WORK IN PROGRESS: PLEASE DO NOT QUOTE WITHOUT PERMISSION OF THE AUTHORS.

Estimating the marginal value of 'better' research output: 'Designed data' v. 'routine data' randomised controlled trials.

David Cohen[1], Professor of Health Economics
Mirella Longo[1], Research Fellow
John Williams[2], Professor of Health Services Research
W Y Cheung[2], Lecturer
Hayley Hutchings[2], Applied Statistician
IT Russell[3], Professor of Health Sciences

1. University of Glamorgan
   Business School
   Pontypridd
   CF37 1DL

2. University of Wales, Swansea
   School of Postgraduate Studies in Medical and Health Care

3. University of York
   Department of Health Sciences and Clinical Evaluation.

Paper presented to the Health Economists' Study Group Meeting, City University, London, September, 2001.

Abstract:

The range and quality of routinely collected NHS data is improving rapidly. We recently completed a study which explored whether routine datasets can reliably be used as a basis for health technology assessment by randomised controlled trial. The study showed that routine data can reduce the cost of RCTs but will also reduce the quality of the resulting information. Here we attempt to place a value on the 'better' information from (primary) designed data studies.

A mock grants committee was presented with two versions of research proposals; more costly versions using designed data and cheaper versions using routine data, and asked *(ex ante)* which it would fund and how much extra it would be willing to pay for the better (designed data) study. Results of both analyses were then presented and the committee asked *(ex post)* how much extra it would now be willing to pay for the better information of the designed data study.

The committee expressed a general 'lack of trust' towards routine data and placed high values on the better information from the designed data studies - particularly information on preferences.

This exercise contained elements of contingent valuation (willingness to pay) and 'implied values' as methods of valuing intangible benefits. In the former case the group (n = 1) valuation is questionable in terms of the utility theory which underpins 'normal' WTP studies conducted across populations. The latter was unusual in that it allowed assessment of both *ex ante* and *ex post* implied values. We are looking to HESG members for ideas on how best to develop this further.

1. Introduction.

The range and quality of routinely collected NHS and other health related data is improving rapidly. In addition to the obvious benefits this may have for epidemiological or audit purposes, attention has recently focused on the role that routine datasets may play in health technology assessment (HTA). In particular, the use of routine datasets in observational and other types of studies is increasingly being seen as way of reducing current reliance on randomised controlled trials (RCT). (Raftery et al, 2001). While such uses of routine datasets may indeed have a role to play in informing evidence based healthcare, the advantages of randomisation in terms of reducing bias and controlling for extraneous effects are well established and the randomised controlled trial is likely to remain the 'gold standard' for health technology assessment.

We recently conducted a study which explored whether routine datasets can reliably be used as a basis for health technology assessment by randomised controlled trial. This was done by replicating four 'conventional' RCT's which were either ongoing or recently completed, and substituting routine data for the designed (i.e. customised) data which had been, or was being, collected for each trial. The two analyses for each study were undertaken by different statisticians working independently. To estimate the cost of running each trial solely with routine data, we modelled the research process from study conception to final writing up and substituted routine for designed data activities at appropriate points. This allowed a direct comparison to be made of both the costs and the outcomes of the two approaches to RCTs.

Details of methods and results are to be published as a Health Technology Assessment report from the NHS R&D HTA Programme which funded the research. As anticipated, the project showed that routine data versions were in all cases less costly than their designed data counterparts. They were able to answer most but not all research questions but overall, the information they provided was consistently inferior to that yielded by their designed data counterpart.

The present paper is not concerned with the specifics of determining difference in costs and outcomes of the two approaches to research. Rather, it addresses a problem common to many economic evaluations - how to deal with results which show one alternative to be both more effective and more costly than another. Such a cost benefit issue (is the better outcomes worth its extra cost?) is often avoided within cost effectiveness / cost utility analyses, by comparing the resulting incremental cost effectiveness ratios (ICER) (e.g. incremental cost/QALY) with those from other interventions or against some pre-determined threshold above which a health service has indicated it is not prepared to pay.

Apart from methodological criticisms of such an approach (e.g. Donaldson, 1998), the multi-dimensional aspects of research output i.e. the information yielded by the studies, make it impossible to specify a single outcome measure and hence derive an ICER. While 'better' research output is clearly valued more highly than 'less good' research output, specification of 'how much better' and 'in which ways better' is only possible in descriptive terms which differ between

studies. Accordingly the cost benefit question cannot be avoided and some attempt to estimate the money value of 'better' is therefore necessary. This paper describes our attempt to value the benefits of HTA research.

1.1 Valuing benefits:

'Implied values' and 'willingness to pay' are two very different approaches to teasing out the money value of intangible benefits. The former, somewhat out-of-fashion approach, is based on the idea that money values are implied every time a resource allocation decision is taken - even when the decision makers are of the opinion that valuing intangible outcomes such as pain or human life in monetary terms is at best distasteful and probably immoral. Essentially, this approach involves using market or shadow prices to value all of the costs and some of the benefits (e.g. resource savings and productivity gains). Subtracting the money valued benefits from the costs identifies the net cost of producing the remaining benefits. A decision to proceed with any proposal implies that the remaining benefits are worth at least the net cost of producing them. Similarly, a decision not to proceed implies that they are worth less than the net cost.

Mooney (1992) has argued that where these benefits are in single output terms such life years or QALYs, then a series of (marginal) implied values from programmes producing these outputs would allow decision makers to become more consistent in their decision making. It would also allow them to become more efficient since any decisions to shift resources from programmes with high marginal implied values to programmes with low marginal implied values, will produce resource neutral net output gains.

Use of the implied values approach has tended to fall out of favour in recent years largely because of the huge range of values implied from past decisions. Mooney's oft quoted table of implied values of a life runs from £50 to £20,000,000 (Mooney, 1979).

The implied value approach can, however, be useful even where there are multiple, non-money valued benefits. Here, decision makers are presented with a description - in whatever terms are most appropriate - of the full range of non-money valued benefits which are then considered against the net cost of producing them. Although this doesn't allow comparison with other programmes, it has the advantage of reducing a complex issue to a single (but by no means simple) question - is this 'package' of benefits worth the net (opportunity) cost of producing them? The trade-off being considered is made explicit and decision makers are now aware that a decision to proceed will imply that they value this benefit package at least at its opportunity cost while a decision not to proceed implies that they value the package at less. In the present study the value of research output was viewed as such an intangible benefit 'package'.

The willingness-to-pay (WTP) method on the other hand, is currently very much *in* fashion and is increasingly being used in health economic research. This approach is firmly rooted in the principles of welfare economics and is widely accepted as the theoretically correct benefit measure for cost benefit analyses (Johannesson and Jonsson 1991). An individual's WTP for something is that

amount of money which, if paid, leaves them on the same utility level i.e. the value of a benefit is equal to the value of what is forgone. If the principle of potential Pareto improvement is accepted, then information on the WTP of all individuals who make up a society can identify where social welfare gains can be made (see for example Sugden and Williams, 1978).

WTP has been used to estimate values of private goods and services (such as most health care) which, for whatever reasons, are provided at zero price (e.g. Ryan, 1996) as well as for public goods (e.g. Shackley and Dixon, 2000) . In our case, we could have regarded research output as a private good since it can in principle be withheld from anyone not willing to pay for it. However, the current emphasis on 'evidence based medicine' and the importance attached to the dissemination of research results suggests that research output ought be regarded as a public good. Valuation of public goods is normally based on a social WTP derived from individual WTPs which in turn are based on expected individual utility gains from consumption of the good. In the present study, however, research outputs benefit the individual only in a very indirect way (better evidence based health care leading to better health outcomes). It would thus clearly be unrealistic to attempt to measure the utility that individuals gain from research output. An alternative conception, in which those who fund research derive utility directly from the value of the output ultimately produced was used in the present study.

2. Scope of Routine Data

Within the definition of routine data we included all electronically stored patient-encounter based data, collected as part of the process of delivery of health care. Some of these data are collected by provider organisations specifically for central returns or contracting, while others are collected by health care professionals or clinical teams to inform the delivery of care to individual patients. Note that the condition that the data be electronically stored is not specified in some definitions of routine data (e.g. Raftery et al 2001).

While use of designed data was not permitted within the routine data studies, it was nevertheless often possible to manipulate routine data to produce surrogates for various designed data items. For example health related quality of life was measured in one study (shared care for inflammatory bowel disease) using a validated disease specific measure, the UKIBDQ (Cheung et al, 2000). To produce a 'surrogate UKIBDQ' we used the original UKIBDQ questionnaire as a template to specify symptoms, signs and diagnostic codes that could be identified in the routine systems available. Surrogate sub-scale were then generated by adding up the relevant items and the calculated surrogate scores transformed to make them comparable with the original UKIBDQ scores. (Details of methods of deriving surrogates will be given in the HTA report and in papers being prepared for publication). As long as such surrogates were developed solely from routine sources they conform to the principles of a routine data HTA and could be used in the routine data studies.

Table 1 lists the routine databases used in the present study. A glossary of terms is given in Appendix A.

Table 1. Sources of Routine Data for the Present Study.

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)	NHS Trusts	Administrative, demographic and some clinical data
CIS (eg GeneCIS)	Clinical Departments of NHS Trust	Administrative, demographic and detailed clinical data on all patient contacts in hospital
Pathology	Hospital Pathology Departments	Demographic and pathological data
Radiology	Hospital Radiology Departments	Demographic data and imaging reports
Laboratory	Hospital Laboratories	Demographic data and test results
TheatreMan	Hospital Operating Theatres	Demographic, administrative and procedure data
Casemix eg CFIS	NHS Trusts	Administrative, demographic, financial and some clinical data
GP	Primary Care	Demographic, administrative and clinical data

3: Method of cost comparisons:

In the case of the designed data studies (i.e. the actual RCTs which were being, or had recently been, carried out), a 'top down' costing approach was used wherein the research staff who had been involved in each study were asked to estimate what proportion of the total time spent on the study had been devoted to a range of data and non-data related activities. Cost was then estimated by apportioning the total time that each researcher had been funded to work on the project to the data related activities and valued at current hourly gross employment costs for the relevant grade. In the case of the routine data studies, a 'bottom up' approach was used where the time spent in replicating the original studies using routing data was monitored prospectively via timing sheets.

4. Brief Description of the 4 RCTs.

The present paper is concerned with valuing research output and not with details of the RCTs themselves. However, to help to understand the valuation process a brief description of each study is given below.

3.1 Study A: Shared Care of Inflammatory Bowel Disease (IBD)

Objective: To evaluate whether open access follow-up of patients with inflammatory bowel disease is better than routine, booked appointments.

Design: Pragmatic two-centre RCT.

Subjects: 180 adult patients (78 Crohn's disease; 77 ulcerative or indeterminate colitis; 25 ulcerative or idiopathic proctitis) recruited from outpatient clinics.

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

Control: Routine outpatient appointments at interval determined by physician at consultation in clinic.

Main outcome measures: HRQoL, resource use, patient and GP preferences.

Designed data sources:: HRQoL by validated questionnaires (generic = SF-36, disease specific = UKIDBQ). Resource use from patient records (electronic + paper) by research team for 2^0 care, and by postal request to GPs for 1^0 care. Patient and GP preferences and patient borne costs by postal questionnaire. Some GPs were interviewed.

Study costs associated with data collection: Designed data = £19,429 Routine data = £5,665

Discussion: Designed data costs were more than treble those for routine data. Routine data or surrogates were available for most health outcomes and aspects of resource use. Both studies provided similar results, but this may have been due at least in part to the fact that no differences between groups were shown using the 'gold standard' method. No routine data could be found for patient and doctor preferences. The cost of this part of the designed data study made up approximately one third of the total data costs.

The aim of the study was to determine whether open access follow-up 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, as here, health outcomes and costs are shown to be similar for the two methods of follow-up.

3.2 Study B: Community Diagnosis of Obstructive Sleep Apnoea (OSA)

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

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

Subjects :182 patients referred with suspected OSA.

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 : Sensitivity and specificity of the home monitoring system, level of agreement between home and inpatient system, mean bias, number of positive diagnosis made, NHS and patient borne costs.

Sources of designed data : Diagnosis data extracted from the individual systems on a patient by patient basis. Resource use and patient borne costs by patient interview.

Study costs associated with data collection: Designed data = £50,440 Routine data = £4,857

Discussion: The large difference in costs 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. Although 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 difference between the two methods was underestimated in the routine study 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..

3.3 Study C: Clinical Effectiveness of Two Surgical Techniques for the Treatment of Stress Urine Incontinence

Objective: To establish clinical efficacy of two surgical techniques of fascial sling for urinary incontinence.

Design: Pragmatic, three-centre RCT.

Subjects: 165 women with clinically proven stress urine incontinence.

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: Leaked urine volume, HRQoL, urinary urgency, voiding difficulty and patient satisfaction.

Sources of designed data: HRQL from validated questionnaires. Operating time from Theatre Man at one site. Operating time at other sites and all clinical data

recorded on designed forms. Patient satisfaction by questionnaire. Blinded reviews of data on complications, re-admissions and voiding difficulty by clinician report.

Study costs associated with data collection: Designed data = £28,747 Routine data = £2,554

Discussion: Cost of the routine data analysis was here lower than for any of the other studies. This study, however, provides an example where routine data may not be capable of producing meaningful results. The routine data study failed to provide surrogates on a wide number of designed data variables, and did not capture the significant difference in pain which was demonstrated with designed data. Operation time differences reached significance with routine data, but only at the 5% level (c.f. $p<.001$ with designed data).

3.4 Study D: Autologous Blood Transfusion in Total Knee Replacement Surgery

Objective: To compare cell salvage and autologous transfusion with use of donor (homologous) blood in total knee replacement surgery.

Design: Single-centre RCT.

Subjects: 231 patients receiving total knee replacement surgery.

Intervention: Perioperative cell salvage and autologous blood transfusion.

Control: Homologous blood transfusion as required. .

Main outcome measures: Operation time, adverse events, transfusion requirements, wound healing, resource use, HRQoL.

Designed Data sources: HRQoL by validated questionnaire (EuroQol). All other data extracted from patient records (electronic + paper) by dedicated research nurse.

Study costs associated with data collection: Designed data = ££23,357 Routine data = £3,251

Discussion: The 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, quality of life measured using EuroQol (EQ-5D) in the designed data study could not be replicated in the routine data study The routine data study performed poorly in identifying differences in resource use between groups.

3.5 Discussion of study results.

As the present analysis is concerned with *differences* in costs between the two methods, all study costs apart from those directly related to data handling have been ignored. A number of sensitivity analyses were performed but these showed little affect on the base case results.

In all cases the routine data studies were considerably less costly but in all cases produced results which were inferior to their designed data counterparts. From a cost effectiveness perspective, therefore, no dominance was demonstrated. Whether or not routine data studies are an efficient way of conducting RCTs thus depends on the value attached to the better research output.

4. Method used to determining the value of better research output.

There are a number of difficulties in assessing whether either analysis "answered the question". For example, with regard to validity, the designed data version of Study A assessed health outcomes using measures which had been previously validated for patients of this type while the routine data counterpart constructed surrogates.

In the case of the UKIBDQ discussed above, the total surrogate score was shown to be significantly correlated with the original score - although only one of four calculated surrogate sub-scales was significantly correlated with the original sub-scale scores. The surrogate scale was shown to be fairly reliable with a Chronbach alpha of 0.56 Thus while we demonstrated that a reasonably valid and reliable surrogate could be constructed using only routine data, the surrogate was clearly an inferior measure.

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 'better' and may well have been capable of identifying significant differences, had they occurred, which might not have been captured by the surrogate measure.

Another problem is completeness. Routine data were often available only for some study patients e.g. in the case of Study A one routine data source (GeneCis) was available at only one of the two study sites (representing 54% of the total sample) while another (PAS) was only available at the other which reduces statistical power and raises concerns about bias. In other cases a summary variable such as 'total cost' included more items in the designed study than in the routine data study.

Decisions on whether or not to pay for better studies are made by research funding bodies who, in reality, often do have to choose between proposals which address the same study questions but which vary in terms of both quality and costs. Such choices must be made *ex ante* i.e. on predictions of how much better able the higher cost proposals will be at answering 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 a goal to maximise the *value* of research output from finite research money, we attempted to estimate the values that funding bodies would place on better research output by setting up a mock Grants Committee with the following composition:

Chair:
Professor or 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
Professor of Nursing
Consultant in Public Health Medicine
Senior Lecturer in General Practice (x 2)
Senior Lecturer in Health Care Research
Senior Lecturer in Statistics
Senior Lecturer in Health Economics

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 particularly relevant points in the discussion.

The exercise was repeated four times; once for each study. Members were initially presented with two versions of a proposal; a more costly version (D) using (D)esigned data, and a cheaper version (R) using (R)outine data. Sources of data in both versions and estimates of completeness were given. (An example of the information presented is given in Appendix B.).

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, we viewed the group as a single entity making a single (group) decision. However, since information on variations between group members could also be revealing, we elicited individual valuations prior to the group discussions.

Members were therefore initially asked to record their own preferences regarding which version to fund. Choosing the more costly option (D) implied that the marginal value of the *anticipated* better output was at least equal to the difference in costs. Choosing the less costly (R) implied that the marginal value of the better output was less than this. In an attempt to refine these valuations, members who chose (D) were also asked to indicate how much cheaper (R) 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 (R) were asked how much extra they would be willing to pay to get the extra output of (D) (this would have to be less than the difference in costs). Group discussion

then took place and the exercise was repeated as above to obtain the group choices and values.

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

5. Results:

Results of the mock grants committee are shown in Table 2

Table2 : Summary of results of mock grants committee's valuation exercise

	Group Decision			**Individual Decisions**	
	Option Chosen	Implied Value	Willingness to Pay	Option Chosen (n = 9)	Median Willingness to Pay (Range)
IBD - Total study cost (*) Designed data version (D) = £97,772 Routine data version (R) = £83,490					
ex ante	D	>£14,282	£40,000	8D, 1R	£73,886 (£2,500 - £97,772)
ex post	D	>14,282	£55,000	9D	£95,000 (£30,000 - £100,000)
OSA - Total study cost (*) Designed data version (D) = £102,664 Routine data version (R) = £57,081					
ex ante	R	<£45,583	£2,500	2D, 7R	£10,000 (0 - £102664)
ex post	D	>£45,583	£65,000	7D, 2R	£65,000 (£22,500 - £101,495)
Sling - Total study cost (*) Designed data version (D) = £110,642 Routine data version (R) = £84,439					
ex ante	D	>£26,203	£70,000	6D, 3R	£70,000 (0 - £110,642)
ex post	D	>£26,203	£30,000	8D, 1R	£26,000 (£22,500 - £37,500)
Autologous Blood - Total study cost (*) Designed data version (D) = £89,143 Routine data version (R) = £69,037					
ex ante (**)	D	>£20,106	£42,500	4D, 4R (**)	£30,000 (0 - £89,143)
ex post (**)	D	>£20,106	£30,000	8D	£30,000 (£10,000 - £75,000)

(*) = total study cost includes all elements of research including those common to both version e.g. recruitment
(**) = only 8 members participated in this round - deciding vote cast by chairman

Study A: IBD

Ex ante choices and valuations

The group opted to fund D implying that the value of the anticipated extra benefit from designed data was at least £14,282 (£97,772 - £83,490). They indicated that they would have chosen R only if it were at least £40,000 cheaper than D. Eight of nine members individually chose D. Four indicated that they would never fund R regardless of how much cheaper it was, with one stating that R would be unethical. Median marginal willingness to pay for D was £73,886 (range = £2,500 - £97,772). (Note: Refusal to fund R at any price implies that the anticipated marginal value of output from D is equal to the total cost of funding that version).

Key reasons for choosing D 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) the inability of routine sources to provide data on professional or patient preferences and 4) concerns about the ability of routine sources to provide valid data on quality of life.

Ex post choices and valuations:

Having been presented with the results of both analyses, the group agreed that the decision to fund D had been justified. Group valuation of the extra benefits of D were increased to £55,000. The results had largely confirmed expectations regarding 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 D had been justified. Median willingness to pay for D was now £95,000 (range £30,000 - £100,000).

Study B: OSA

Ex ante choices and valuations

The group opted to R implying that anticipated marginal value of D was less than its extra cost of £45,583 (£102,664 - £57,081). It would have been willing to fund D only if its cost were no more than £2,500 more than R. In terms of individual choices, seven of nine members chose R with a median willingness to pay of £10,000 (range = 0 - £102,664) for the extra benefit from D which was less than its incremental cost. The two members who chose D indicated that they would not fund R at any price. These individuals clearly had little influence in the group valuation.

While recognising that D 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 the decision to fund D (which they had not supported) i.e. to pay an extra £45,583 had in fact been justified. They now felt that the extra benefit of D 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 routine data's inability to provide a surrogate for quality of life, recognition of the importance of lost productivity which could not be gleaned from routine data (home monitoring involved much less time off work) and serious concerns about the validity of much of the routine data e.g. 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 D implying that the marginal value of the better study was at least £26,203 (#110,642 - £84,439). It would have been willing to fund R up to an incremental cost of £70,000. Individually, six of nine members agreed with this decision and median willingness to pay was £70,000 (range = 0 - £110,642), There was considerable disagreement over the value of preference data which could only be gleaned from D and to the extent with which quality of life could be proxied by 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 the decision to fund D, i.e. to pay an extra £26,203 had been justified. They now felt that the extra benefits of D were worth at least £30,000. This lower *ex post* valuation was generally due to a feeling that D had not produced as impressive results as had been anticipated. Individually eight of nine members agreed that the decision to fund D had been justified with a median willingness to pay of £26,000 (range = £22,500 - £37,500).

Study D - Autologous Blood

Ex ante choices and valuations

The group opted to fund D implying that the marginal value of the better study was at least £20,106 (£89,143 - £69,037). It would have been willing to fund D up to an incremental cost of £42,500. Individually, four of eight members (one

had to leave at this point of the exercise) agreed with this decision and the chairman had to cast the deciding vote. Median willingness to pay 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 D 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 R 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 the decision to fund D i.e. to pay an extra £20,106 had been justified. They now felt that the extra benefits of D 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 any theatre data from being obtained. Given the importance of length of surgical operation in this study, this was judged to be a major problem. Individually all agreed with the choice of D with median willingness to pay of £30,000 (range £10,000 - £75,000).

6. Discussion:

In the mock grants committee exercises, the *ex ante* exercises represent choices between proposals and choices between valuations of anticipated outputs. In 3 of 4 cases, the choice was made to fund D i.e. the implied value of anticipated extra benefits was at least equal to the differences in costs. For these 3 proposals, the group WTP was roughly double the implied values. For the single decision not to fund D, the group WTP was only $1/20^{th}$ of the implied value. While implied values can only represent either minima (opting for D implies that the value of extra benefits is at least equal to the cost of producing them) or maxima (opting for R implies that the value of extra benefits is less than the cost of producing them) these results suggest that implied values are poor proxies for the more theoretically sound WTP valuations - subject to the potential problems of using a group as the valuation unit discussed above.

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 nevertheless vary considerably from the mean of the individual members. This is not surprising given the very wide ranges of values and the fact that one or two outliers can significantly influencing the mean while having little influence on the group decision. (It is for this reason that medians have been reported).

The *ex post* exercises represent choices between known outcomes with details of the reasons why these results emerged, e.g. 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 study in question

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

1) A '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 to a degree artificial.

2) 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.

3) 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 2nd 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. Ordering the exercises with 4 *ex ante* choices between proposals being followed by 4 *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.

4) Availability of information. For practical reasons, the amount of information given to members with respect to both details of proposals and results had to be kept within certain constraints. A real grants committee would inevitably have more information including comments from referees.

5) Prejudices. There is considerable concern amongst clinicians regarding the validity of routine data, and the discussion reflected this. The views expressed may have been different if there was greater trust in the data.

While obviously limited in its ability to draw firm conclusions, this exercise suggests that the cost reductions which can be achieved through routine data RCTs are unlikely to be justified - given the poorer quality information which such trials provide

Further research is needed to verify these results and to show how this conclusion may change as the range and quality of routine datasets improves over time. It will also be important to examine more closely how marginal benefit/marginal

cost ratio change as individual designed data items are added incrementally to a routine data study.

References

Cheung WY, Garratt AM, Russell IT, Williams JG. The UKIBDQ – a British version of the inflammatory bowel disease questionnaire: development and validation. Journal of Clinical Epidemiology 2000;53:297-306.

Donaldson C. The (near) equivalence of cost-effectiveness and cost-benefit analysis: Fact or fallacy. Pharmacoeconomics 1998; 13:389-396.

Johannesson M., Jonsson B. Economic evaluation in health care: is there a role for cost benefit analysis? Health Policy 1991:17:1-23

Mooney GH. Economics Medicine and Health Care. Harvester Press Hemel Hempstead. 1992

Mooney GH, Values in healthcare. In Lee, K. (ed) Economics and Health Planning. Croom Helm, London, 1979.

Raftery J, Roderick P, Stevens A. Potential Use of Routine Databases in Health Technology Assessment. NHS R&D HTA Programme Report. In press (2001).

Ryan M. Using willingness to pay to assess the benefits of assisted reproductive techniques. Health Economics 1996;5:543-558.

Shackley P, Dixon S. Using contingent valuation to elicit public preferences for water fluoridation. Applied Economics, 2000;32:777-787.

Sugden R, Williams A. The Principles of Practical Cost-Benefit Analysis. Oxford University Press. Oxford. 1978.

Appendix A : List of routine dataset terms

CFIS	Contracting for Information Services
GeneCIS	Generic Clinical Information System
ICD10	International Classification of Disease version 10
PAS	Patient Administrative System
PEDW	Patient Episode Database (Wales)
PIMS	Patient Information Management System
PMS	Patient Management System
QS1	Administrative data on outpatients
SCOPE	System Care Orientated Patient Environment
Theatre Man	Operating theatre information system

Appendix 2 : Information Given to Mock Grants Committee

Part One

You are a member of a Commissioning Panel which 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 v. Routine follow-up for patients with Inflammatory Bowel Disease. This proposal has been banded alpha plus (the highest banding) by all external referees. The panel agrees with the banding 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 patients 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."

The applicants have produced the proposal in a way which offers two options. Their preferred option (D) will use designed data and will cost £97,772. They also offer a second option (R) 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 (D).

The table below identifies the research questions and the 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 D (Designed)	Option R (Routine)
Health Status	SF-36/UK-IBDQ	GeneCIS (Neath* only) PEDW (30% of patients only)
Resource Use (Secondary Care)	Hospital Notes (paper records)	Pathology GP notes PAS (Morriston** 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 AA motoring costs	AA route finder (distance) and AA motoring costs (cost/mile)
Patient preferences	Patient self-report by questionnaire	Cannot be examined
GP preferences	Interviews	Cannot be examined
Total Project Cost	**£97,772**	**£83,490**

*Neath patient represent 54% of the sample
**Morriston patients represent 46% of the sample

Which of the two options will you fund? D _____ R _____

IF you chose option D, how much cheaper would option R have to be to persuade you to change your mind. (Note option R is already £14,282 cheaper than option D)

- □ up to £15,000 cheaper
- □ up to £30,000 cheaper
- □ up to £50,000 cheaper
- □ more than £50,000 cheaper (please specify _____)

IF you chose option R, how much extra would you be willing to pay to get option D. (Note that you were not willing to pay the extra £14,282 for option D).

- □ 0
- □ up to £5,000
- □ up to £10,000
- □ up to £14,281

Part Two

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

Attached are **[note: these are not provided in this HESG paper]**

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) summaries of the conclusions of designed and routine data studies

4) '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

5) 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 R

- □ 0
- □ up to £5,000
- □ up to £15,000
- □ up to £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. _____