

# **MEASURING BENEFIT IN ECONOMIC EVALUATIONS: QALYS ARE NOT ENOUGH**

Nigel Armstrong  
Centre for Health Services Research  
University of Newcastle

The author acknowledges the *fpa* (formerly known as the Family Planning Association) for providing financial support for the research reported in the paper.  
The views expressed are those of the author and not the *fpa*.

**WORK IN PROGRESS: PLEASE DO NOT QUOTE OR CITE WITHOUT THE PERMISSION OF  
THE AUTHOR**

# MEASURING BENEFIT IN ECONOMIC EVALUATIONS: QALYS ARE NOT ENOUGH

## BACKGROUND

In 2001, the Department of Health published the first National Strategy for Sexual Health and HIV [1]. Although various policy recommendations were made in this landmark document, very little economic evidence was used to back these up, particularly in the area of *fertility control* (contraception and abortion). This raised questions of whether such economic evidence exists, and, if there are gaps, in terms of interventions covered or in terms of quality of studies, can research be conducted to mitigate these situations. Reflecting the importance of these questions, the last key report, on the economics of contraception, was published 10 years ago [2].

In light of the above, the *fpa* commissioned a Review of the Economics of Sexual Health in England, overseen by an Expert Panel to take place during April 2003-March 2005. This Review was to have two main aims:

- to assess the literature on the economics of contraception and abortion services in terms of both its quality and relevance to current policy questions; and
- building on the limitations identified, to assess the economic impact of changes in contraception and abortion services.

This project has acted as a vehicle for practical application of the conclusions from a theoretical model for measuring evidence quality for resource allocation decisions. In summary, it implies the need for evidence:

- on both cost and benefit
- for technologies for all populations over which a decision maker (DM) has choice
- on what constitutes cost and what constitutes benefit from the perspective of decision-makers (DMs).

The purpose of this paper is to firstly present progress in the exposition of theory, focussing mostly on the deduction of the need to provide evidence on benefit from the perspective of DMs. In this respect, sexual health seemed to be an ideal area to use given that ‘Current approaches largely fail to recognize the nonmedical benefits of sexual and reproductive health interventions and thus undervalue these interventions’ [3]. Therefore, secondly, the corresponding evidence from the sexual health project is presented. This evidence is presented in two parts: the application of qualitative research to obtain an idea of what constitutes benefit from a DM perspective (Obtaining DM goals) and a systematic review (The review) to discover the extent to which existing economic evaluation studies measure benefit in this way. Only economic evaluations were reviewed on the basis that, according to the aforementioned theoretical model, evidence on both cost and benefit is necessary.

It is important to note early in the paper that the methodology of producing QALYs is not discussed in depth. This is for two main reasons: the purpose is to show from the

research conducted that it became apparent a) that QALYs are insufficient even if necessary and b) other measures might actually be necessary (which might include QALYs).

## THE THEORY

The need for evidence in the form of cost and benefit implies the need for economic evaluation given that economic evaluation provides this [4]. However, economic evaluation can take several forms, depending on whether grounded in *welfare* economic theory or not [5].

Welfare economic theory begins with the assumption that individuals make choices of bundles of goods according to their *preferences* such that the one with the highest *utility* is *preferred*. In this sense it can be stated in such decision-making that the *objective function* of the individual is to *maximise utility*, which is equivalent to saying that the *goal of the decision maker* (DM) is to make the preferred choice. This definition of utility is uncontroversial since it only depends on some notion of preference. Utility simply permits some expression of the relationship between preference and the *attributes* of the choice. Problems arise when one attempts to *estimate* this relationship. This occurs when one wants to estimate, for a change<sup>1</sup> from the status quo, the extent to which any individual will be able to make choices according to preference (i.e. the degree of utility maximisation), in the context of a group of individuals *competing for* a fixed amount of resources. The need to perform an economic evaluation and to measure the extent to which any individual will be able to maximise utility has been driven by one of two *normative* goals:

1) *Pareto optimality* where resources have been allocated such that any other allocation would result in a decrease in utility of at least one individual.

The recognition that any policy is likely to imply a decrease in utility for at least some has led to the goals of *allocative efficiency*<sup>2</sup> (the degree to which there could be a 'Pareto improvement' (the sum across all individuals of the losses and gains in utility))[6, 7], according to one of two other goals:

2) a) *maximising Net Benefit*, which is the marginal total societal gain in welfare (benefit) as measured by marginal willingness to pay (WTP), minus the marginal societal cost, where both are measured in monetary units. This constitutes a Cost Benefit Analysis (CBA). Given that WTP is a function of an individual's utility and resource constraints, the goals of allocative efficiency and maximum net benefit would produce the same resource allocations between individuals.

b) the utilitarian goal of *maximisation of the sum of individual utilities* of each member of a society. This and the need also to allow for uncertainty in individual utility led to the development of 'expected utility theory' [8], which provides the axioms (assumptions) necessary for a cardinal measure of utility i.e. one that

---

<sup>1</sup> This change can be said to be *incremental* or *marginal*, which will be assumed to be the same here.

<sup>2</sup> Allocative efficiency is often abbreviated to 'efficiency', particularly when discussed in the context of 'equity goals'.

measures the extent to which, on average, some will gain and others lose by any policy.

Partly the difficulty of estimating allocative efficiency, but also the belief that welfarist goals are not necessarily consistent with the goals of society led to the development of *extra-welfarism* and, in particular, the goal of *health maximisation* [9]. This is part of the justification for the quality adjusted life year (QALY), which weights expected life years by utility values measured as above [8], but given a limited set of attributes i.e. only those thought to be relevant to health related quality of life. This goal has also been associated with the *decision maker (DM) approach* [10], assuming that a DM would only want to maximise health.

However, it is argued here that the above goals are insufficient and perhaps not necessary *given the purpose of economic evaluation*, which is *to inform decision-making* and given *evidence* on the goals of the actual DMs.

The rejection of, for example, Pareto optimality is not on the grounds of difficulty in achieving, or the questionable validity of, a cardinal measure of utility. It is simply that if economic evaluations are for decision-making then they *ought to measure the extent of maximisation of the goals of DMs*. This might be some function of the utility of others and even be Pareto optimality. However, whether the DM's goal is Pareto optimality is an *empirical question*: it requires evidence on the goals of DMs. For example, although Pareto optimality is not necessarily incompatible with some level of *distribution* (of e.g. utility (through some SWF) or health) it does not by itself *imply* any level. Similarly, the maximisation of health as measured by QALYs (even if assumed to be valid), does not imply equality of health. In fact, given resource constraints and the fairly 'safe' assumption of heterogeneity of production function of health between individuals, achievement of equal distribution of health is logically incompatible with health maximisation [9].

Of course, one might question, on ethical grounds, the desirability of any of the DM goals and indeed one might also question the achievability of the goals of DMs. However, if we accept the assumption that the purpose of research evidence is to inform decision-making then it needs to be in a form that allows those making the decisions to make choices that they *believe* will better fulfil *their* goals. This conclusion led to the following empirical research to obtain goals that would be appropriate for decision-making in the area of fertility control.

## **OBTAINING DM GOALS**

### ***Background to qualitative methods***

It was on this basis that it was decided to gather evidence on the goals of DMs in the area of *fertility control* (contraception and abortion). From this point onwards it is proposed that *benefit* be defined as the extent of fulfilment of DM goals. For example, a goal of health maximisation would imply a measure of health and a goal of maximisation of health equality would imply a measure of health equality (or inequality). This is generalisable to any DM (including any member of society) and

any number of goals (although any decision would imply some weight attached to each goal).

### ***Production of main categories***

In order to facilitate the obtaining of goals, the above theoretical framework was initially drawn on to begin to produce broad categories. The framework essentially arises in the meeting of economics and ethics where, put simply, economics requires an *objective function* and ethics shows what it *ought to be* [9, 11]. These were in terms of *some function of health* given that we are to compare health technologies, the *distribution of this function of health* and *the distribution of the technologies themselves* or *the resources used to produce them*. The aspects of health outcome, distribution and cost could each be expressed as a measure of benefit (e.g. health, health inequality, cost) or transformed into a goal (e.g. ‘improve health’, ‘reduce health inequality’, ‘reduce cost’). It can easily be seen that the general rule for converting a measure of benefit to a goal is to add an expression to indicate the direction of change. There was also a need to allow for goals that were not related to the obvious output of health i.e. what are referred to as *process measures* e.g. politeness. Of course they might be instrumental in achieving another goal; for example politeness, at least for some people, might not be an end in itself, but is important because it contributes to well being. A glossary of these main categories is presented in the results section.

However, such a basic framework can only be a starting point in that it does not tell us whether individuals *actually* do want to achieve these things and it is not exhaustive: in particular it has no specific connection with sexual health. Hence there was a need for sampling, in order to produce the categories of measures of benefit that might be meaningful to DMs in the context of sexual health.

### ***Sampling***

To discover goals, both *key documents* and DM interviews were sampled. In essence statistical sampling rests on sampling theory where we can quantify the probability of obtaining a random sample of individual occurrences of a defined phenomenon (in this case a particular goal) with particular quantifiable characteristics (e.g. frequency) from the population. However, one begins with the assumption that one is unsure what the goals are and also their value (the weights we might attach to make a decision). Hence one is unable at this stage to use statistical sampling. Therefore, the process of compiling a *corpus* with the intention to move towards saturation [12] was used, where one can no longer by further analysis of document or interview find any more goals. One definition of a corpus is “finite collection of materials, which is determined in advance by the analyst, with (inevitable) arbitrariness, and on which he is going to work” (p.96, [13])

Hence Bauer and Aarts [12] advocating a step-wise procedure where one gathers some information, analyses it, and then considers gathering further information. Clearly, it will always be a matter of *judgment* (taking into account the *likely*<sup>3</sup>

---

<sup>3</sup> ‘Likely’ is used here to mean ‘most believable given the evidence’.

*opportunity cost*<sup>4</sup>) as to when to stop sampling, but this is no different to statistical sampling. However the efficiency of the process is increased by attempting to identify domains, the variability of which are likely to be associated with variability in goals. For example, one domain that is relevant to DMs is ‘level in the health service managerial chain leading to the user’, which would include clinicians and sexual health service leads. Therefore, in sampling both documents and interviewees, we attempted to provide as broad a coverage as possible within our resource constraints.

#### Key Documents:

Prior to approaching any DMs in order to discover their goals, the goals explicit or implicit in *Key Documents* were analysed. These documents represented the major sources of policy or policy advice in the area of sexual health: Department of Health [1, 14, 15]; Royal College [16, 17] and charity/user perspective [18, 19]. The results were used to help to decide the appropriateness of the categories, the method of producing subcategories and in designing the interview schedule. These preliminary results are therefore not presented here.

#### Interviews:

DMs were chosen to generate coverage on two domains, decision making at various levels within health care, from the Department of Health through Chief Executive of a PCT to managing a service and service delivery (seeing service users/patients) and the various services/methods. The characteristics of the interviewees are listed in Table 1.

**Table 1** Characteristics of interviewees in survey of decision makers

Interviewee	Decision making level	Services/methods
1	Consultant Physician, GUM	Contraception
2	Co-Director of Contraceptive Services	Contraception
3	Consultant, Contraception and Sexual Health	Contraception, abortion, other sexual health
4	Sexual Health Program Manager	Contraception, abortion, other sexual health
5	PCT sexual health lead	Sexual health
6	Public health	All local health services
7	Campaign organisation representative	Sexual health

#### *Method of data production and extraction*

The following applies to the work with both interviews and documents, unless otherwise stated. Although the aim was to obtain data in the form of goals it was clear from the outset that the information within documents and interviewee output would not always be in a ‘ready to use’ form e.g. ‘the/my goal is...’. The Oxford Paperback Thesaurus defines goal as “objective, aim, end, target, design, intention, intent, plan, purpose; (holy) grail; ambition, aspiration, wish, dream, desire, hope”[20]. Therefore, in order to maximise coverage, any expression in any of these terms was initially included.

<sup>4</sup> ‘Opportunity cost’ is the potential loss in benefit by engaging in one activity (here, sampling) instead of some other.

It was expected that there would also be different ways of expressing the same goal and some uncertainty in knowing the difference. Therefore, one of the aims of the process was to develop the framework for the goals that might be described as a *lexicon* (way of categorising). This led to the creation of sub-categories, as discussed more fully below.

### ***Interview methods***

These could be classified as individual (in contrast to group) and with questions that were open-ended in that the interviewees' responses were not constrained deliberately by the interviewer. The interviewer had a script for introducing the topic, a list of questions and, for each question, a set of prompts. The first set of questions concerned the individual's characteristics:

#### **[Role questions]**

- 1) What is your position in the management chain leading to the service user?
  - 2) Do you have responsibility for a budget?
  - 3) What services/methods does the budget cover?
- Does your role/budget cover more than contraception/abortion services? If so, what else does it cover?

The second concerned their goals:

#### **[Goal questions]**

- 1) Could you give me some insight into what you are trying to achieve/would like to achieve from contraceptive and abortion services [**prompts: health outcome (intermediate, final), QoL, cost/resource use, individual vs societal, distribution/equity, efficiency, other**].
- 2) Do you have any broader goals that you would like to achieve from your resources/budget?
- 3) I would also like you to tell us what you think are useful outcomes that we could use to measure how successful we might be in achieving these goals.
- 4) What kind of information would you like in order to help you to better make decisions regarding these services? [**types of evidence e.g. effectiveness, cost, c-e, systematic reviews/HTAs, observational, health/social/other, individual/patient preferences, institutional**]
- 5) What do you think that research evidence should provide in order to help make decisions?

Only the first in each set were read out. The rest were essentially used as prompts, to encourage conversation and enhance coverage of predicted (from theory and previous interviews) categories of goals. The interviews were recorded in order not to miss anything important and to enhance transparency. Notes were also taken in order for the interviewer to return on any issues that arose during the interview for pursuing at a convenient time during the interview. Also the interviewer wrote a note at the beginning of the interview with interviewee 3 to as a reminder to focus on abortion given the previous two interviewees not having responsibility for this area. Generally, the interviewer judged what balance to strike between encouraging interviewees to

explore freely what they considered to be goals and outcomes and covering the predicted concepts.

The interviews were transcribed without noting pauses, interruptions or non-word utterances. Benefit measures were produced and categorised according to the above framework. This was done in (4) main stages:

- 1) highlighting phrases/passages within each interview transcript that indicate the expression of a goal or benefit measure
- 2) each expression was extracted to produce a list for each interview
- 3) each expression was interpreted to produce a *reduced form*, which appeared to represent the essential meaning of the expression in terms of a measure of benefit. This was inferred from the researcher's interpretation of the phrase as well as the context of the expression within a transcript. For example, in interview 1, "a timely service" and "seen within a reasonable time-frame" were reduced to 'waiting time', which implies the goal of *to reduce waiting time*.
- 4) All reduced forms were collated from all interviews to produce the corpus of sub-categories.

Although the fitting of expressions into the broad categories was largely by deduction, given the definitions provided, the reduction of expressions to the reduced form was a subjective process. However, the path from transcript to corpus is traceable.

## **Results**

Glossary of main categories (generated from theoretical framework):

*Health* and *QoL* are the most obvious outputs of a health care technology and therefore an obvious goal is to improve health or its constituent parts. It is, however, not easy to define, although it should logically include *sexual* health. What is important to state here is that we have included prevention of unintended or unwanted pregnancy. This is because, although they might not easily fit with a definition of health, they are the most obvious outputs of contraception. Health is also differentiated from *QoL* (*quality of life*), which is just as difficult to define [21]. Generally, QoL would be accepted as being broader such that one component of it could be health. The World Health Organization (WHO) defines health as 'A state of complete physical, mental, and social well-being not merely the absence of disease . . .'. WHO defines QoL as '...an individual's perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns.' (<http://www.who.int/evidence/assessment-instruments/qol/ql1.htm>). Here, because we are using benefit to capture fulfilment of goals, QoL is less broad. In fact, it might be closer to the WHO definition of health, which we might call 'health-related QoL', with health referring more to the most obvious outputs of health care technologies and the absence of disease.

*Equity* is yet another difficult concept. A recent discussion paper considering whether the NHS is equitable states: '...an equitable service is defined as one that offers equality of access to health care to individuals in equal need (often referred to as horizontal equity)' (p.6 [22]). However, at its most general, it is simply defined as



‘fairness’, which might be interpreted as ‘treating people equally’. The difficulties in defining the term have been discussed at length [23]. Essentially, equity is used to refer to a set of terms for describing *distributional equality* i.e. how some attribute (*equality of:*) is distributed between individuals who differ by some measure (*equality by:*), such as gender, age or socioeconomic status. This attribute could be, as above, access to health care, but it could be health itself or expenditure. Although these concepts are challenging to define, it is crucial to understand which one is the goal of the DM since it can be shown that fulfilment of one might prevent fulfilment of the others [9]. It is clear that equity terms, like social ones, qualify others. For example one can have a goal of improving health and one of improving health equality.

*Social* is more a qualifier for other terms in that it refers to goals for any people other than the service user. These include family, but also non-family, such as society as a whole. For example, ‘public health’ is translated to *social health*.

*Accessibility* is used interchangeably with *access*. It is distinguished from equity in that, although one DM goal might be to increase equality of access, another might be to increase access irrespective of its distribution. Yet again it is not straightforward to define and includes utilization and costs incurred in receiving health care [24].

*Individual benefit* refers to measures, which are related to the way in which the extent of fulfilment of a goal of the DM is derived from the benefit of other individuals or, in economic terms, by maximising the utility of others i.e. a DM might have a goal of fulfilling the goals of individual service users or society as a whole. Therefore measures of preference or the values of individual members of defined groups would be relevant here.

Sub-categories generated from interviews:

Table 2 provides a summary of the corpus in terms of categories and sub-categories obtained from analysis of the interviews.

**Table 2 Benefits (and disbenefits) of contraception and abortion services**

<b>Main category:</b>	<b>Sub-category:</b>
Health	Sexual health, fertility control, prevention of: unwanted/unplanned pregnancies and births, ectopic pregnancies, infections, stroke, thrombosis, uterine wall perforation, intra-uterine device (IUD) expulsion, etc
Quality of life	Happy/reasonable/good life, self-esteem, knowledge/understanding/education, ability to work or study and avoidance of feeling ill, stress/hassle, pain, worry
Equality of:	Provision, access, benefit, standards, availability, fertility control, choice, health, resources, competencies/skills of commissioners, skills of clinicians
Equality by:	Age, location, capacity to benefit, probability to access, initial health, ethnic group, service, culture, choice, deprivation, disability
Access to:	Fertility control, service, abortion, sex education in schools/sex and relationship education, abortion services, contraception, knowledge/understanding
Other access issues	Waiting time, availability
Social	Benefit, cost, choice, health, preference, values
Individual benefit	Choice, fulfilment of aspirations, goal achievement, respect of individual perspective/values/importance, satisfaction, empowerment/control, happiness

**Discussion**

It is clear from the theoretical framework that DM goals are not only health related and therefore QALYs (even if a valid measure of health) would not be sufficient. However, empirically, the diversity of goals can be seen. Of course, some of the measures in the health or QoL categories could theoretically be aggregated to produce QALY weights. For example, ‘fertility control’ could be seen as another dimension of health related QoL, thus providing a basis for QALYs even if in a different form. This would at least make QALYs more applicable to this area.

Some goals in the individual benefit category might be interpreted as pertaining to welfare, for example choice or fulfilment of aspirations, since they seem to suggest some relationship to the ability to make choices according to preference. Therefore, QALYs would certainly not be sufficient.

Measures of equity are also prevalent. They cover distributional concerns for several population sub-groups, which again are not only health related. However, even if they were all health related and if QALYs were valid measures of health, as pointed out earlier, if the goal is one of distribution by population sub-group, then simply measuring QALYs for a given population is insufficient.

Although this research does not measure the *relative value* of these measures, their expression does imply that it is non-zero.

## LITERATURE REVIEW

### *Methods*

The literature search was also part of a project commissioned by the *fpa* to review economic evaluations in the area of fertility control. It was decided, given the a priori belief that the number of economic evaluations would be small enough to accomplish the task within a year, to review economic evaluations for all populations within the area of fertility control (abortion and contraception)<sup>5</sup>. Therefore we could test the hypothesis that current literature is lacking both in quality of *individual* studies and *as a body* for helping DMs fulfil their goals. A review of the use of benefit measures was therefore important, but formed only one part of the whole review.

Each population was essentially defined according to the technology used, for example *contraceptive method* users (with particular characteristics e.g. female, aged 18-44, not intending to get pregnant). Therefore, studies were searched by *technology category* (corresponding to a broad population). These are individual methods of abortion or contraception (referred to as *contraception methods* and *abortion methods*) as well as what we called 'service delivery modes'. This latter category included comparisons of whole *contraceptive services* as in the study by McGuire and Hughes [25], which might be described as consisting of a contraceptive profile (reflecting the prevalence of individual methods). Other categories added during the review were *contraception other* or *abortion other*, to contain components of provision (e.g. direct referral for laparoscopic sterilisation or prophylactic antibiotic treatment for abortion) and *emergency contraception* (EC) reflecting comparison of methods and means of delivery (e.g. pharmacy prescribed). The search strategy was designed in order to be broad enough to encompass all of these categories as well as not exclude on the basis of standard quality criteria e.g. randomisation or systematic review to obtain parameter estimates for modelling. A pilot search was done in the belief that the number of economic evaluations would be so small as to warrant retrieval additionally of effectiveness only studies. However, this produced over 40,000 studies, which was considered far too many to screen. The strategy was then revised to incorporate a highly sensitive economic filter used to locate economic evaluations for the NHS Economic Evaluation Database (NHS EED) [26]. The database is recommended to be searched to increase sensitivity in Health Technology Assessment (HTA) [27]. This filter was then attached to one that was devised in collaboration with the Expert Group to locate the relevant interventions. The strategy was then adapted for use in four databases, as recommended [27]: Medline, EMBASE, SCI and Popline to cover the period from 1980 to December 2003. The NHS EED was also searched and the pharmaceutical company Schering provided a list of studies from a prior search of the Health Economic Evaluation Database

---

<sup>5</sup> This approach *contains* a typical systematic review of literature, which involves a *pooling* of studies *within the same population*. The basis for doing this is that the more studies that can be examined the better on the same principle as applies to a single study: the larger the sample the better. Therefore, the probability of getting an estimate that is unrepresentative of the population reduces. A review of studies for a given population also allows an analysis of heterogeneity (variability) i.e. different studies producing different results. The sensitivity of the result (place in the decision matrix) to such variability is investigated in the section on key methodological issues.

(HEED); this was used to validate the search strategy. An update was also performed to cover the period from January until end of August 2004.

#### Retrieval and screening:

References were retrieved from Medline first, then EMBASE, SCI, Popline and finally NHS EED, removing duplicates at each stage. In fact, unsurprisingly since we used the same filter as for NHS EED, no studies were found additional to those found in the other databases. Where references included an abstract, these were sorted out. Another filter was then applied to include only those that included the free text items: 'cost', 'economic' or 'price' in either the title or the abstract. The basis for this was the experience of the author in filtering studies for the NHS EED that if a study had an abstract it would contain some details on the method and/or results in terms of cost and that any study that did not contain such information in the abstract was very unlikely to contain any useful cost information. A sample of the references with abstracts revealed no loss of sensitivity. All titles were examined in the references without abstracts. Studies that appeared to be economic evaluations and/or that required viewing of the full paper were screened according to the exclusion criteria as set out below, and if in doubt about relevance to the UK a second reviewer from the Expert Group saw them.

#### Exclusion criteria:

Studies were excluded prior to seeing the full text version if they were in a foreign language. From the full text article, studies were not counted as economic evaluations if they contained *insufficient cost data* on the basis for that an economic evaluation is useful in that it provides an estimate of the association between benefit and cost. Therefore resource use data should either be derived from the same sample as the benefit data or, in a modelling study, cost should be a product of the model and not simply taken from another source. Economic evaluations were excluded in three categories: interventions not relevant to the UK, use of condoms only to affect HIV transmission and. However, ostensibly *cost only* studies were included, which measured benefit as savings as mentioned in the section on costs above. This is because, for contraception studies, either service or method comparisons, assuming that pregnancy prevention outweighs any bad effects of contraception, savings due to pregnancy prevention *imply* an overall benefit increase. On a similar basis, abortion service studies were also included, although the assumption of an overall benefit increase through prevention of birth (as opposed to pregnancy) might be more controversial.

Table 3 shows the results of the retrieval and screening process:

**Table 3: Results of retrieval**

Database	Total retrieved	Full text screened	Economic evaluations	Not relevant to UK	HIV only	Reviewed
Medline	3092	105	35	8	7	20
EMBASE	2382	37	24	10	6	8
SCI	131	20	1	0	0	1
Popline	16	4	4	2	0	2
Update	225	2	1	0	0	1
Total	5846	168	64	20	13	32

To this were added two studies not referred to on any of the databases [25, 28] to make a total of 34 studies.

### ***Results: summary of characteristics of studies reviewed***

The 34 studies have been data extracted. Of these, 11 compared whole contraceptive services with no service, 1 compared a service with over the counter (OTC)<sup>6</sup> oral contraception to without, 9 compared individual contraceptive methods, 3 contraceptive other, 3 EC, and 2 abortion services and 2 each of abortion methods or abortion other. One study [29] was largely a contraceptive method comparison, but did also compare *contraceptive profiles* i.e. in the manner of contraceptive service comparisons. Another [25] was essentially a service comparison, but incorporated a comparison of individual methods, which was reproduced (with minor alterations) in another publication [30]. 6 were from the 1980s, 18 from the 90s and 10 from 2000 to 2004. 18 were conducted in the US and 10 in the UK with 1 each from Canada, Mexico, Peru, Portugal, Thailand and Turkey. The vast majority, 27, derived estimates of benefit and/or cost via modelling as opposed to sample. All 25 contraceptive method, EC or service studies used modelling. 4 out of 6 abortion studies were sample based: only the service ones used modelling.

### ***Results: benefit measures***

In this section comparison should be made to the standard of the measures of benefit elicited from interviews with DMs and shown in Table 2. 19 out of the 25 contraceptive modelling studies where savings were made by prevention of pregnancies (the other modelling study by Hendrix et al [31] assumed tubal ligation and vasectomy had equal benefit in this way and therefore only measured savings due to avoidance of complications) were essentially cost only studies. In fact, nine out of the twelve contraception service studies provided a summary measure anyway in terms of either number of unintended pregnancies/births or incremental change in number of unintended pregnancies or births over the given time horizon for provision of the service. Of the four contraceptive method studies that were essentially cost only [32-35], only Trussell et al 1995 [32] provided a summary measure, in terms of number of pregnancies averted. Of the three EC studies, only one [36] provided a

<sup>6</sup> 'OTC' should be contrasted with 'pharmacy prescribed' in that, although both imply purchase at a pharmacy, the former implies no prescription and prescription implies a *consultation* with the pharmacist and its associated costs.

summary measure, in terms of absolute risk (of unintended pregnancy) difference, from the original trial of pharmacy versus no pharmacy provision. Even so all imply benefit due to pregnancy prevention and, as stated above, a cost saving implies benefit increase, given that the saving is through reduction in unintended pregnancies.

This leaves six studies, which were all contraceptive method comparison. They provided measures of cost-effectiveness in terms of cost per pregnancy averted only [35, 37], or additionally, cost per couple-year of protection (CYP) [25, 30, 38] or cost per QALY [29].

Of the five studies measuring savings also by STI prevention, no summary measure of benefit in these terms was given. However, Sonnenberg et al [29] multiplied expected duration in various 'health states' including PID and HIV, as well as pregnancy, by utility values in order to calculate QALYs. The technique used to elicit these values (time trade-off) is well established [4], although not without controversy regarding validity and the sample was only one of convenience.

Patient satisfaction was measured in only two studies both examining tubal ligation [39, 40], as was the only other non-modelling contraception study [41]. In only the McKessock et al study [39] was there any kind of process measure, that of mean waiting time: clinician satisfaction was also measured in this study. The other two tubal ligation studies were about the surgical procedure itself and also measured rate of complications amongst other things. One of the abortion method studies [42] used process measures (number of gemprost pessaries, induction to abortion interval) as well as rate of incomplete abortions. The other [43] used the proportion that would opt for the other procedure in the future. This showed that of those women who were randomised to either surgical or medical, a statistically significantly greater proportion would switch from medical to surgical than the reverse. However, it was shown that this was critically dependent on gestational age, i.e. the earlier the more likely that a woman would prefer medical, and that reduced pain might play a large part in this. Both abortion other studies [44, 45] used rate of PID as a measure of benefit.

In none of the studies was any notion of equity or a measure of the distribution between individuals (e.g. of health or access) used.

## **DISCUSSION**

It has been deduced that, if we make the weak assumption that economic evaluations are for decision-making, then they need to measure benefit insofar as it is the extent of fulfilment of DM goals. It has also been shown empirically, in the area of sexual health, the diversity of DM goals, only very few of which could be regarded as related to the goals for which economic evaluations have been designed and that existing economic evaluations have barely scratched the surface in measuring benefit.

In fact there is a perception that health care economic evaluations are of poor quality [46]. Indeed, a recent survey of health care decision makers has shown that the perception of poor quality is a major reason for lack of use in health care decision-making [47]. On this basis Jefferson and Demicelli concluded that urgent action is

required to impose stricter quality criteria at the journal peer review stage. Could it not be that economic evaluations are simply not designed to measure what is relevant to DMs?

Clearly, there are several notable limitations to the empirical work. Firstly, given the type of sampling, it is not possible to say, in objective terms, how representative the sample was of any population. What we can say is that measures of benefit other than QALYs and those that are health related cannot be ruled out empirically, given the assumption of the purpose of economic evaluations to inform DMs.

The second is that it was limited to the area of sexual health. However, if one considers health technology assessment, particularly that by NICE, the measure of benefit that is being demanded is the QALY, implying the goal of QALY maximisation. However, even NICE admit to making their decisions not entirely on the basis of QALYs. Also, NICE can be seen as representing basically one DM. The goals obtained above from documents and interviews pertained to several other DM perspectives (stakeholders). Unless NICE imposes decisions at these other levels, it is a tacit acknowledgement that their goals are also of value. Also, interviewees were not limited to only having goals that were to do with sexual health or fertility control, and indeed did express broader ones, such as 'health'. A review of studies examining access to health and social care following hospital discharge found use of a wider range of benefit measures, including social support and self esteem [48]. The question is how the preferences of all DMs (including service users and society as a whole) over all of these putative goals are taken into account.

Thirdly, the review was limited in that it did not include unpublished sources and we had no access to industry studies. However, discussion with members of an expert group convened by the *fpa* and representatives from industry revealed that ongoing research did not measure benefit in any other way.

Not only do studies need to have appropriate measures, but they must be appropriately *designed* to estimate the relationship between technology change and outcome [49]. Very few studies, including effectiveness only have been designed to measure the effect of technologies on inequalities: most research on inequalities has been explanatory [50]. This is of course not to say that all studies should be intervention type or that all intervention studies should measure benefit as defined. It can be argued that, through decision modelling, the full range of study designs and intermediate measures can be synthesised to produce information that is relevant to DMs [51]. However, most of the studies reviewed ostensibly used models, but the most prevalent outcome was cost and most prevalent measure of benefit, which was usually implicit, was reduction in unintended pregnancy rate. Similarly, most HTAs now require decision modelling, but the only outputs are cost and QALYs. Equity issues are still not accounted for explicitly in terms of any objective measure.

In conclusion, this paper has shown the theoretical basis and some empirical evidence for measuring benefit in economic evaluations in terms of the extent of fulfilment of DM goals. Clearly, if there is doubt as to the applicability of these results to other decision-making contexts, the relevant DMs should be identified and their goals sought. What also logically follows, given the existence of multiple DMs and multiple goals, is the continued need for preference elicitation techniques, but where

individuals are shown the consequences of their choices given best available evidence (through decision modelling). There remains at least one concession to welfare economic theory: if individuals were *fully informed* (knew all consequences for themselves in the long run) then utility maximisation might be consistent with a universally acceptable distributional goal. What is clear now is that QALYs cannot be relied upon as the sole measure of benefit.



## REFERENCES

1. Department of Health, *The national strategy for sexual health and HIV*. 2001, London: Department of Health.
2. McGuire, A. and D. Hughes, *The economics of family planning services*. 1994, London: FPA.
3. Singh, S., et al., *Adding it up: the benefits of investing in sexual and reproductive health care*. 2003, New York: The Alan Guttmacher Institute.
4. Drummond, M.F., et al., *Methods for the Economic Evaluation of Health care programmes*. Oxford medical Publications. 1997, Oxford: Oxford University press.
5. Hurley, J., *An overview of the normative economics of the health sector*, in *North-Holland handbook of health economics*, A.J. Culyer and J.P. Newhouse, Editors. 2000, Elsevier: Rotterdam.
6. Hicks, J., *The foundation of welfare economics*. *Economic Journal*, 1939. **49**: p. 696-712.
7. Kaldor, N., *Welfare propositions and interpersonal comparisons of utility*. *Economic Journal*, 1939. **49**: p. 549-552.
8. Loomes, G. and L. McKenzie, *The use of QALYs in health care decision making*. *Social Science and Medicine*, 1989. **28**(4): p. 299-308.
9. Culyer, A.J. and A. Wagstaff, *Equity and equality in health and health care*. *Journal of Health Economics*, 1993. **12**: p. 431-457.
10. Brouwer, W.B.F. and M.A. Koopmanschap, *On the economic foundations of CEA. Ladies and gentlemen, take your positions!* *Journal of Health Economics*, 2000. **19**: p. 439-459.
11. Sen, A., *On ethics and economics*. 1987, Oxford: Blackwell Publishers.
12. Bauer, M.W. and B. Aarts, *2: Corpus construction: a principle for qualitative data collection*, in *Qualitative researching with text, image and sound: a practical handbook*, B.M. W and G. G, Editors. 2000, Sage Publications Ltd.: London.
13. Barthes, R., *Elements of semiology*. 1967, New York: Hill and Wang, the Noonday Press.
14. Department of Health, *The national strategy for sexual health and HIV: implementation action plan*. 2002, London: Department of Health.
15. Department of Health, *Effective commissioning of sexual health and HIV services: a sexual health and HIV commissioning toolkit for primary care trusts and local authorities*. 2003, London: Department of Health.
16. Faculty of Family Planning and Reproductive Health Care of the Royal College of Obstetricians and Gynaecologists, *Service standards for sexual health services*. 2003, London.
17. Royal College of Obstetricians and Gynaecologists, *The care of women requesting induced abortion: evidence-based guideline (No. 7)*. 2000, London: RCOG Press.
18. FPA, *Annual review*. 2001, London: FPA.
19. Mawer, C. and M. McGovern, *Early abortions: promoting real choice for women*. 2003, London: FPA.
20. Waite, M., ed. *The Oxford Paperback Thesaurus*. 2001, Oxford University Press: Oxford.
21. Bowling, A., *Measuring health: a review of quality of life measurement scales*. 2nd ed. 1997, Buckingham: Open University Press.

22. Dixon, A., et al., *Is the NHS equitable? A review of the evidence. LSE Health and Social Care Discussion Paper No. 11.* 2003, London School of Economics: London. p. 1-38.
23. Wagstaff, A. and E. van Doorslaer, *Equity in health care finance and delivery*, in *Handbook of Health Economics*, A.J. Culyer and J.P. Newhouse, Editors. 2000, Elsevier Science B.V.: Amsterdam.
24. Mooney, G., *Key issues in health economics.* 1994, Hemel Hempstead: Prentice Hall/Harvester Wheatsheaf.
25. McGuire, A. and D. Hughes., *The economics of family planning services.* 1994, FPA: London.
26. NHS Centre for Reviews and Dissemination, *Improving Access to Cost-effectiveness Information for Health Care Decision-making: The NHS Economic Evaluation Database, CRD Report No 6.* 2 ed. Vol. 2. 2001, York: University of York.
27. Royle, P. and N. Waugh, *Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out by the National Institute for Clinical Excellence appraisal system.* Health Technology Assessment, 2003. **7**(34).
28. Laing, W.A., *Family planning: the benefits and the costs.* 1982, Policy Studies Institute: London.
29. Sonnenberg, F., et al., *Costs and net health effects of contraceptive methods.* Contraception, 2004. **69**: p. 447-459.
30. Hughes, D. and A. McGuire, *The cost-effectiveness of family planning service provision.* Journal of Public Health Medicine, 1996. **18**(2): p. 189-196.
31. Hendrix, N.W., S.P. Chauhan, and J.C. Morrison, *Sterilization and its consequences. [Review] [67 refs].* Obstetrical & Gynecological Survey, 1999. **54**(12): p. 766-777.
32. Trussell, J., et al., *The economic value of contraception: a comparison of 15 methods.[comment].* American Journal of Public Health, 1995. **85**(4): p. 494-503.
33. Trussell, J., et al., *Medical care cost savings from adolescent contraceptive use.* Family Planning Perspectives, 1997. **29**(6): p. 248-255.
34. Ortmeier, B.G., et al., *A cost-benefit analysis of four hormonal contraceptive methods.* Clinical Therapeutics, 1994. **16**(4): p. 707-713.
35. Ashraf, T., S.B. Arnold, and M. Maxfield, Jr., *Cost-effectiveness of levonorgestrel subdermal implants. Comparison with other contraceptive methods available in the United States.* Journal of Reproductive Medicine, 1994. **39**(10): p. 791-798.
36. Marcianite, K.D., et al., *Modeling the cost and outcomes of pharmacist-prescribed emergency contraception.* American Journal of Public Health, 2001. **91**(9): p. 1443-1445.
37. French, R.S., et al., *Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: Two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.* Health Technology Assessment (Winchester, England), 2000. **4**(7).
38. Phillips, C.J., *Economic analysis of long-term reversible contraceptives. Focus on Implanon.[comment].* Pharmacoeconomics, 2000. **17**(2): p. 209-221.
39. McKessock, L., et al., *A randomized controlled trial of direct access for laparoscopic sterilization.* Family Practice, 2001. **18**(1): p. 1-8.

40. Garcia, F.A., et al., *Economic and clinical outcomes of microlaparoscopic and standard laparoscopic sterilization. A comparison.* Journal of Reproductive Medicine, 2000. **45**(5): p. 372-376.
41. Lemos, P., et al., *The economic benefits of ambulatory surgery relative to inpatient surgery for laparoscopic tubal ligation.* Ambulatory Surgery, 2003. **10**(2): p. 61-65.
42. Tewari, S.K., et al., *Mifepristone in mid trimester termination of pregnancy-value for money?* British Journal of Family Planning, 1995. **21**(1): p. 7-9.
43. Henshaw, R.C., et al., *A prospective economic evaluation comparing medical abortion (using mifepristone and gemeprost) and surgical vacuum aspiration.* British Journal of Family Planning, 1994. **20**(3): p. 64-68.
44. Blackwell, A.L., et al., *Health gains from screening for infection of the lower genital tract in women attending for termination of pregnancy.[comment].* Lancet, 1993. **342**(8865): p. 206-210.
45. Penney, G.C., et al., *A randomised comparison of strategies for reducing infective complications of induced abortion.[comment].* British Journal of Obstetrics & Gynaecology, 1998. **105**(6): p. 599-604.
46. Jefferson, T. and V. Demicheli, *Quality of economic evaluations in health care.* British Medical Journal, 2002. **324**: p. 313-314.
47. Hoffman, C. and J. Matthias Graf von der Schulenburg, *The influence of economic evaluation studies on decision making: A European survey.* Health Policy, 2000. **52**: p. 179-192.
48. Richards, S. and J. Coast, *Interventions to improve access to health and social care after discharge from hospital: a systematic review.* Journal of Health Service Research Policy, 2003. **8**: p. 171-179.
49. Birch, S., *As a matter of fact: evidenced-based decision-making unplugged.* Health Economics, 1997. **6**: p. 547-559.
50. Mackenbach, J.P. and M.J. Bakker, *Tackling socioeconomic inequalities in health: analysis of European experiences.* The Lancet, 2003. **362**: p. 1409-1414.
51. Philips, Z., et al., *Review of guidelines for good practice in decision-analytic modelling in health technology assessment.* Health Technology Assessment, 2004. **8**(36).