome measures:

- Primary: reduction in leaked urine volume during a 1 hour pad test, changes in HRQoL scores: Urogenital Distress Inventory (UDI) and Incontinence Impact Questionnaire (IIQ), the incidence of symptoms of urinary urgency and voiding difficulty, patient satisfaction.
- Secondary: postoperative pain, operative time, length of stay in hospital, immediate and late complications including readmissions to hospital within 12 months.

Results: At 12 months there was no significant difference between the two groups in reduction of urine leakage by the pad test (mean 41 to 9.5 ml versus 37 to 4.6 ml), in changes in HRQoL scores by UDI or IIQ, stress incontinence (0.16 versus 0.16), urge incontinence (0.46 versus 0.38), voiding difficulty (0.23 versus 0.20) or patient satisfaction (0.78 versus 0.76).

Pain: Operative details were similar for the two groups, although there was a significant time advantage to the sling on a string (group A mean 62.4 minutes, group B mean 54.0 minutes, $p = 0.001$). This difference was consistent when patients who had had concurrent procedures carried out were excluded. There were no significant differences in length of stay or immediate (0.38 versus 0.40) and late (0.30 versus 0.20) complications. Readmissions during the follow-up year totalled 35 for group A and 25 for group B (0.43 versus 0.29).

Conclusion: These results confirm that there is no significant difference between the efficacy and morbidity of the two procedures at 12 months. The short sling-on-a-string technique is quicker to perform.

Introduction

The surgical approach to the control of stress incontinence in women is bedevilled by choice. Over the last 100 years many surgical procedures have been described but despite the improved understanding of the anatomy and the pathophysiology of vesicourethral dysfunction and a wealth of surgical literature, no procedure has gained universal acceptance. A meta-analysis of the literature[1] revealed that of 263 papers reviewed there were only seven randomised controlled trials and only 23.5% of patients had been assessed postoperatively. In another cohort study,[2] although most women appeared to be improved by surgery and were usually declared cured by the surgeon, critical analysis showed that only 28% were completely dry at 1 year. Both authors stressed the need for properly constructed clinical trials to compare the treatments used for stress incontinence.

The treatment of patients who have previously failed with surgery is particularly difficult as the results of repeat surgery are worse, but recent series reporting the use of sling procedures have given 90% success rates even in this group.[3,4] A pilot study conducted by one of the investigators in this project showed a 94% early success rate following correction by pubovaginal sling fashioned from anterior rectus sheath.[5] Theoretical advantages of the sling technique include the fact that it achieves urethral support without deforming the vagina, it works as well for patients with urethral hypermobility as for those with intrinsic sphincter weakness[6] and it can be carried out as successfully in patients requiring repeat surgery as in those undergoing their first procedure.[7] Based on this group's early results and the theoretical advantages of the procedure, the pubovaginal sling became their operation of choice as both a primary and secondary procedure.

The sling procedure is relatively time consuming owing to the time taken to harvest autologous

fascia, and patients often experience pain in the region of the fascial dissection. Therefore an alternative technique was adopted, using the same principle of placement of a fascial sling around the posterior urethra, but using a short piece of fascia mounted on Nylon threads. Theoretically this approach ought to reduce the incidence of rectus sheath pain. Early results from this procedure have reported successful outcomes.[8,9]

A randomised prospective trial was therefore set up to compare the modification of the sling technique ('sling on a string') with a conventional pubovaginal sling to test a null hypothesis that there would be no difference in the relative morbidity and efficacy of the two techniques. The study was planned to be carried out in three hospitals in south Wales.

Patients and methods

Population

The study commenced following review and approval from the two local research ethics committees. Patients who were deemed suitable for the study were invited by their consultant to participate, and were assured that if they declined to enter the study their management would not be affected. Patients were also informed of their right to withdraw from the study at any time. They were assured that their medical notes would be treated in strict confidence and that all information collected on them would be reported using a unique research number, ensuring their confidentiality.

All women who met the entry criteria and gave informed consent were entered into the study. Eligibility for entry into the study required women aged 18 years and older with clinically and urodynamically proven stress urinary incontinence who needed corrective surgery. Exclusion criteria included women who refused or who were unable to give consent, women with evidence of neurological disease and women with urodynamic evidence of detrusor instability or hypocompliance (assessed urodynamically as a pressure rise of more than 20 cmH_2O at a capacity of 500 ml using a filling speed of 50 ml/minute).

Intervention

The investigators at each centre agreed to an operation standard for the two procedures.

Group A (standard pubovaginal sling)

A 20 × 1.5 cm strip of fascia was fashioned from a transverse strip of rectus sheath taken through a transverse suprapubic incision. After opening of the retropubic space through two laterally placed vaginal incisions, the sling was passed through a tunnel at the level of the bladder neck and the ends were passed through the retropubic space and sutured with absolutely no tension to the rectus fascia on each side.

Group B (sling on a string)

The technique for this study was similar to that of group A except that the length of the sling obtained was between 8 and 10 cm, thus requiring a smaller skin incision. This was mounted at each end with a no. 1 Nylon thread and passed through the retropubic space to be tied over the rectus fascia aponeurosis, again with no tension.

Other procedures, such as prolapse repair, perineal repair, abdominoplasty and hysterectomy, were permitted and were fully documented. Antibiotic prophylaxis, analgesia and thromboprophylaxis were standardised within each unit.

Outcome measures

Primary outcomes

The primary outcome measures were assessed at 3, 6 and 12 months; in each case the study was looking for no significant difference in outcome:

- 1 hour pad test (undertaken with a minimum of 200 ml in the bladder assessed by a bladder scan)
- HRQoL scores. the UDI and IIQ[10]
- patient history of urinary urgency or voiding difficulty
- patient satisfaction.

Secondary outcomes

Operative time ('skin to skin') was recorded. Pain scores were measured on a four-point scale at 24 hours, 4 or 5 days and 3 months after the operation. Other end-points included the number and type of immediate complications, the length of postoperative stay in hospital, readmissions related to surgery, the number of days from surgery to successful voiding and details of patients who were taught intermittent self-catheterisation.

Statistical analysis

A sample size of 164 patients was required to show (with 80% power) that there was no difference in efficacy (less than 15% difference assuming that the standard operation would be 95% successful in alleviating stress incontinence). Statistical analysis was performed by intention to operate. To counteract the effect of possible differences in baseline HRQoL scores and volume of leakage,

changes in individual measurements from baseline were compared using *t* tests. Covariance analysis and regression analysis, following the recommendation of Berry and Armitage,[11] were performed to see whether any observed significant change was a statistical artefact. Operative time and length of stay in hospital were also analysed by *t* tests. Postoperative pain scores, incidence of symptoms of urinary urgency, voiding difficulty, patient satisfaction, complications and readmissions were analysed by chi-squared tests.

An independent evaluator was appointed by the funding body to review clinical outcomes at 6 and 12 months into recruitment to ensure that there were no major differences between the two groups.

Assignment

A computer-generated randomisation schedule was used for each centre and each individual. Remote telephone randomisation by the independent research team was undertaken after obtaining written informed consent. The study was not blinded because the type of operation performed was obvious to all medical and nursing personnel involved in assessment.

Results

Recruitment

In all, 168 patients were recruited (41 Morriston, 72 Singleton, 55 Cardiff) over a 2-year period. A further nine patients declined to enter the study. Three of the 168 patients recruited into the study were subsequently found not to have genuine stress incontinence confirmed by video-urodynamics and therefore they were not eligible for the study. Data on the remaining 165 patients are reported.

Demographic details, baseline symptomatology and clinical findings are described in *Table 1*. There was no difference between the groups. Forty-six patients had undergone 49 previous operations for incontinence (*Table 1*) and these were equally distributed between the two groups.

Intervention

Table 2 compares the operative details of time and blood loss for the two groups. Whether or not concurrent procedures were included, the time taken for group B was significantly shorter than that for group A ($p = 0.003$). There was no significant difference in blood loss or the use of vaginal packing.

TABLE 1 *Demography, presenting symptoms, clinical findings and details of previous surgery by group*

Description	Long sling (group A)	Short sling (group B)
Demography		
Mean age (range) (years)	51.64 (31–71)	52.1 (33–73)
Mean height (range) (cm)	160 (146–180)	160.4 (142–177)
Mean weight (range) (kg)	73.1 (45–100)	69.1 (51–98)
Mean symptom years	7.1 (1–30)	7.8 (1–35)
Mean frequency (times between voids) (hours)	1.9	2.1
Urge incontinence (*n*)	65	72
Clinical findings		
Hypermobility of bladder	57 = mild 15 = gross11 = gross	65 = mild
Associated anterior vaginal wall prolapse	38 = grade 1 2 = grade 2/3	35 = grade 1 3 = grade 2/3
Vault prolapse	17 = grade 1 2 = grade 2/3	13 = grade 1 2 = grade 2/3
Rectocele	14 = grade 1 3 = grade 2/3	23 = grade 1
Vaginal squeeze	7 = absent 8 = normal 64 = weak	3 = absent 9 = normal 69 = weak
Vaginal scarring	22 = yes 3 = gross	17 = yes 2 = gross
Urodynamic findings	81 GSI	84 GSI
Previous surgery		
Anterior/vaginal repair/colporhaphy	16	19
Endoscopic needle suspension of bladder neck	2	2
Retropubic bladder neck suspension	6	4
Clam cystoplasty	1	0
Posterior repairs/anorectal	10	8

TABLE 2 *Operative details, relevant early and late adverse events associated with surgery for the two groups*

Description	Long sling (group A)	Short sling (group B)	*p* Value	95% CI of the difference
Operative details				
Mean operative time (skin–skin) (minutes)	62.44 (38–135)	54.01 (25–140)	0.001	(–13.34 to –3.52)
Mean operative time (excludes concurrent procedures)	60.72 (38–94)	51.82 (25–90)	0.001	(–13.32 to –4.48)
Mean blood loss (ml)	273.64 (50–800)	229.81 (50–700)	0.07	(–91.26 to 3.61)
No. of patients transfused	2/81	2/84	1.00	(–5% to 5%)
No. of vaginal packs inserted	53/81	57/84	0.87	(–12% to 17%)
Number and type of concurrent procedures				
Vaginal repair	7	9	0.65	(–7% to 11%)
Adominoplasty	7	5	0.51	(–11% to 5%)
Hysterectomy	3	4	1.00	(–7% to 7%)
Others: minor procedures	4	4	1.00	(–13% to 13%)
Number and type of perioperative complications				
Bladder injury	2	3	1.00	(–4% to 6%)
Bleeding	10	8	0.56	(–12% to 7%)
Urinary tract infection	10	6	0.26	(–14% to 4%)
Suprapubic catheter problems	2	4	0.71	(–3% to 8%)
Respiratory	3	3	1.00	(–6% to 6%)
Cardiac	2	5	0.47	(–3% to 10%)
Wound	7	5	0.72	(–11% to 5%)
Readmissions related to surgery				
Within 3 months	19/79	9/83	0.03	(–25% to –2%)
3–6 months	6/79	10/84	0.36	(–5% to 13%)
>6 and up to 12 months	10/73	6/80	0.21	(–16% to 4%)
From discharge up to 12 months	35/81	25/84	0.07	(–28% to 1%)
Adverse events not requiring readmission				
Within 3 months	33/79	41/83	0.33	(–8% to 23%)
3–6 months	39/79	33/84	0.20	(–25% to 5%)
> 6 and up to 12 months	18/83	27/80	0.22	(–5% to 23%)
Complications related to operation from discharge up to 12 months	24/81	17/84	0.16	(–23% to 4%)
Pertinent voiding difficulty at 12 months	19/81	17/84	0.62	(–16% to 9%)

Primary outcomes

The results of the 3-, 6- and 12-month assessments are summarised in *Table 4*. No significant differences were observed in the number of patients who were satisfied with the outcome of surgery, or in the incidence of urgency and persistent stress incontinence between the groups. Pad test results are summarised in *Table 3* and were significantly improved compared to baseline but, again, there was no difference between the operations.

Patients who underwent sling surgery as a primary procedure were compared with those who had had previous surgery. This revealed a 23% difference in group B between those who had sling on a string as a primary procedure and those who had it done as a secondary or redo procedure (*Table 4*). However, the numbers involved were too small to be able to state whether this operation truly performs worse when used as a secondary procedure.

There were no significant differences in changes in HRQoL scores between the two groups at 3 or 6 months. Group A patients improved significantly more in HRQoL at 12 months by the UDI ($p = 0.04$). However, group A patients had significantly worse HRQoL at recruitment than group B patients ($p = 0.003$) and had more room for improvement. Covariance analysis was performed by regressing patients' UDI scores at 12 months on their scores at recruitment for the two study groups separately. The slopes of the two resultant regression lines were compared by the *t* test. There were no significant differences between the slopes ($p > 0.05$). This suggested that the observed difference in relative improvements in UDI between the two groups could be due to their different scores at recruitment.

Secondary outcomes

There was no significant difference in pain scores, or the site of pain, between the two groups at either 24

TABLE 3 *Comparison of primary outcomes with status at baseline*

	Long sling (group A)				Short sling (group B)			
Description	**Baseline**	**3 months**	**6 months**	**12 months**	**Baseline**	**3 months**	**6 months**	**12 months**
Pad test: mean leakage	39.49	6.92	17.60	7.71	36.97	8.61	2.02	4.61
Mean reduction in leakage (ml)	NA	–37.50	–25.79	–33.37	NA	–31.14	–34.27	–34.91
Patient satisfaction	NA	62/78	56/78	57/73	NA	66/83	63/79	62/82
Stress incontinence	81	12/78	13/78	12/74	84	11/82	10/79	13/80
Urge syndrome	65/81	39/77	36/78	34/74	72/84	37/83	32/80	30/80

Quality-of-life questionnaires

Description	**Group A**	**Group B**	**Two-tail significance level**	**95% CI of the difference**
UDI				
Baseline	181.22	156.98	0.003	(–40.38 to –8.11)
3 month difference from baseline	–126.00	–111.19	0.21	(–8.29 to 37.91)
6 month difference from baseline	–121.29	–111.87	0.44	(–14.59 to 33.44)
12 month difference from baseline	–128.40	–101.34	0.04	(1.12 to 52.99)
IIQ				
Baseline	236.25	226.56	0.51	(–38.75 to 19.36)
3 month difference from baseline	–146.19	–162.02	0.45	(–57.38 to 25.71)
6 month difference from baseline	–149.53	–164.09	0.49	(–56.01 to 26.88)
12 month difference from baseline	–167.90	–127.07	0.08	(–4.32 to 85.96)

NA: not applicable.

TABLE 4 *Comparison of primary surgery and repeat surgery for the two groups*

	Long sling (group A)	**Short sling (group B)**	**Difference (B – A)**	***p* Value (95% CI of the difference)**
Primary surgery				
n	56/81	63/84	6%	0.40 (–8% to 20%)
Improved	44/56	55/63	9%	0.20 (–5% to 22%)
Persistent GSI	7/56	6/63	–3%	0.60 (–14% to 8%)
De novo urge	5/56	2/63	–6%	0.35 (–14% to 3%)
Repeat surgery				
n	25/81	21/84	–6%	0.40 (–20% to 8%)
Improved	18/25	14/21	–5%	0.70 (–32% to 21%)
Persistent GSI	6/25	7/21	9%	0.48 (–17% to 36%)
De novo urge	1/25	0/21	–4%	1.00 (–12% to 4%)

hours or at day 4 or 5 (*Table 5*). In the immediate postoperative period 75 (93%) patients in group A and 72 (86%) patients in group B used a patient-controlled intravenous analgesia technique. In group A 11 (14%) patients also received intramuscular opiates compared with 17 (21%) in group B. Oral analgesia was recorded for 80 patients in group A and 83 patients in group B, and was comparable.

Time to successful void for both groups was a median of 3 days postoperation. Intermittent self-catheterisation was taught to 21 patients in group A and 13 patients in group B, while still inpatients. At the time of discharge 59 group A patients and 69 group B patients were voiding normally. This difference was not statistically significant.

There was no significant difference in mean postoperative length of stay between the two groups (pubovaginal sling 6.48 days versus sling on a string 6.73 days, $p = 0.62$).

Early and late adverse events and reasons for readmission to hospital were evaluated and those considered to be related in any way to the surgery are summarised in *Table 2*, but no difference was apparent between the two operations.

Discussion

Since this study was initiated there has been a great deal more interest in sling procedures, which have now been widely adopted as a primary surgical treatment in the USA,[4] and have gained advocates in the UK as well.[7] There has been much written about the short-term results and the complications of various techniques, but still remarkably little evidence from RCTs informs this debate. A Cochrane Centre review of RCTs involving the use of suburethral slings has identified only five studies which met the criteria for assessment, yet even these were deemed methodologically poor and totalled only 206 women, with 126 who had undergone a sling procedure.[12] The review criticised the inadequate power of studies, failure to describe the method of randomisation and baseline comparisons, and the absence of third party outcome assessments. None of these criticisms applies to this study.

The hypothesis underlying the design of this study developed from the realisation that occasionally when harvesting a long strip of fascia, insufficient material was obtained to achieve continuity of fascia from the rectus sheath down around, below the bladder neck and back up again to the rectus

TABLE 5 *Postoperative pain scores and the site of pain recorded at three assessment points*

	Long sling (group A)	Short sling (group B)	Difference (B – A)	*p* Value (95% CI of the difference)
24 hours after operation				
Patients with any type of pain	73/80	75/79	4%	0.36 (–12% to 4%)
Patients with wound pain	43/80	45/79	3%	0.68 (–12% to 19%)
Patients with angle pain	28/80	29/79	2%	0.82 (–13% to 17%)
Patients with other pain	2/80	1/79	–1%	1.00 (–5% to 3%)
4 or 5 days after operation				
Patients with any type of pain	34/81	48/84	17%	0.051 (0% to 30%)
Patients with wound pain	15/81	21/84	6%	0.31 (–6% to 19%)
Patients with angle pain	18/81	24/84	6%	0.45 (–8% to 18%)
Patients with other pain	0/81	3/84	4%	0.86 (0% to 8%)
3 months after operation				
Patients with any type of pain	52/78	42/82	–16%	0.047 (–30% to 0%)
Patients with wound pain	13/78	18/82	5%	0.52 (–7% to 17%)
Patients with angle pain	35/78	22/82	–19%	0.011 (–34% to –5%)
Patients with other pain	3/78	2/82	–2%	0.68 (–7% to 4%)

sheath. Suspension threads have to be used under these circumstances and there appeared to be no detrimental effect on the surgical outcome.

There was no difference in outcome at 12 months between the two procedures. The sling-on-a-string technique is a quicker procedure and the authors feel justified in adopting this modification as their standard procedure. It is premature to make claims of equal efficacy until longer term follow-up is available, although there is some evidence that if sling techniques are successful within the first postoperative year then the long-term survival of the procedure is good.[13]

It would appear that the mode of action is dependent on fixation of the autologous rectus fascia to the tissues around the bladder neck and the pelvic floor in that region. The fascia also becomes adherent to side-wall structures in the cave of Retzius. Once the tissues are fixed, attachment or continuity with the anterior abdominal wall structures is no longer relevant.

The choice of suture material used to secure the short free sling was a non-absorbable monofilament Nylon thread in this study, but other slowly absorbed materials [vicryl and polydioxanone suture (PDS)] have also been used without any apparent effect on long-term results.

The symptom of urgency was common both pre- and postoperatively in this series. Previous studies have not focused on this particular problem, although it is known that women find that urge incontinence interferes with their quality of life (QoL) more than stress incontinence.[14] With both operations, 53% of patients were cured of urgency. This has been shown to be in part a result of achieving a closed bladder neck at rest.[15] Six patients in group A and two in group B developed *de novo* urge incontinence. This may be a result of a sling becoming too tight due to shrinkage[16] and may resolve after vaginal division of the sling in the midline.

All patients were asked to undergo postoperative urodynamics. However, only 107 patients consented to this repeat investigation, and the study was performed at varying intervals (median 12 months range 4–23 months; 96 had the study performed between 11 and 15 months). The purpose of the study was simply to identify detrusor instability when present and confirm the presence or absence of genuine stress incontinence. Although previous studies[17,18] have utilised additional urodynamic outcome parameters such as functional length and maximum urethral closure pressure, these measures were not used here because they have been shown not to correlate with clinical status.[19,20] Thus, the inclusion of postoperative urodynamic assessment in this study added little to the outcome evaluation.

When planning this study the UDI and IIQ had been promoted as standard QoL tools for this population.[10] This group would now choose to use both condition-specific and generic QoL scores and attempt to shorten the number of questions involved. Several such measures have been proposed and used in other studies,[21–25] but there is still an urgent need for standardisation in this field. All condition-specific scoring systems suffer from being too focused on one area of dysfunction, often ignoring the impact that one organ system can have on another. For instance, surgery for stress incontinence usually impacts on bowel and sexual function as well as urinary function, and objective measures should take these factors into account.

There was a problem with randomisation at entry, in which group A had a worse QoL than group B on recruitment for both the UDI and the IIQ (*Table 1*). The difference was highly significant with the UDI ($p = 0.003$, i.e. a 3 in 1000 chance that this difference is a chance finding). The difference between the two groups with the IIQ was less pronounced and did not reach statistical significance. A detailed investigation of the methodology and randomisation process was carried out by an independent statistician to determine the reasons for the differences between the two groups at recruitment. Covariance analysis and regression analysis suggested that the observed differences in relative improvements in UDI between the two groups at 12 months were due to their different scores at recruitment, hence, a statistical artefact. This may have been a result of some patients performing the assessment after randomisation had taken place, when the patient knew which operation she was going to have. This prior knowledge may have subtly modified their perceptions of QoL. There is no evidence for this contention but nonetheless it is recommended that all baseline QoL assessment take place before randomisation in future studies of this type.

The subjective results for this series are not as good as some previously reported.[3,8,26] Indeed, the results were subjectively worse that those reported previously by two of the authors in which the evaluations were done by the surgeon and not by an independent team of evaluators.[5,9] It is possible that our surgery has been getting worse, but more likely that, as has been shown elsewhere, independent evaluators uncover a higher level of

morbidity and dissatisfaction with outcomes than do the surgeons themselves.[2,27] The inclusion of independent evaluation is a vital component of any RCT in this field.

References

1. Jarvis GJ. Stress incontinence. In Mundy AR, Stephenson TP, Wein AJ, editors. Urodynamics; principles, practice and application. New York: Churchill Livingstone: 1994; pp. 299–326.
2. Black NA, Griffiths JM, Pope C, Bowling A, Abel PD. Impact of surgery for stress incontinence on morbidity – cohort study. *BMJ* 1993;**315**:1493–8.
3. Chaikin DC, Rosenthal J, Blaivas JG. Pubovaginal fascial sling for all types of stress incontinence: long term analysis. *J Urol* 1998;**160**:1312–16.
4. Leach GE, Dmochowski RR, Appell RA, Blaivas JG, Hadley HR, Lubar KM, *et al.* Female Stress Urinary Incontinence Clinical Guidelines panel summary report on surgical management of female stress urinary incontinence. *J Urol* 1997;**158**:875–80.
5. Barrington JW, Fulford SCV, Bales G, Stephenson TP. The modified rectus fascial sling for genuine stress incontinence. *J Obst Gynaecol* 1998;**18**:61–2
6. Blaivas G Jacobs B. Pubovaginal fascial sling for the treatment of complicated stress incontinence. *J Urol* 1991;**145**:1214–18.
7. Bidmead J, Cardozo L. Sling techniques in the treatment of genuine stress incontinence. *Br J Obstet Gynaecol* 2000;**107**:147–56.
8. Mason R, Roach M. Modified pubovaginal sling for the treatment of intrinsic sphincter deficiency. *J Urol* 1996;**156**:1991–4.
9. Lucas MG, Emery SJ, Thomas JA, Holt BA, Wareham K, Roberts C. Sling on a string: for the surgical treatment of stress incontinence [abstract]. *Br J Urol*. 1996;**77**(Suppl 1):3.
10. Shumaker SA, Wyman JF, Uebersax JS, McLish D, Fantl JA, for the Continence Program in Women (CPW) Research Group. Health related quality of life measures for women with urinary incontinence: the Incontinence Impact Questionnaire and the Urogenital Distress Inventory. *Qual Life Res* 1994;**3**:291–306.
11. Armitage G, Berry P. Statistical methods in medical research. Oxford: Blackwells; 1994.
12. Bezerra CA, Bruschini H. Suburethral sling operations for urinary incontinence in women (Cochrane Review). In: The Cochrane Library 2000; Issue 3. Oxford: Update Software.
13. Morgan J, Farrow G, Stewart F. The Marlex sling operation for the treatment of recurrent stress urinary incontinence; a sixteen year review. *Am J Obstet Gynecol* 1985;**151**:224–6.
14. Frazer MI, Haylen BT, Sutherst JR. The severity of urinary incontinence in women. Comparison of subjective and objective tests. *Br J Urol* 1989; **63**:14–15.
15. Fulford SC, Flynn R, Barrington J, Appanna T, Stephenson TP. An assessment of the surgical outcome and urodynamic effects of the pubovaginal sling for stress incontinence and the associated urge syndrome. *J Urol* 1999;**162**:135–7.
16. Mcguire E, Lytton B. Pubovaginal sling procedure for stress incontinence. *J Urol* 1978;**119**:82–4.
17. Hilton P, Stanton SL. Urethral pressure measurement by microtransducer: the results in symptom free women and in those with genuine stress incontinence. *Br J Obstet Gynaecol* 1983; **90**:919–33.
18. Bump RC, Copeland WE, Jr, Hurt WG, Fantl JA. Dynamic urethral pressure/profilometry pressure transmission ratio determinations in stress incontinent and stress continent subjects. *Am J Obstet Gynecol* 1988;**159**:749–55.
19. Meyer S, De Grandi P, Schmidt N, Sanzeni W, Spinosa JP. Urodynamic parameters in patients with slight and severe genuine stress incontinence: is the stress profile useful? *Neurourol Urodyn* 1994;**13**:21–8.
20. Swift SE, Ostergard DR. Evaluation of current urodynamic testing methods in the diagnosis of genuine stress incontinence. *Obstet Gynaecol* 1995; **86**:85–91.
21. The EuroQol Group. EuroQol – a new facility for the measurement of health related quality of life. *Health Policy* 1990;**16**:199–208.
22. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF36 health survey questionnaire. *Qual Life Res* 1993;**2**:169–80.
23. Jenkinson C, Layte R, Jenkinson D, Lawrence K, Petersen S, Paice C, Stradling J. A shorter form health survey: can the SF12 replicate results from the SF36 in longitudinal studies? *J Public Health Med* 1997;**19**:179–86.
24. Jackson S, Donovan J, Brookes S, Eckford S, Swithinbank L, Abrams P. The Bristol Female Lower Urinary Tract Symptoms questionnaire: development and psychometric testing. *Br J Urol* 1996;**77**:805–12.
25. Kelleher CJ, Cardozo LD, Khullar V, Salvatore S. A new questionnaire to assess quality of life in urinary incontinence in women. *Br J Obstet Gynaecol* 1997; **104**:1374–9.
26. Carr L, Walsh P, Abraham V, Webster G. Favourable outcome of pubovaginal slings for geriatric women with stress incontinence. *J Urol* 1997;**157**:125–8.
27. Black NA, Downs SH. The effectiveness of surgery for stress incontinence in women; systematic review. *Br J Urol* 1996;**78**:497–510.

Appendix 4

Autologous blood transfusion

From: Thomas D, Wareham K, Cohen D, Hutchings H. Autologous blood transfusion in total knee replacement surgery. *Br J Anaesth* 2001;**86**:669–73.

Abstract

This study compared allogeneic blood usage for two groups of patients undergoing total knee replacement surgery. Patients were randomised either to receive their postoperative wound drainage as an autotransfusion (n = 115) after processing or to have this wound drainage discarded (n = 116). Allogeneic blood was transfused in patients of either group whose haemoglobin fell below 90 g/l. Only 7% of patients in the autotransfusion group required an allogeneic transfusion compared with 28% in the control group ($p < 0.001$). There was no hospital mortality and only a 3% mortality from all causes at the study completion, which spanned 6 months to 3 years. There was a higher incidence of infection requiring intervention in the allogeneic group ($p < 0.036$). Total patient costs were £113 higher in the autotransfusion group. In conclusion in this type of surgery postoperative cell salvage is a safe and effective method for reducing allogeneic blood use.

Introduction

The identification of transfusion transmitted diseases, such as HIV, hepatitis C and new variant Creutzfeldt–Jakob disease (nvCJD), has led to an increasing number of tests that need to be performed before allogeneic blood transfusion. The current allogeneic blood supply is probably the safest ever produced. However, this statement does not take account of the evolution and subsequent identification of other illnesses that may be transmitted by blood transfusion. The most recent problem relates to the transmission via blood transfusion of the prion thought to cause nvCJD. This has led to the leucodepletion of all blood products within the UK. The increased cost of blood products has added considerable enthusiasm in medical circles to decrease their use, and subsequent patient exposure to allogeneic blood.

There is some evidence,[1] although inconclusive, suggesting that patients may suffer fewer perioperative infections if they avoid allogeneic blood transfusion at the time of surgery. Using autotransfusion may be one way of maintaining perioperative haemoglobin levels and reducing the need for allogeneic blood transfusion.[2–5] If perioperative transfusion practices were aimed at a minimum haemoglobin level, rather than trying to maintain a preoperative value, this would further decrease perioperative transfusions.

Owing to the paucity of evidence in the literature from randomised controlled studies, on RCT was undertaken to confirm the findings from a small pilot study,[6] which suggested that a marked reduction in allogeneic blood transfusion could be safely achieved using postoperative red cell salvage (PRCS) and a haemoglobin transfusion trigger.

Patients and methods

This was a single-centre RCT of patients undergoing TKR surgery. Following review and approval by the local research ethics committee, informed consent was obtained from 231 patients. Patients were randomly allocated to one of two treatment groups; one group received allogeneic blood (if their haemoglobin fell below a preset transfusion trigger of 90 g/l) and the other group received autotransfusion of wound drainage if the volume was above 125 ml postoperatively. The collected blood was washed and resuspended in saline before reinfusion, using a centrifugal cell washing machine (Cell Saver 5 Haemonetics). The patients in the cell salvage group were also transfused if their haemoglobin fell below the preset trigger after autotransfusion. A transfusion trigger was chosen to standardise the transfusion incidence in both groups. Although the American Society of Anaesthesiology recommended a trigger of 7 g/l it was felt this was perhaps too aggressive and it would be difficult to apply. Many anaesthetists would be reluctant to withhold blood at this level of anaemia knowing the correlation with an optimum oxygen delivery and haemoglobin of 10 g/l. Haemoglobin levels were estimated on days 1, 2, 3, 4 and 7 for all patients. The TKR was conducted as routine. Data were collected by research nurses for postoperative length of stay, perioperative and post-hospital discharge infection rates, adverse

TABLE 1 *Demography of patients undergoing a routine TKR by randomised group*

	Autologous (cell salvage)		Allogeneic (homologous)		*p* Value
Description	**Females (*n* = 71)**	**Males (*n* = 44)**	**Females (*n* = 61)**	**Males (*n* = 55)**	**0.18[a]**
Mean age (range) (years)	70.5 (32–95)	67.4 (38–85)	70.2 (40–87)	69.7 (48–88)	0.59[b]
Mean ASA grade (range)	2.1 (1–3)	2.0 (1–3)	2.1 (1–3)	1.9 (1–3)	0.56[b]
Smokers	9	10	6	9	0.35[a]
Previous transfusion	15	9	17	12	0.09[a]
Aspirin/non-steroidal anti-inflammatory use	25	19	24	31	0.16[a]

[a] Fisher's exact test.
[b] Independent samples *t* test.

TABLE 2 *Methods of anaesthesia*

Description	Group 1	Group 2
General alone	67	67
General + femoral block	24	23
General + spinal block	6	9
General + epidural	7	7
Spinal alone	7	6
Epidural alone	0	1
Femoral alone	3	3
Spinal + epidural	1	0

events, wound healing rates and quality of life (EuroQol EQ-5D).[7]

One of the investigators scrutinised all adverse events in a blinded fashion (details of group were withheld) to determine which were possibly related to transfusion effects (e.g. wound infection, embolic events, myocardial ischaemic events and cardiopulmonary complications).

Data were loaded onto an SPSS version 7.5 computer program (SPSS, Chicago, IL, USA) and all statistical analysis was performed using this program. The level of statistical significance for all tests was set up at a *p* value of less than 0.05. For bivariate analysis a two-tailed test of significance was used. Baseline demographics of the two groups were examined using Fisher's exact test (categorical data) or the independent sample *t* test (numerical data). In respect of patients with adverse events the comparison was examined using the chi-squared test.

Results

Of the 231 patients 98 were males and 133 were females, 115 were randomised into the cell salvage group, who received their own blood after processing (collected blood was washed and resuspended in saline before reinfusion) and 116 controls. Characteristics of patients in the two groups were comparable (*Table 1*). Although the method of anaesthesia was not standardised, *Table 2* depicts the type of anaesthesia used and the similarity between the groups. The majority of patients had a Johnson and Johnson prosthesis (75 in the autologous arm and 77 in the allogeneic arm), with the remaining patients in both groups having a De Puy prosthesis. All knee replacements were performed under tourniquet, with the pressure set according to systolic blood pressure. Patients in both groups were transfused with allogeneic blood if their haemoglobin fell below the preset trigger of 90 g/l.

The study was analysed on an intention-to-treat basis. Of the 115 patients randomised into the autologous arm of the study, 85% received an autologous transfusion. The remaining 18 patients were not transfused owing to technical problems and a lack of technical staff to operate the cell salvage equipment (13 patients), insufficient blood collection (four patients) and tourniquet failure (one patient).

Twelve patients in the autologous arm of the study received an allogeneic transfusion. Two were inappropriate, as both patients had a haemoglobin level above 90 g/l and were asymptomatic, and could be classed as a protocol deviation. The remaining ten patients had haemoglobin levels between 76 and 89 g/l. Of the 10 patients whose transfusion was warranted, four of these were from the 18 patients in whom cell salvage failed and a further three patients had only a small amount of blood salvaged (<150 ml).

In the control group 33 patients received allogeneic blood. The majority (76%) received 2 units, 6% received 3 units, 6% received 4 units and 12% of patients had a 1 unit transfusion.

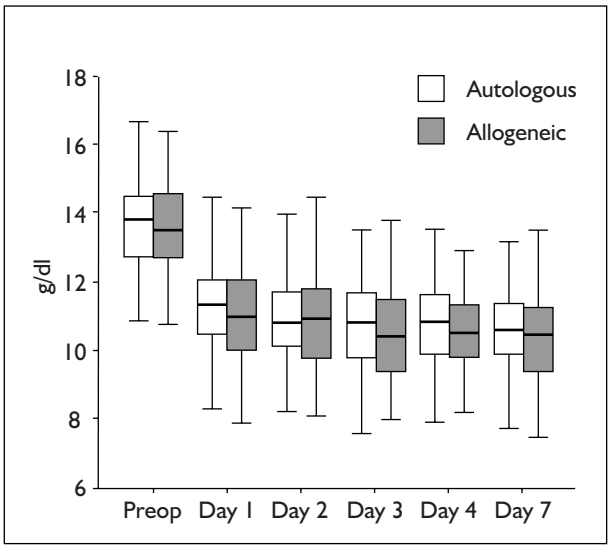

FIGURE 1 *Pre-operation – day 7 post-operation*

There was no significant difference in clinical outcome when analysed for length of stay, wound healing, serious adverse events or mortality, or health-related quality of life (using EuroQol) 6 months following surgery. There was a significantly lower (7%) incidence of allogeneic blood transfusion in the cell salvage group, compared with 28% in the controls ($p < 0.001$). There was no difference in postoperative mean haemoglobin levels between the two groups (*Figure 1*).

In relation to transfusion practice there were significantly fewer readmissions to hospital ($p = 0.008$) and visits to GPs ($p = 0.043$) among patients in the autologous blood transfusion group (*Table 3*). Infective complications were increased in allogeneic recipients ($p = 0.036$), with increasing significance ($p = 0.025$) if all patients receiving allogeneic blood were placed in the allogeneic group.

TABLE 3 *Average per patient costs (£, 1998) of allogeneic and autologous blood transfusion*

Description	Cost of allogeneic transfusion	Cost of autologous transfusion
Allogeneic blood	27.96	12.20
Staff time		49.34
Capital and servicing		24.12
Disposables	00.74	80.12
Total direct cost	**28.70**	**165.78**
Readmission	34.65	11.66
GP Consultation	01.55	00.72
Total indirect costs	**36.20**	**12.38**
Total per patient cost	**64.90**	**178.16**

A comparison of the cost difference between allogeneic transfusion and autologous transfusion was made by one of the authors (DC). A summary of the findings is shown (*Table 4*).

Discussion

This study showed that a decrease in allogeneic blood use could be achieved using PRCS. A previous study[6] showed that the use of PRCS could dramatically reduce patients' exposure to allogeneic blood (from 82 to 18%) without clinical detriment, as all patients left hospital. In that study, as with this one, there was no statistical difference in the discharge haemoglobin levels. The design of the original pilot study did not apply a strict transfusion trigger to both groups. That study was trying to show how the use of PRCS could improve transfusion practice over routine clinical practice. This study, however, assessed the difference in transfusion of allogeneic blood between the two groups while applying a strict transfusion trigger to both randomised groups. There is no doubt that by applying rational transfusion principles, a large decrease in allogeneic transfusion can be achieved without the use of any autologous transfusion methods.

The present study shows that further significant reductions can be achieved by the use of PRCS, decreasing the overall use of allogeneic transfusion to below 7% in the autologous group. Despite the publication of other studies showing a similar trend,[2–6,8] this study is one of the largest RCTs yet performed. Criticism has also been levelled at the lack of outcome measures applied to many studies assessing the practice of PRCS. In this study no patients failed to leave hospital from either group. It was reassuring that the end of study mortality (in some cases indicating a 2-year follow-up and/or a minimum of 6 months post-hospital discharge) was similar in both groups, when deaths from all causes were considered. This mortality rate compared very favourably with a large orthopaedic audit reported from the Mayo Clinic.[9]

The data did not support a difference in immediate postoperative infection or earlier hospital discharge, as had been supported by earlier publications. The well-recognised effect of immunomodulation due to allogeneic blood transfusion was not apparent. This may add weight to the argument that the universal leucodepletion offers only minor benefit. The

TABLE 4 *Readmission by group: conditions considered possibly related to blood products*

Males/females (autologous group)		Males/females (allogeneic group)	
Age/Sex	**Problem**	**Age/Sex**	**Problem**
83 F	Recurrent pulmonary embolism/wound abscess: died	72 M	Cellulitis – ?DVT – doppler
66 M	Wound infection	65 F	DVT
72 F	Wound infection persisted: sinus explored which did not connect to knee replacement	70 M	Infected wound
88 M	Cellulitis	70 F	Wound infection
65 M	Acute anaemia: 4 units blood transfused	78 F	Superficial wound infection: antibiotics commenced
60 M	? PE – heparinised overnight	68 M	Wound infection

length of stay was consistent with data from other Welsh hospitals (Department of Public Health, personal communication). Assessment using the EuroQol Health State score could not show a difference between groups, in contrast to the proposal that patients receiving autologous blood have improved health and well-being compared with those receiving allogeneic blood. It was noted that the EuroQol scores improved in both groups when preoperative and 12 weeks postoperative scores were compared.

The only area where a statistical difference between groups was found was in the post-hospital infective complications, with allogeneic recipients having increased infective problems. This effect became even more significant if those who received rescue transfusion were included in the allogeneic group. These findings are supported by the reduced readmissions and visits to GPs by patients who had been randomised to receive an autologous transfusion.

This study has shown that a reduction in allogeneic transfusion can be achieved safely using a combination of PRCS and limiting the transfusion when the patient has a haemoglobin higher than 9 g/dl. This can be considered by most as a conservative haemoglobin trigger and appropriate for even those patients with significant cardiac disease.[10–15] The patients undergoing this type of joint replacement are more elderly than the general population and thus more likely to suffer from heart disease. The authors believe therefore that such blood conservation techniques are clinically indicated in the light of present evidence. There is a need to seek safe alternatives to allogeneic blood, both to decrease the risk of future unknown blood-borne transmitted disease and to increase the availability of allogeneic blood supplies where there are no available alternatives.

The cost analysis showed that autologous transfusion was overall more expensive, despite having lower readmission and postoperative GP costs. At the time of the study the unit cost of allogeneic blood was £50.83. In addition, staff time of £49.34 was estimated on a cell-salvage operator being present throughout the postoperative collection period. In practice this is not necessary. Processing of the drained blood had a mean time of 20 minutes. These two factors would now make a cost comparison more favourable.

Moreover, although autologous transfusion was not shown to be cost-effective, it should be noted that this analysis was short term and ignores the value attached to the reduced risk of transmission of virus-related illness. A recent US study[16] has estimated median willingness to pay for autologous blood to reduce this risk at $900 per patient, which is considerably more than the excess cost per patient in the experimental arm of the present study. It seems that the reluctance to adopt such techniques in routine practice is due to a number of factors, including cost, organisation and perhaps motivation. The recent increase in the cost of production of all red cell products, due to improved testing for hepatitis C (nucleic acid testing) and the leucodepletion of all blood products to decrease the risk of nvCJD transmission, may make PRCS more attractive to hospitals.

It is hoped that if cost is the most important driver, then significant reduction in red cell use, without increased morbidity or mortality, might aid the motivation and organisation of such transfusion alternatives.

References

1. Murphy P, Heal JM, Blumberg N. Infection or suspected infection after hip replacement surgery with autologous or homologous blood transfusion. *Transfusion* 1991;**31**:212–17.
2. Heddle NM, Brox WT, Klama LN, Dickson LL, Levine MN. A randomised trial on the efficacy of an autologous blood drainage and transfusion device in patients undergoing elective knee arthroplasty. *Transfusion* 1992;**32**:8:742–6.
3. Clements DH, Thomas MD, Sculco P, *et al.* Salvage and reinfusion of postoperative sanguineous wound drainage. *J Bone Joint Surg* 1992;**74A**:5:646–51.
4. Goulet JA, Arbor A, Michigan TJ, *et al.* Intraoperative autologous transfusion in Orthopaedic patients. *J Bone Joint Surg* 1989;**71A**:3–8.
5. Lisander B, Ivarsson I, Jacobsson SA. Intraoperative autotransfusion is associated with modest reduction of allogeneic transfusion in prosthetic hip surgery. *Acta Anaesthesiol Scand* 1998;**42**:707–12.
6. Shenoliker A, Wareham K, Newington D, Thomas D, Hughes J, Downes M. Cell salvage auto transfusion in total knee replacement surgery. *Transfus Med* 1997;**7**:277–80.
7. EuroQol Group. EuroQol. A new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**;199–208.
8. Newman JH, Bowers M, Murphy J. The clinical advantages of autologous transfusion. *J Bone Joint Surg* 1997;**79B**:630–2.
9. Warner DO, Warner DR, Schroeder KP, *et al.* Changing transfusion practices in hip and knee arthroplasty. *Transfusion* 1998;**38**:738–44.
10. Wahr J. Risk of under transfusion. Thematic Reports. 1st Annual Meeting Network for Advancement of Transfusion Alternatives. 31 January–1 February 2000.
11. Carson JL, Duff A, Berlin JA, Lawrence VA, Poses RM, Huber EC, *et al.* Perioperative blood transfusion and postoperative mortality. *JAMA* 1998;**279**:199–205.
12. Spahn DR, Zollinger A, Schlumpf RB, Stohr S, Seifert B, Schmidt ER, *et al.* Hemodilution tolerance in elderly patients without known cardiac disease. *Anaesth Analges* 1996;**82**:681–6.
13. Goodnough LT, Shafron D, Marcus RE. The impact of preoperative autologous blood donation on orthopaedic surgical practice. *Vox Sang* 1990; **59**:65–9.
14. Nelson AH, Fleisher LA, Rosenbaum SH. Relationship between postoperative anaemia and cardiac morbidity in high-risk vascular patients in the intensive care unit. *Crit Care Med* 1993; **21**:860–6.
15. Czer LS, Shoemaker WC. Optimal haematocrit value in critically ill postoperative patients. *Surg Gynaecol Obstet* 1978;**147**:363–8.
16. Lee LS, Neumann P, Hallowell C, Cannon M, Weinstein M, Johannesson M. Patients' willingness to pay for autologous blood donation. *Health Policy* 1997;**40**:1–12.

Appendix 5

SF-36

Instructions

This survey asks for your views about your health. This information will help keep track of how you feel and how well you are able to do your usual activities.

Answer every question by marking the answer as indicated. If you are unsure about how to answer a question, please give the best answer you can.

1. In general, would you say your health is: (circle one)

Excellent	1
Very good	2
Good	3
Fair	4
Poor	5

2. Compared to one year ago, how would you rate your health in general: (circle one)

Much better now than one year ago	1
Somewhat better now than one year ago	2
About the same as one year ago	3
Somewhat worse now than one year ago	4
Much worse now than one year ago	5

3. The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?
(circle one number on each line)

Activities	**Yes, limited a lot**	**Yes, limited a little**	**No, not limited at all**
(a) **Vigorous activities**, such as running, lifting heavy objects, participating in strenuous sports	1	2	3
(b) **Moderate activities**, such as moving a table, pushing a vacuum cleaner, bowling or playing golf	1	2	3
(c) Lifting or carrying groceries	1	2	3
(d) Climbing **several** flights of stairs	1	2	3
(e) Climbing **one** flight of stairs	1	2	3
(f) Bending, kneeling or stooping	1	2	3
(g) Walking **more than a mile**	1	2	3
(h) Walking **half a mile**	1	2	3
(i) Walking **one hundred yards**	1	2	3
(j) Bathing or dressing yourself	1	2	3

4. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?
(circle one number on each line)

	Yes	No
(a) Cut down on the **amount of time** you spent on work or other activities	1	2
(b) **Accomplished less** than you would like	1	2
(c) Were limited in the **kind** of work or other activities	1	2
(d) Had **difficulty** performing the work or other activities (for example, it took extra effort)	1	2

5. During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?
(circle one number on each line)

	Yes	No
(a) Cut down the amount of time you spent on work or other activities	1	2
(b) Accomplished less than you would like	1	2
(c) Didn't do work or other activities as carefully as usual	1	2

6. During the past 4 weeks, to what extent has your physical health or emotional problems interfered with your normal social activities with family, friends, neighbours, or groups?
(circle one)

Not at all	1
Slightly	2
Moderately	3
Quite a bit	4
Extremely	5

7. How much bodily pain have you had during the past 4 weeks? (circle one)

None	1
Very mild	2
Mild	3
Moderate	4
Severe	5
Very severe	6

8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? (circle one)

Not at all	1
A little bit	2
Moderately	3
Quite a bit	4
Extremely	5

9. These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks – (circle one number on each line)

	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
(a) Did you feel full of life?	1	2	3	4	5	6
(b) Have you been a very nervous person?	1	2	3	4	5	6
(c) Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
(d) Have you felt calm and peaceful?	1	2	3	4	5	6
(e) Did you have a lot of energy?	1	2	3	4	5	6
(f) Have you felt downhearted and low?		2	3	4	5	6
(g) Did you feel worn out?	1	2	3	4	5	6
(h) Have you been a happy person?	1	2	3	4	5	6
(i) Did you feel tired?	1	2	3	4	5	6

10. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting with friends, relatives, etc.)? (circle one)

All of the time	1
Most of the time	2
Some of the time	3
A little of the time	4
None of the time	5

11. How TRUE or FALSE is each of the following statements for you? (circle one number on each line)

	Definitely true	Mostly true	Don't know	Mostly false	Definitely false
(a) I seem to get ill more easily than others	1	2	3	4	5
(b) I am as healthy as anybody I know	1	2	3	4	5
(c) I expect my health to get worse	1	2	3	4	5
(d) My health is excellent	1	2	3	4	5

Appendix 6

The UK Inflammatory Bowel Disease (IBD) Questionnaire

The following questions ask about your bowel problem and how it affected your life over the last two weeks. Please tick **one** answer for each of the questions. If you are unsure about how to answer any question, just give the best answer you can. Do not spend too much time answering, as your first thoughts are likely to be the most accurate.

1. On how many days over the last two weeks have you had loose or runny bowel movements?
   (a) none
   (b) on one or two days only
   (c) on three to seven days
   (d) on eight to fourteen days (i.e. more than every other day)

2. On how many days over the last two weeks have you felt tired?
   (a) none
   (b) on one or two days only
   (c) on three to seven days
   (d) on eight to fourteen days (i.e. more than every other day)

3. In the last two weeks have you felt frustrated?
   (a) No, not at all
   (b) Yes, some of the time
   (c) Yes, most of the time
   (d) Yes, all of the time

4. In the last two weeks, has your bowel condition prevented you from carrying out your work or other normal activities?
   (a) No, not at all
   (b) Yes, for one or two days
   (c) Yes, for three to seven days
   (d) Yes, for eight to fourteen days (i.e. more than every other day)

5. On how many days over the last two weeks have you opened your bowels more than *three* times a day?
   (a) none
   (b) on one or two days only
   (c) on three to seven days
   (d) on eight to fourteen days (i.e. more than every other day)

6. On how many days over the last two weeks have you felt full of energy?
   (a) none
   (b) on one or two days only
   (c) on three to seven days
   (d) on eight to fourteen days (i.e. more than every other day)

7. In the last two weeks did your bowel condition prevent you from going out socially?
   (a) No, not at all
   (b) Yes, some of the time
   (c) Yes, most of the time
   (d) Yes, all of the time
   (e) does not apply to me

8. On how many days over the last two weeks have your bowels opened accidentally?
   (a) none
   (b) on one or two days only
   (c) on three to seven days
   (d) on eight to fourteen days (i.e. more than every other day)

9. On how many days over the last two weeks have you felt generally unwell?
   (a) none
   (b) on one or two days only
   (c) on three to seven days
   (d) on eight to fourteen days (i.e. more than every other day)

10. In the last two weeks have you felt the need to keep close to a toilet?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

11. In the last two weeks, has your bowel condition affected your leisure or sports activities?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time
    (e) does not apply to me

12. On how many days over the last two weeks have you felt pain in your abdomen?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

13. On how many nights over the last two weeks have you been unable to sleep well (days if you are a shift worker)?
    (a) none
    (b) on one or two nights only
    (c) on three to seven nights
    (d) on eight to fourteen nights (i.e. more than every other night)

14. In the last two weeks have you felt depressed?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

15. In the last two weeks have you had to avoid attending events where there was no toilet close at hand?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

16. On how many days over the last two weeks, have you had a problem with large amounts of wind?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

17. On how many days over the last two weeks have you felt off your food?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

18. Many patients with bowel problems have worries about their illness. How often during the last two weeks have you felt worried?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

19. On how many days over the last two weeks has your abdomen felt bloated?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

20. In the last two weeks have you felt relaxed?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

21. In the last two weeks have you been embarrassed by your bowel problem?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

22. On how many days over the last two weeks have you wanted to go back to the toilet immediately after you thought you had emptied your bowels?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

23. In the last two weeks have you felt upset?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

24. On how many days over the last two weeks have you had to rush to the toilet?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

25. In the last two weeks have you felt angry as a result of your bowel problem?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

26. In the last two weeks, has your sex life been affected by your bowel problem?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time

(d) Yes, all of the time
(e) does not apply to me

27. On how many days over the last two weeks have you felt sick?
    (a) none
    (b) on one or two days only
    (c) on three to seven days
    (d) on eight to fourteen days (i.e. more than every other day)

28. In the last two weeks have you felt irritable?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

29. In the last two weeks have you felt lack of sympathy from others?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

30. In the last two weeks have you felt happy?
    (a) No, not at all
    (b) Yes, some of the time
    (c) Yes, most of the time
    (d) Yes, all of the time

Appendix 7

Contingent valuation exercise parts 1 and 2

Study A: Inflammatory bowel disease

Part 1

You are a member of a Commissioning Panel that funds research out of public money (e.g. Wales Office for R&D). The total amount of money available to you to spend on research is fixed. A call for proposals of up to a maximum of £200,000 each has gone out.

The total cost of all proposals received that are fundable (in the sense that they address important issues, are methodologically sound, etc.) far exceeds the available funds and you will have to make funding choices.

One proposal is to investigate open-access versus routine follow-up for patients with IBD. This proposal has been banded alpha plus (the highest banding) by all external referees. The panel agrees with this banding and are keen to fund it.

The abstract to the proposal reads as follows:

> "This project will test the hypothesis that open access follow-up of patients with chronic relapsing disease is more effective and more responsive to patient and general practitioner needs than conventional follow-up by pre-booked appointments. The study will build on work already done to improve the shared care of patients with gastrointestinal disease. The optimum method of follow-up for those who need to remain under joint care will be identified by detailed study of patients with inflammatory bowel disease (IBD) attending outpatients in a busy district general hospital. A randomised controlled trial will compare conventional follow-up at booked appointments with open access follow-up at the request of the patient or GP. Cost, clinical effectiveness, patient and carer preference will be evaluated. A cumulative summary of the patient's progress will be used to ensure that both primary and secondary carers are fully informed of all events. The guidelines which result from the study and the methods developed to improve communication will be applicable to other specialties where patients may come under prolonged follow-up."

Design

Pragmatic two-centre RCT.

Subjects

One hundred and eighty adult patients (78 Crohn's disease; 77 ulcerative or indeterminate colitis; 25 ulcerative or idiopathic proctitis) recruited from outpatient clinics.

The applicants have produced the proposal in a way that offers two options. Their preferred option (a) will use designed data and will cost £97,772. They also offer a second option (b) which will be restricted solely to data that is collected routinely in electronic format, but will cost £83,490, i.e. £14,282 less than option (a).

The table overleaf identifies research questions and data sources to be used under each option. Some routine data sources will not be able to provide data on all patients.

Data item	Option a (designed)	Option b (routine)
Research question = Health outcomes		
Health status	SF-36/UK-IBDQ	GeneCIS (Neath[a] only) PEDW (30% of patients only)
Research question = costs		
Resource use (secondary care)	Hospital notes (paper records)	Pathology GP notes PAS (Morriston[b] only) Radiology (Neath only) Theatre (Neath only) GeneCIS (Neath only)
Resource use (primary care)	GP notes (electronic)	GP notes (electronic)
Patient travel costs	Patient self-report by questionnaire AA motoring costs	AA Route Finder (distance AA motoring costs
Research question = Doctor and patient preferences		
Patient preferences	Patient self-report by questionnaire	Cannot be examined
GP preferences	Interviews	Cannot be examined
Total project cost	£97,772	£83,490

[a] Neath patient represent 54% of the sample.
[b] Morriston patients represent 46% of the sample.

Which of the two options will you fund?
a _____ b _____

IF you chose option (a), how much cheaper would option (b) have to be to persuade you to change your mind? [Note that option (b) is already £14,282 cheaper than option (a)]

☐ up to £15,000 cheaper
☐ between £15,001 and £30,000 cheaper
☐ between £30,001 and £50,000 cheaper
☐ more than £50,000 cheaper (please specify _____)
☐ would not choose option (b) at any price

IF you chose option (b), how much extra would you be willing to pay to get option (a)? [Note that you were not willing to pay the extra £14,282 for option (a)]

☐ 0
☐ up to £5,000
☐ between £5001 and £10,000
☐ between £10,001 and £14,281

Part 2

In the event, option (a) (designed data) was funded. Just for fun, however, the research team also undertook option (b) as a parallel study.

Attached are

(1) a list of variables used in the designed study together with their routine data surrogates (where available) and sources
(2) a list of routine data sources, their level of completeness and comments
(3) 'reconciliation' tables comparing results obtained by each of the two studies showing intervention versus control group differences (95% confidence interval of differences in means) for
    - health outcomes
    - resource use in secondary care
    - resource use in primary
(4) an overall 'reconciliation' table showing how each study answered the study questions.

On the basis of this information

(1) Do you feel the decision to pay the extra £14,282 was justified?
(2) How much extra would you now be willing to pay for option (a)?

☐ 0
☐ up to £5,000
☐ between £5,001 and £15,000
☐ between £15,001 and £30,000
☐ more than £30,000

If >£30,000, please specify figure, which must be no more than £129,265, which would bring the project to the limit of £200,000. _____

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Kent Woods,
Director, NHS HTA Programme & Professor of Therapeutics, University of Leicester

Professor Bruce Campbell, Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Professor Shah Ebrahim, Professor in Epidemiology of Ageing, University of Bristol

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Kent Woods, Director, NHS HTA Programme, Department of Medicine and Therapeutics, Leicester Royal Infirmary, Robert Kilpatrick Clinical Sciences Building, Leicester

Chair,
Professor Shah Ebrahim,
Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol

Deputy Chair,
Professor Jenny Hewison,
Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine, Leeds

Professor Douglas Altman, Professor of Statistics in Medicine, Centre for Statistics in Medicine, Oxford University, Institute of Health Sciences, Cancer Research UK Medical Statistics Group, Headington, Oxford

Professor John Bond, Professor of Health Services Research, Centre for Health Services Research, University of Newcastle, School of Health Sciences, Newcastle upon Tyne

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield, ScHARR Regent Court, Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Cloudside, Rossendale, Lancs
and
Principal Research Fellow, Clinical Therapeutics in the School of Pharmacy, Bradford University, Bradford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York, Research Section, Seebohm Rowntree Building, Heslington, York

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen, Lilian Sutton Building, Foresterhill, Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen

Professor Alastair Gray, Director, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Mark Haggard, Director, MRC ESS Team, CBU Elsworth House, Addenbrooke's Hospital, Cambridge

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham, Primary Care and Clinical Sciences Building, Edgbaston, Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University, Coventry

Dr Donna Lamping, Senior Lecturer, Health Services Research Unit, Public Health and Policy, London School of Hygiene and Tropical Medicine, London

Professor David Neal, Professor of Surgical Oncology, Oncology Centre, Addenbrooke's Hospital, Cambridge

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine, London

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital NHS Trust, Bramwell Dott Building, Edinburgh

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health & Social Care, St George's Building, Portsmouth

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Dr Paul Cockcroft, Consultant Medical Microbiologist/ Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Jane Franklyn, Professor of Medicine, University of Birmingham

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust, Devon

Mr Tony Tester, Chief Officer, South Bedfordshire Community Health Council, Luton

Dr Andrew Walker, Senior Lecturer in Health Economics, University of Glasgow

Professor Martin J Whittle, Head of Division of Reproductive & Child Health, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital

Professor Tony Avery, Professor of Primary Health Care, University of Nottingham

Professor Iain T Cameron, Professor of Obstetrics & Gynaecology, University of Southampton

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre, Bushey, Herts.

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Professor Alastair Gray, Professor of Health Economics, Institute of Health Sciences, University of Oxford

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter

Professor Terence Stephenson, Professor of Child Health, University of Nottingham

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London

Professor Dame Jenifer Wilson-Barnett, Head of Florence Nightingale School of Nursing & Midwifery, King's College, London

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell,
Consultant Vascular and General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of Clinical Support & Health Protection, Directorate of Health and Social Care (North), Department of Health, Manchester

Professor John Bond, Head of Centre for Health Services Research, University of Newcastle upon Tyne

Mr Michael Clancy, Consultant in A & E Medicine, Southampton General Hospital

Dr Carl E Counsell, Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital, Derby

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Ms Bec Hanley, Freelance Consumer Advocate, Hurstpierpoint, West Sussex

Professor Alan Horwich, Director of Clinical R&D, The Institute of Cancer Research, London

Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health, London

Mr George Levvy, Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay, Professor of Psychiatry for the Elderly, University of Leicester

Dr Mike McGovern, Senior Medical Officer, Heart Team, Department of Health, London

Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust

Professor Mark Sculpher, Professor of Health Economics, Institute for Research in the Social Services, University of York

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, University of Aberdeen

Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British Health-Care Industries, London

Ms Judith Brodie,
Head of Cancer Support Service, Cancer BACUP, London

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury, Bucks

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Mr John A Cairns,
Professor of Health Economics, Health Economics Research Unit, University of Aberdeen

Professor Howard Stephen Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, University of York

Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric Epidemiology, London

Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield, West Sussex

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Servs., West Middlesex University Hospital, Isleworth, Middlesex

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen, Dorset

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead, Surrey

Dr Andrew Mortimore,
Consultant in Public Health Medicine, Southampton City Primary Care Trust

Dr Sue Moss,
Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, Surrey

Professor Jon Nicholl,
Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield

Mrs Julietta Patnick,
National Co-ordinator, NHS Cancer Screening Programmes, Sheffield

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health, London

Professor Sarah Stewart-Brown,
Director HSRU/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278